Optimal Zero-Delay Jamming Over an Additive Noise Channel

Emrah Akyol, *Member, IEEE*, Kenneth Rose, *Fellow, IEEE*, and Tamer Başar, *Life Fellow, IEEE*

***Abstract*— This paper considers the problem of optimal zero-delay jamming over an additive noise channel. Building on a sequence of recent results on conditions for linearity of optimal estimation, and of optimal mappings in source-channel coding, the saddle-point solution to the jamming problem is derived for general sources and channels, without recourse to Gaussianity assumptions. The linearity conditions are shown to play a pivotal role in jamming, in the sense that the optimal jamming strategy is to effectively force both the transmitter and the receiver to default to linear mappings, i.e., the jammer ensures, whenever possible, that the transmitter and the receiver cannot benefit from non-linear strategies. This result is shown to subsume the known result for Gaussian source and channel. The conditions and general settings where such unbeatable strategy can indeed be achieved by the jammer are analyzed. Moreover, a numerical procedure is provided to approximate the optimal jamming strategy in the remaining (source-channel) cases where the jammer cannot impose linearity on the transmitter and the receiver. Next, the analysis is extended to vector sources and channels. This extension involves a new aspect of optimization: the allocation of available transmit and jamming power over source and channel components. Similar to the scalar setting, the saddle-point solution is derived using the linearity conditions in vector spaces. The optimal power allocation strategies for the jammer and the transmitter have an intuitive interpretation as the jammer allocates power according to water-filling over the channel eigenvalues, while the transmitter performs water-pouring (reverse water-filling) over the source eigenvalues.**

***Index Terms*— Correlated jamming, zero-sum games, zero-delay source-channel coding, linearity conditions, water-filling power allocation.**

I. INTRODUCTION

THE interplay between communication and game theory has been an important research area for decades, e.g., an explicit formulation of communication problem as a game was first proposed more than 50 years ago by Blachman [1]. We consider in this paper the problem of optimal jamming, by a power constrained agent, over an additive noise channel. The jamming problem has traditionally been studied in the asymptotically high delay communication scenarios, using the mutual information of the input and output of the channel as the payoff function see [1]–[8]. Reference [1] is one of the earliest papers to address such a problem using a mutual information payoff. A two-player zero-sum game was explicitly adopted in [2] yielding the Gaussian distribution as a saddle point. In [3], vector strategies were considered in a game-theoretic formulation of communication over channels with block memory, where it was found that memoryless jamming and transmission constitute a saddle point. Such jamming problems, with mutual information payoff, are essentially identical to the Shannon theoretic studies of the capacity of arbitrarily varying channels (AVCs) [9]–[11]. The scalar and vector Gaussian AVCs were studied in [12] and [13] respectively, where a saddle point in the vector case was achieved by a water-filling solution for the jammer and for the transmitter.

This paper builds on our prior work [14], [15], and in contrast with most previous contributions, considers the zero-delay setting, motivated by current and emerging applications such as sensor networks and the smart grid where delay is a critical constraint. We consider the setting where the shared objective of the transmitter and the receiver is to minimize the mean squared error (MSE) between the source and the reconstruction at the receiver, while the jammer aims to maximize this MSE. Another important distinction of this work, beside the zero-delay constraint, is that we consider joint source and channel coding, to avoid the suboptimality of separate source-channel coding at zero-delay settings. This problem of interest was solved in [16] and [17] for the special case of scalar Gaussian source and scalar Gaussian channel, and under various types of information available to the jammer. The analysis was extended to vector Gaussian settings in [18] with linear encoding and decoding mappings. From a game theoretic perspective, the problem can be viewed as a two-player zero-sum game, where allowing mixed strategies for the transmitter and the jammer in this strictly competitive game, one can show that a saddle-point solution exists. Consider for example the specific case of the scalar Gaussian source-channel pair, where the jammer has access to the source, and where the transmitter and the receiver cooperate through a side channel carrying randomization information. The saddle-point solution of this zero-sum game was shown in [17] to comprise: randomized linear mapping for the transmitter; an independent, Gaussian noise as the

Manuscript received May 7, 2014; revised May 19, 2015; accepted May 26, 2015. Date of publication June 15, 2015; date of current version July 10, 2015. This work was supported in part by NSF under Grant CCF-1016861, Grant CCF-1111342, and Grant CCF-1118075, Grant CCF-1320599, and in part by the Air Force Office of Scientific Research through MURI under Grant FA9550-10-1-0573. This paper was presented at the 2013 Allerton Conference on Communication, Control, and Computing, and the 2013 IEEE Conference on Decision and Control.

E. Akyol and T. Başar are with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Champaign, IL 61801 USA (e-mail: akyol@illinois.edu; basar1@illinois.edu).

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

Communicated by S. Avestimehr, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2015.2445344

jammer output; and a linear (conditioned on the randomizing sequence of the transmitter) decoder. In this paper, by leveraging recent results on conditions for linearity of optimal estimation and communication mappings, [19], [20], we extend this analysis to non-Gaussian sources and non-Gaussian channels. The contributions of this paper are thus the following:

- We show that linearity is essential to characterizing the optimal jamming. The jammer, whenever possible, forces the transmitter and the receiver to be linear (conditioned on the randomization random variable). While in the Gaussian source-channel setting, this result is not surprising and corresponds to generating a Gaussian jamming noise, it is quite surprising in general, where the optimal jamming noise is not Gaussian.
- We derive the necessary and sufficient condition (called the "matching condition") on the jamming noise density to ensure linearity of the optimal transmitter and the receiver.
- Based on the matching condition, we derive asymptotic (in terms of low and high channel signal-to-noise ratios (CSNRs)), optimal jamming strategies.
- We present a numerical procedure to approximate the optimal jammer strategy, in cases where a matching jamming density does not exist and the jammer hence cannot force the transmitter and the receiver to be exactly linear.
- Using the necessary and sufficient condition (called the "vector matching condition") on the jamming noise density to ensure linearity of the optimal transmitter and the receiver, within the vector setting, we extend our analysis to vector spaces. The condition is much more involved than in the scalar case due to dependencies across source and channel components.

The paper is organized as follows. In Section II, we present the problem definition and preliminaries. In Section III, we review prior results related to jamming, estimation and communication problems. In Section IV, we derive the linearity result, which leads to our main result. In Section V, we study the implications of the main result, and in Section VI, we present a procedure to approximate the optimal jamming density in the non-matching case. In Section VII, we present our results on the vector extension. We discuss future directions in Section VIII.

II. Problem Definition

Let $\mathbb{R}$ and $\mathbb{R}^+$ denote the respective sets of real numbers and positive real numbers. Let $\mathbb{E}(\cdot)$, $\mathbb{P}(\cdot)$ and $*$ denote the expectation, probability and convolution operators, respectively. Let $Bern(p)$ denote the Bernoulli random variable, taking values in $\{-1, 1\}$ with probability $\{p, 1-p\}$. The Gaussian density with mean μ and variance σ^2 is denoted as $\mathcal{N}(\mu, \sigma^2)$. Let $f^{'}(x) = \frac{df(x)}{dx}$ denote the first-order derivative of the continuously differentiable function $f(\cdot)$. Let $\delta(\cdot, \cdot)$ denote the Kronecker delta function. All logarithms in the paper are natural logarithms and may in general be complex, and the integrals are, in general, Lebesgue integrals. Let us define $\mathcal{S}_m^k$ to denote the set of Borel measurable, square integrable functions $\{\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^k\}$, and use $\mathcal{S}$ for $\mathcal{S}_1^1$ for brevity.

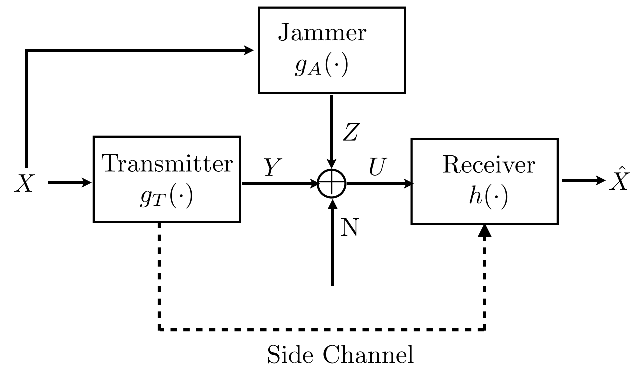

Fig. 1. The jamming problem where mixed strategies are allowed and randomization is transmitted over a side channel.

In general, lowercase letters (e.g., c) denote scalars, boldface ones (e.g., $\boldsymbol{x}$) indicate vectors, and uppercase letters (e.g., C, X) stand for matrices and random variables. I denotes the identity matrix. R_X, and R_{XZ} denote the auto-covariance of $\boldsymbol{X}$ and cross covariance of $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively. Let A^T and $\text{tr}(A)$ respectively denote the transpose and the trace of a matrix A. The elements of a diagonal $m \times m$ matrix A are denoted as $A(i)$, for $i = 1, \ldots, m$. Let $(x)^+$ and $F^\alpha(X)$ denote respectively the function $\max(0, x)$ and α-th power of F, i.e., $(F(X))^\alpha$.

We consider the general communication system whose block diagram is shown in Figure 1. Source $\boldsymbol{X} \in \mathbb{R}^m$ is mapped into $\boldsymbol{Y} \in \mathbb{R}^k$ which is fully determined by the conditional distribution $p(\cdot|\boldsymbol{x})$. For the sake of brevity, and at the risk of slight abuse of notation, we refer to this as a randomized (stochastic) mapping $\boldsymbol{y} = \boldsymbol{g}_T(\boldsymbol{x})$ (i.e., we allow mixed strategies in the problem formulation as in [17] and [18]) so that

$$\mathbb{P}(\boldsymbol{g}_T(\boldsymbol{x}) \in \mathcal{Y}) = \int\limits_{\boldsymbol{y}' \in \mathcal{Y}} p(\boldsymbol{y}'|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad \forall \mathcal{Y} \subseteq \mathbb{R}^k. \tag{1}$$

The adversary has access to the same source signal $\boldsymbol{X}$ and generates the jamming signal $\boldsymbol{Z}$ through a stochastic mapping $\boldsymbol{g_A(\cdot)}$ which is added to the channel output, and aims to compromise the transmission of the source. In addition to the message initiated by the encoder, the decoder also has access to a side channel that provides the randomization sequence, denoted as $\{\gamma\}$ which allows the encoder and the decoder to employ mixed (randomized) strategies. The received signal $\boldsymbol{U} = \boldsymbol{Y} + \boldsymbol{Z} + \boldsymbol{N}$ is mapped by the decoder to an estimate $\hat{\boldsymbol{X}}$ via a function $\boldsymbol{h}(\cdot) \in \mathcal{S}_k^m$. The channel noise $\boldsymbol{N}$ is assumed to be independent of the source $\boldsymbol{X}$ and the randomization signal. The source density is denoted by $f_X(\cdot)$ and the noise density by $f_N(\cdot)$, with characteristic functions $F_X(\boldsymbol{\omega})$ and $F_N(\boldsymbol{\omega})$, respectively. All random variables are assumed to be zero mean.[1] All the statistical properties are given to all agents (the encoder, the decoder and the jammer).

[1]The zero-mean assumption is not essential but simplifies the presentation, and therefore it is made throughout the paper.

The overall cost, measured as the mean squared error (MSE) between source X and its estimate at the decoder, $\hat{X}$, is a function of the transmitter, jammer and the receiver mappings:

$$J(\boldsymbol{g}_T(\cdot), \boldsymbol{g}_A(\cdot), \boldsymbol{h}(\cdot)) = \mathbb{E}\{||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2\}, \tag{2}$$

where expectation is over the statistics of all random variables. Transmitter $\boldsymbol{g}_T(\cdot)$ and receiver $\boldsymbol{h}(\cdot)$ seek to minimize this cost while the adversary (jammer) seeks to maximize it by an appropriate choice of $\boldsymbol{g}_A(\cdot)$. Pre-specified power constraints must be satisfied by the transmitter

$$\mathbb{E}\{||\boldsymbol{g}_T(\boldsymbol{X})||^2\} \leq P_T, \tag{3}$$

and the jammer

$$\mathbb{E}\{||\boldsymbol{g}_A(\boldsymbol{X})||^2\} \leq P_A. \tag{4}$$

The conflict of interest underlying this problem implies that the optimal transmitter-receiver-adversarial policy has to be sought as a saddle-point solution $(\boldsymbol{g}_T^*(\cdot), \boldsymbol{g}_A^*(\cdot), \boldsymbol{h}^*(\cdot))$ satisfying the set of inequalities

$$J(\boldsymbol{g}_T^*, \boldsymbol{g}_A, \boldsymbol{h}^*) \leq J(\boldsymbol{g}_T^*, \boldsymbol{g}_A^*, \boldsymbol{h}^*) \leq J(\boldsymbol{g}_T, \boldsymbol{g}_A^*, \boldsymbol{h}), \tag{5}$$

provided that such a saddle point exists. In this paper, we show the existence of such a saddle point, and its essential uniqueness.[2]

Remark 1: The setting considered here (also the one in [16]) involves a jammer accessing the source input but not the channel input (transmitter output). Such scenarios are of interest where accessing the source is relatively easy (such as a sensor network setting where all sensors observe the same source). The setting where the jammer taps into the communication channel is beyond the scope of the current paper and left as future work (see [17], [18] for analysis of such settings in the presence of scalar Gaussian sources and channels).

Remark 2: The focus of this paper is on communicating a memoryless source, i.e., independent and identically distributed (i.i.d.) random variables as done in prior work (see [16], [17]). Hence, all the mappings and random sources are memoryless. This is in contrast with many conventional settings in control theory, where sources with memory are of interest. Note that for communication purposes, it is reasonable to first solve the memoryless case, since one can block the source process into a sequence of vectors to be sent over a vector channel. Alternatively, the encoder can apply whitening transforms or perform predictive filtering over random processes with memory and treat the independent innovations as an effective memoryless source, consistent with the setting considered here.

[2]The optimal transmitter and receiver mappings are not strictly unique, in the sense that multiple trivially "equivalent" mappings can be used to obtain the same MSE cost. For example, a scalar unit variance Gaussian source and scalar Gaussian channel with power constraint P, can be optimally encoded by either $y = \sqrt{P}x$ or $y = -\sqrt{P}x$. To account for such trivial, essentially identical solutions, we use the term "essentially unique" when a solution is unique up-to sign differences.

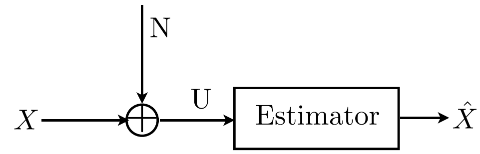

Fig. 2. The estimation problem.

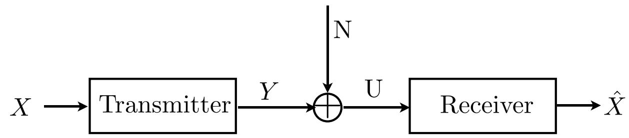

Fig. 3. The communication setting.

III. Prior Work

The jamming problem, in the form defined here, was studied in [16] and [17], for a scalar Gaussian source and a scalar Gaussian channel and in [18] for the vector case. In this paper, we show that jamming problem is intrinsically connected to fundamental problems in estimation theory and the theory of zero-delay source-channel coding. In particular, conditions for linearity of optimal estimation [19] and optimal mappings in communications [20] play a key role in the solution of the general (non-Gaussian) jamming problem considered in this paper. We start with the estimation theoretic viewpoint.

A. A Fundamental Estimation Problem

Consider the one dimensional setting in Figure 2. The estimator receives U, the noisy version of the source X and generates the estimate $\hat{X}$ by employing a function $h : \mathbb{R} \to \mathbb{R}$, selected such that MSE, $\mathbb{E}\{(X - \hat{X})^2\}$ is minimized. It is well-known that, when a Gaussian source is contaminated with Gaussian noise, a linear estimator minimizes the MSE. Recent work [19] analyzed, more generally, the conditions for linearity of optimal estimators. Given a noise (or source) distribution, and a specified channel signal to noise ratio (CSNR), conditions for existence and uniqueness of a source (or noise) distribution for which the optimal estimator is linear were derived.

Here, we restate the basic result whose relevance to the jamming problem will become evident. Specifically, we present the necessary and sufficient condition on source and channel distributions such that the linear estimator $h(U) = \frac{\kappa}{\kappa+1}U$ is optimal, where $\kappa = \frac{\sigma_X^2}{\sigma_N^2}$ is the CSNR.

Theorem 1 [19]: Given a CSNR level κ, and noise N with characteristic function $F_N(\omega)$, there exists a source X for which the optimal estimator is linear if and only if

$$F_X(\omega) = F_N^{\kappa}(\omega). \tag{6}$$

Given a valid characteristic function $F_N(\omega)$, and for some $\kappa \in \mathbb{R}^+$, the function $F_N^{\kappa}(\omega)$ may or may not be a valid characteristic function, which determines the existence of a matching source. For example, the existence of a matching source density is guaranteed for integer κ and it is also guaranteed for infinitely divisible N. More comprehensive discussion of the conditions on κ and $F_N(\omega)$ for $F_N^{\kappa}(\omega)$ to be a valid characteristic function can be found in [19].

B. A Communication Problem

In [20], a communication scenario whose block diagram is shown in Figure 3 was studied. We will focus on a scalar setting (results on the vector extension are presented in Section VII) where source $X \in \mathbb{R}$ is mapped into $Y \in \mathbb{R}$ by a function $g(\cdot) \in \mathcal{S}$, and transmitted over an additive noise channel. The channel output $U = Y + N$ is mapped by the decoder to estimate $\hat{X}$ via function $h(\cdot) \in \mathcal{S}$. The zero-mean noise N is assumed to be independent of the source X and the randomization signal. The source density is denoted by $f_X(\cdot)$ and the noise density by $f_N(\cdot)$, with characteristic functions $F_X(\omega)$ and $F_N(\omega)$, respectively. Note that there is no jammer in the problem formulation and hence a side channel does not exist (no randomization needed).

The objective is to minimize, over the choices of encoder $g(\cdot)$ and decoder $h(\cdot)$, the distortion

$$D = \mathbb{E}\{(X - \hat{X})^2\}, \tag{7}$$

subject to the average transmission power constraint,

$$\mathbb{E}\{g^2(X)\} \leq P_T. \tag{8}$$

A result pertaining to the simultaneous linearity of optimal mappings is summarized in the next theorem.

Theorem 2 [20]: The optimal mappings are either both linear or they are both nonlinear.

The necessary and sufficient condition for linearity of both mappings is given by the following theorem.

Theorem 3 [20]: For a given power constraint P_T, noise N with variance σ_N^2 and characteristic function $F_N(\omega)$, source X with variance σ_X^2 and characteristic function $F_X(\omega)$, the optimal encoder and decoder mappings are linear if and only if

$$F_X(\alpha\omega) = F_N^{\kappa}(\omega) \tag{9}$$

where $\kappa = \frac{P_T}{\sigma_N^2}$ and $\alpha = \sqrt{\frac{P_T}{\sigma_X^2}}$.

Remark 3: In [20], it was shown that the cost function associated with the communication problem is convex in the channel input density $f_Y(\cdot)$ and the set of encoding mappings is compact due to the constraint $\mathbb{E}\{g^2(X)\} \leq P$. Hence, a minimizer for this problem, within the set of encoding mappings, exists. The cost is concave in the density of the channel noise, $f_N(\cdot)$, and $\mathbb{E}\{N^2\} = \sigma_N^2$; hence, a maximizer exists, within this set of channel noises.

C. The Gaussian Jamming Problem (The Scalar Case)

The problem of transmitting independent and identically distributed (i.i.d.) Gaussian random variables over a Gaussian channel in the presence of an additive jammer, depicted in Figure 1, was considered in [16] and [17]. In [17], a game theoretic approach was developed and it was shown that the problem admits a mixed saddle-point solution where optimal transmitter and receiver employ a randomized strategy. The randomization information can be sent over a side channel between the transmitter and the receiver or it could be viewed as the information generated by a third party and observed by both the transmitter and the receiver.[3] Surprisingly, the optimal jamming strategy ignores the input to the jammer and merely generates Gaussian noise, independent of the source. Here we state the main relevant result of [17] which derives the optimal strategy for the transmitter, the adversary and the receiver in Figure 1.

Theorem 4 [16], [17]: The optimal encoding function for the transmitter is randomized linear mapping:

$$g(X) = \gamma\, \alpha_T X, \tag{10}$$

where $\{\gamma\}$ is i.i.d Bernoulli ($\frac{1}{2}$) over the alphabet $\{-1, 1\}$ and $\alpha_T = \sqrt{\frac{P_T}{\sigma_X^2}}$. The optimal jammer generates i.i.d. Gaussian output $\{Z\}$

$$Z \sim \mathcal{N}(0, P_A). \tag{11}$$

where Z is independent of the source X. The optimal receiver is

$$h(U) = \frac{\sigma_X^2}{P_T + P_A + \sigma_N^2} \alpha_T \gamma\, U, \tag{12}$$

and total cost is

$$J = \frac{\sigma_X^2 (P_A + \sigma_N^2)}{P_T + P_A + \sigma_N^2}. \tag{13}$$

Remark 4: In this paper, we study the generalized jamming problem which does not limit the set of sources and channels to Gaussian random variables. As we show in Section IV-B, the linearity property of the optimal transmitter and receiver at the saddle-point solution still holds, while the Gaussianity of the jammer output in the early special case was merely a means to satisfy this linearity condition, and does not hold in general.

Remark 5: The proof of Theorem 4 relies on the fact that for a Gaussian source over a Gaussian channel, zero-delay linear mappings achieve the performance of the asymptotically high delay optimal source-channel communication system [21]. This fact is unique to the Gaussian source-channel pair given that the channel cost constraint is a power constraint and the distortion measure is MSE (see [22] for necessary and sufficient conditions for this fact to hold for general distortion and channel cost measures), hence it is tempting to conclude that the saddle-point solution in Theorem 4 can only be obtained in the "all Gaussian" setting. Perhaps surprisingly, we show that there are infinitely many source-noise pairs that yield a saddle-point solution of this type (see Remark 7).

IV. Main Results-Scalar Setting

A. A Simple Upper Bound From Linear Mappings

In this section, we present a new lemma that is used to upper bound the distortion of any zero-delay communication system. Although the main idea is quite simple, it is nevertheless presented as a separate lemma, due to its operational significance here.

[3] In practice, randomization can be achieved by (pseudo) random number generators at the transmitter and the receiver using the same seed.

Lemma 1: Consider the problem setting in Figure 1. For any given jammer output satisfying the power constraint (4), the minimum distortion achievable by a transmitter-receiver, D, is upper bounded by $D_L = \frac{\sigma_X^2(P_A+\sigma_N^2)}{P_T+P_A+\sigma_N^2}$ *which is determined by second moments, regardless of the shape of the densities.*

Proof: Clearly, encoder and decoder can achieve D_L by utilizing linear mappings that satisfy the power constraint P_T for any source and channel density. Hence, it is straightforward to achieve $D = D_L$ in any source-channel density by using linear mappings. □

Lemma 1 connects the recent results on linearity of optimal estimation and communication mappings to the jamming problem. It implies that the optimal strategy for a jammer that can only control the density of the overall additive noise in the channel (as is the case in our problem due to the side channel allowing randomization) is to force the transmitter and the receiver to use (randomized) linear mappings.

B. The Saddle-Point Solution of the Scalar Jamming Problem

Our main result concerns the optimal strategy for the transmitter, the adversary and the receiver in Figure 1. Let us introduce the quantity

$$\beta \triangleq \frac{P_A + \sigma_N^2}{P_T}. \tag{14}$$

The CNSR associated with the jamming problem is redefined in conjunction with β as $CSNR = \frac{1}{\beta}$. In this section, we make the following assumption.

Assumption 1: $\frac{F_X^\beta(\alpha_T\omega)}{F_N(\omega)}$ *is a valid characteristic function for a given* $\beta \in \mathbb{R}^+$ *and* $\alpha_T \in \mathbb{R}^+$.

The case where Assumption 1 does not hold is analyzed in Section VI. Next, we present our result which pertains to optimal jamming.

Theorem 5: For the jamming problem, the optimal encoding function for the transmitter is randomized linear mapping:

$$g(X) = \gamma\alpha_T X, \tag{15}$$

where $\{\gamma\}$ *is i.i.d. Bernoulli* $(\frac{1}{2})$ *over the alphabet* $\{-1, 1\}$ *and* $\alpha_T = \sqrt{\frac{P_T}{\sigma_X^2}}$. *The optimal jamming function is to generate i.i.d. output* $\{Z\}$ *with characteristic function*

$$F_Z(\omega) = \frac{F_X^\beta(\alpha_T\omega)}{F_N(\omega)} \tag{16}$$

where Z *is independent of the source input* X.

The optimal receiver is

$$h(U) = \frac{\sigma_X^2}{P_T + P_A + \sigma_N^2}\alpha_T\gamma U, \tag{17}$$

and total cost is

$$J = \frac{\sigma_X^2(P_A + \sigma_N^2)}{P_T + P_A + \sigma_N^2}. \tag{18}$$

Moreover, this saddle-point solution is (almost surely) unique.

Proof: We prove this result by verifying that the mappings in this theorem satisfy the saddle-point inequalities given in (5), following the approach in [16]. First, we note that this saddle point exists due to Remark 3.

RHS of (5): Suppose the policy of the jammer is given as in Theorem 5. The communication system at hand becomes identical to the communication problem considered in Section II.B, for which the linear encoder, i.e., $Y = \alpha_T X$ is optimal (see Theorem 3). Any randomized encoder in the form of (15) (irrespective of the density of γ) yields the same cost as the corresponding deterministic encoder and hence is optimal.

LHS of (5): Let us derive the overall cost conditioned on the randomization sequence (i.e., $\gamma = 1$ and $\gamma = -1$) used in conjunction with the decoder given in (17). If $\gamma = 1$,

$$D_1 = \frac{\sigma_X^2(P_A + \sigma_N^2)}{P_T + P_A + \sigma_N^2} + \Psi\mathbb{E}\{ZX\} + \psi\mathbb{E}\{ZN\} \tag{19}$$

for some constants Ψ, ψ, and similarly if $\gamma = -1$,

$$D_2 = \frac{\sigma_X^2(P_A + \sigma_N^2)}{P_T + P_A + \sigma_N^2} - \Psi\mathbb{E}\{ZX\} - \psi\mathbb{E}\{ZN\} \tag{20}$$

where the overall cost is

$$J = \mathbb{P}(\gamma = 1)D_1 + \mathbb{P}(\gamma = -1)D_2. \tag{21}$$

Clearly, for $\gamma \sim Bern(\frac{1}{2})$ overall cost $J = \frac{\sigma_X^2(P_A+\sigma_N^2)}{P_T+P_A+\sigma_N^2}$ is only a function of the second-order statistics of the adversarial outputs, irrespective of the higher order moments of Z; hence the solution presented here is a saddle point.

Toward showing (almost sure) uniqueness, we start by restating the fact that the optimal solution for the transmitter is in the randomized form given in (15). Let us prove the properties that were not covered by the proof of the saddle point:

i) Characteristic Function of Z and Independence of Z of X and N: The choice $F_Z(\omega) = \frac{F_X^\beta(\alpha_T\omega)}{F_N(\omega)}$ renders the transmitter and receiver mappings linear in conjunction with independence of Z and X and N, due to Theorem 3, and it maximizes the overall cost due to Lemma 1.

ii) Choice of Bernoulli Parameter: Note that the optimal choice of the Bernoulli parameter for the transmitters is $\frac{1}{2}$ since other choices will not cancel out the cross terms in (19) and (20). These cross terms can then be exploited by the adversary to increase the cost, and hence an optimal transmitter strategy is to set $\gamma = Bern(1/2)$. □

Remark 6: Theorem 5 subsumes the previous results that focus on the special case of Gaussian source for this setting [16], [17]. When $X \sim \mathcal{N}(0, \sigma_X^2)$, *the unique matching noise, determined by (16) is also Gaussian* $Z \sim \mathcal{N}(0, P_A)$ *for all power levels* P_A *and* P_T. *Hence, Theorem 5 can be viewed as a generalization of Theorem 4. We note in passing that optimality at all power levels,* P_A *and* P_T, *is unique to the Gaussian source-channel pair setting, i.e., the shape of the matching jamming density will, in general, depend on the power levels.*

V. Implications of the Main Result

In this section, we explore some special cases utilizing the matching condition (16). We start with a simple but perhaps surprising result on the existence of linearity achieving jammer.

Corollary 1: If the source X and the channel noise N are identically distributed and β is an integer, then there exists a jammer policy that enforces the optimal mappings to be linear.

Proof: From (16), given that β is an integer and Z is independent of X and N, the matching jammer can be written as

$$Z = \sum_{i=1}^{\beta-1} \nu_i \tag{22}$$

where ν_i are independent and distributed identically to X. Hence, there exists a matching Z. □

A more direct existence result is presented in the following corollary.

Corollary 2: In the case of identically distributed source and channel noise, i.e., $X \sim N$, and $P_T = P_A = \sigma_N^2$, optimal jamming strategy would be generating a random variable identically distributed with X (and N), and optimal transmitter functions are as given in Theorem 5.

Proof: It is straightforward to see from (16) that, at $\beta = 2$, the characteristic functions must be identical, $F_Z(\omega) = F_X(\omega)$, almost everywhere. Since characteristic function uniquely determines the density [23], $Z \sim X$. □

Next, we recall the concept of infinite divisibility, which is closely related to the problem at hand.

Definition [24]: A distribution with characteristic function $F(\omega)$ is called infinitely divisible, if for each integer $k \geq 1$, there exists a characteristic function $F_k(\omega)$ such that

$$F(\omega) = F_k^k(\omega) \tag{23}$$

Alternatively, $f_X(\cdot)$ is infinitely divisible if and only if the random variable X can be written for any k as $X = \sum_{i=1}^{k} X_i$ where $\{X_i, i = 1, \ldots, k\}$ are independent and identically distributed.

Infinitely divisible distributions have been studied extensively in probability theory [24], [25]. It is known that Poisson, gamma, and geometric distributions (and their *mixed* variations) as well as the set of stable distributions (which includes the Gaussian distribution) are infinitely divisible. In the following, we present another matching case.

Corollary 3: In the setting where there is no channel noise N, i.e., $\sigma_N^2 = 0$, if the source is infinitely divisible, there exists a matching jamming noise Z for all $P_T \in \mathbb{R}^+$ and $P_A \in \mathbb{R}^+$.

Proof: We first note that if $f_X(\cdot)$ is infinitely divisible, $F_X^{1/j}(\alpha_T\omega)$ is a valid characteristic function for all natural j and $\alpha_T \in \mathbb{R}^+$, as follows directly from the definition of infinite divisibility. Then, by using the arguments in the proof of Corollary 1, one can show that $F_X^{i/j}(\alpha_T\omega)$ is also a valid characteristic function, which implies that so is $F_X^r(\alpha_T\omega)$ for all positive rational $r > 0$, since a rational r implies that $r = i/j$ for some natural i and j. Using the fact that every $\beta \in \mathbb{R}^+$ is a limit of a sequence of rational numbers r_n, and by the continuity theorem [23], we conclude that $F_Z(\omega) = F_X^{\beta}(\alpha_T\omega)$ is a valid characteristic function, and hence a matching jamming noise exists. □

However, note that at a given CSNR, there may exist a matching jamming noise, even though $f_X(\cdot)$ is not infinitely divisible. For example, a finite alphabet discrete random variable V is not infinitely divisible but still can be k-divisible, where $k < |V| - 1$ and $|V|$ is the cardinality of V. Hence, when $\beta = 1/k$, there may exist a matching jamming density, even when the source distribution is not infinitely divisible.

Remark 7: Corollaries 1, 2, and 3 demonstrate that there is indeed a rich set of source and channel densities that make the optimal mappings linear. Hence, the Gaussianity assumption of the source and channel is not necessary to achieve the saddle-point solution.

Let us next consider a case where the jammer does not need to know the density of the source, i.e., it can perform optimally regardless of the source density.

Corollary 4: At asymptotically low CSNR level, i.e., as $\beta \to \infty$, for a Gaussian channel, the optimal jamming strategy, regardless of the source density, is to generate Gaussian noise that is independent of the source.

Proof: As we have shown in the proof of Theorem 5, the jammer's aim is to force the transmitter and the receiver to use linear mappings. Hence, the matching jamming noise (if exists) satisfies the following:

$$F_Z(\omega)F_N(\omega) = F_X^{\beta}(\alpha_T\omega). \tag{24}$$

As $\beta \to \infty$, RHS of (24) converges to Gaussian characteristic function, due to central limit theorem [23], and hence (16) is asymptotically satisfied. □

Another interesting case is the high CSNR level ($\beta \to 0$) and Gaussian source. The following corollary states our result associated with this setting.

Corollary 5: At an asymptotically high CSNR level, i.e., as $\beta \to 0$, for a Gaussian source, the optimal jamming strategy is to generate noise independent of the source regardless of (either channel or jamming) noise density.

Proof: Again, the matching jamming noise (if exists) must satisfy

$$(F_Z(\omega)F_N(\omega))^{\frac{1}{\beta}} = F_X(\alpha_T\omega). \tag{25}$$

As $\beta \to 0$, LHS of (25) converges to the Gaussian characteristic function and, hence (16) is asymptotically satisfied. □

VI. The Non-Matching Case

We note that given valid characteristic functions $F_X(\alpha_T\omega)$ and $F_N(\omega)$ and for some $\beta \in \mathbb{R}^+$ and $\alpha_T \in \mathbb{R}^+$, the function $\frac{F_X^{\beta}(\alpha_T\omega)}{F_N(\omega)}$ may or may not be a valid characteristic function, which determines the existence of a matching jamming noise density f_Z that enforces linearity on the communication mapping. For example, the existence of a matching jamming density is guaranteed for integer β with $F_X(\alpha_T\omega) = F_N(\omega)$ almost everywhere. Conditions on β and $F_X(\omega)$ for $F_X^{\beta}(\omega)$ to be a valid characteristic function were studied in detail in [19], to which we refer to avoid repetition. In the following, we address this question: what is the optimal jamming noise density $f_Z(\cdot)$, when the jammer cannot make the optimal mappings linear, i.e., $\frac{F_X^{\beta}(\omega)}{F_Z(\omega)}$ is not a valid characteristic function? We first examine the case of the basic estimation setting, and then extend our analysis to the jamming setting.

A. Estimation Setting

The problem of interest as described above appears to be open even in the more fundamental setting, i.e., for estimation problem depicted in Figure 2. Note that there is no encoder or a jammer, and hence no randomization is needed in this problem. We are particularly interested in the noise density $f_N(\cdot)$ that maximizes minimum mean square error, $\mathbb{E}((X-\mathbb{E}(X|U))^2)$. Clearly, if $F_X^{\beta}(\omega)$ is a valid characteristic function, the worst-case noise will have the characteristic function $F_N(\omega) = F_X^{\beta}(\omega)$, and make the optimal (MMSE) estimator linear. Intuitively, it is expected that in the case where $F_X^{\beta}(\omega)$ is *not* a valid characteristic function, the worst-case noise would be the one that forces the optimal estimator to be as close to linear as possible in some sense. The optimal estimator $h(u) = \mathbb{E}\{X|U=u\}$ is given by:

$$h(u) = \frac{\int x\, f_X(x)\, f_N(u-x)\, dx}{\int f_X(x)\, f_N(u-x)\, dx}, \qquad (26)$$

which can also be written as:

$$h(u) = \frac{\int F_X'(\omega) F_N(\omega) e^{ju\omega} d\omega}{\int F_X(\omega) F_N(\omega) e^{ju\omega}\, d\omega}. \qquad (27)$$

We next replace $F_N(\omega)$ and $F_X(\omega)$ with their polynomial expansions, particularly Gram-Charlier expansion over the Gaussian densities $\mathbb{N}(0,\sigma_N^2)$ and $\mathbb{N}(0,\sigma_X^2)$ respectively (see [26] for details):

$$F_N(\omega) = \sum_{m=0}^{M} \left(1 + \frac{\alpha_m}{m!}(jw)^m\right) e^{-\sigma_N^2 w^2/2}, \qquad (28)$$

$$F_X(\omega) = \sum_{m=0}^{M} \left(1 + \frac{\theta_m}{m!}(jw)^m\right) e^{-\sigma_X^2 w^2/2}, \qquad (29)$$

where α_m and θ_m are the polynomial coefficients associated with $F_N(\omega)$ and $F_X(\omega)$, respectively. It is known that these polynomial expansions converge (in mean) to $F_N(\omega)$ and $F_X(\omega)$ as $M \to \infty$. In the following, to render the analysis exact, we assume $M \to \infty$. Plugging (28) and (29) in (27), the optimal estimator is expressed by a ratio of two polynomials:

$$h(u) = \frac{P_a(u)}{P(u)}, \qquad (30)$$

where polynomial $P(u)$ approximates the probability density function $f_U(\cdot)$, i.e., the density of $U = X+N$, and $P_a(u)$ can be computed in terms of α_m and θ_m, for $m = 1, 2, \ldots, M$. Let $\{P_n(u)\}$ be a sequence of polynomials that are orthonormal with respect to $P(u)$, that is

$$\int P_k(u) P_m(u) P(u) du = \delta(m,k), \quad m,k = 0,1,\ldots \qquad (31)$$

Next, $h(u)$ is expanded in terms of $P_m(u)$:

$$h(u) = \sum_{m=0}^{M} c_m P_m(u), \qquad (32)$$

where

$$c_m = \int P_m(u) P_a(u) du. \qquad (33)$$

Then, the MMSE is

$$\begin{aligned} J &= \mathbb{E}((X-\mathbb{E}(X|U))^2) \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X|U))^2 \\ &= \sigma_X^2 - \sum_{m=0}^{M} c_m^2. \end{aligned} \qquad (34)$$

where (34) follows from (31). The worst-case noise maximizes J and hence, minimizes $\sum_{m=0}^{M} c_m^2$. Note that $c_0 = 0$ and $c_1 = \sqrt{\frac{\sigma_X^2}{\sigma_X^2+\sigma_N^2}}$. These two coefficients are determined by the second-order statistics of the source and the noise, while higher order coefficients, $c_m, m \geq 2$, depend on higher order statistics. Note also that the polynomials associated with these coefficients are $P_0(u) = 1$ and $P_1(u) = \sqrt{\frac{\sigma_X^2}{\sigma_X^2+\sigma_N^2}}u$. We next present our main result regarding this setting.

Lemma 2: The worst-case noise minimizes $\sum_{m=0}^{M} c_m^2$, where $\{c_m\}$ are the coefficients of the orthonormal polynomial expansion with density $f_U(\cdot)$.

Given the source density, we can approximate the optimal estimator used in conjunction with the worst-case noise. In the following, we focus on finding the worst-case noise that matches this estimator.

Let us assume $h(u) = \sum_{m=0}^{M} b_m u^m$ for $b_m \in \mathbb{R}$. Then, the following holds

$$\sum_{m=0}^{M} b_m u^m = \frac{\int x\, f_X(x)\, f_N(u-x)\, dx}{\int f_X(x)\, f_N(u-x)\, dx}. \qquad (35)$$

Expanding (35), and expressing integrals as convolutions, we have

$$\sum_{m=0}^{M} b_m u^m (f_X(u) * f_N(u)) = (u f_X(u)) * f_N(u). \qquad (36)$$

Taking the Fourier transforms of both sides, we obtain

$$\sum_{m=0}^{M} b_m \frac{d^m}{d\omega^m}(F_X(\omega)F_N(\omega)) = F_X'(\omega) F_N(\omega) \qquad (37)$$

Hence, given the optimal estimator, we can find the approximate worst-case noise by solving the differential equation given in (37). Note that throughout our analysis we assumed that $M \to \infty$. In practice, M is a fixed, finite quantity, which is determined by the allowed complexity of the approximation scheme.

B. Jamming Setting

Let us now return to the original problem of jamming. Throughout our analysis, for brevity in the notation, we omit the randomization factor γ in the encoding and decoding policies, although we still study the same problem where randomization is allowed, and hence the source input is useless to the jammer and the jammer's action space is limited (without any loss of generality) to additive noise. We carry out a similar analysis to approximate the best M^{th} order polynomial expansion of the decoder, given the encoder. For simplicity, and to enable the subsequent derivations, we have the following assumption throughout this section.

Assumption 2: The transmitter function is linear, i.e., $g(X) = \alpha_T X$.

Note however that as Theorem 2 implies, if the optimal decoder is nonlinear so must be the encoder. However, the jammer tries to render the optimal mappings linear, hence it is expected that the encoder at the saddle-point solution is close to linear. Even though taking the encoder to be linear is not accurate in the strict sense, it is a reasonable assumption used in the development of our approximation to the optimal solution.

We follow the same steps as in the estimation setting, starting with the derivation of the optimal M^{th} order polynomial approximation of the decoder. We again assume that $M \to \infty$ throughout the analysis to make the approximations exact. The optimal decoder can be expressed as:

$$h(u) = \frac{\displaystyle\int x\, f_X(x)\, f_{N+Z}(u - \alpha_T x)\, dx}{\displaystyle\int f_X(x)\, f_{N+Z}(u - \alpha_T x)\, dx} \tag{38}$$

where $\alpha_T = \sqrt{\frac{P_T}{\sigma_X^2}}$ and f_{N+Z} is the density of $N + Z$. Noting that X and Z are independent, we have

$$h(u) = \frac{\displaystyle\int F'_X(\alpha_T\omega) F_N(\omega) F_Z(\omega) e^{ju\omega} d\omega}{\displaystyle\int F_X(\alpha_T\omega) F_N(\omega) F_Z(\omega) e^{ju\omega}\, d\omega} \tag{39}$$

Plugging the appropriate polynomial expressions for $F_X(\omega)$, $F_N(\omega)$, and $F_Z(\omega)$, we express $h(u)$ as:

$$h(u) = \frac{P_b(u)}{P_c(u)}, \tag{40}$$

where $P_b(u)$ and $P_c(u)$ are polynomials. Again, expanding $h(u)$ in terms of the polynomials which are orthonormal under the measure $P_c(u)$ (which is the density of $\alpha_T\ X + Z + N$), and following the same steps that led to (34), we obtain

$$J = \sigma_X^2 - \sum_{m=0}^{M} c_m^2, \tag{41}$$

where c_m's are the coefficients of the polynomials that are orthonormal with respect to the density of the channel output $U = \alpha_T X + Z + N$. Hence, we can approximate the optimal jamming density as the one that minimizes $\sum_{m=0}^{M} c_m^2$. Note that in our analysis we assumed that $M \to \infty$, while in practice a fixed, finite M can be used for approximation purposes. Similar to the estimation setting, once the best polynomial approximation is found, an approximation to the optimal jamming density can be obtained by solving a differential equation which can be obtained following the same steps that yielded (37).

A Numerical Example: Let us demonstrate this numerical approximation procedure with an example. We emphasize that the objective of this example is not to obtain theoretical optimality results or devise a numerically optimal jamming system, but rather to present a picture of how the numerical approach here can be used. We focus on a simple setting, e.g., let the source density deviate from a Gaussian density as follows:

$$f_X(x) = \left(1 + \frac{\epsilon}{4!} H_4(x)\right) f_G(x) \tag{42}$$

where $f_G(\cdot)$ is the Gaussian density with zero-mean and unit variance, $H_4(x)$ is the Hermite polynomial of order 4 and we have[4] $0 \le \epsilon \le 4$.

For simplicity, let us first assume that the channel is noiseless (we will briefly consider the noisy channel case later in the example), and also $P_A \approx P_T \approx 1$, i.e., deviations ($\epsilon$) are small enough. In the following, we will approximate the jamming density, the (nonlinear) decoder and MSE cost at the saddle-point solution, for a simple case of $M = 3$. First, we note that by employing a linear encoder approximation, we effectively end up with the estimator approximation problem analyzed in Section VI-A. Let us denote the jamming noise density also in the form of Gaussian density perturbed by Hermite polynomials as

$$f_Z(z) = \left(1 + \sum_{m=4}^{\infty} \frac{\eta_m}{m!} H_m(z)\right) f_G(z) \tag{43}$$

where H_m is the Hermite polynomial of order m and $\eta_m = 0$ for $m = 1, 3, 5 \ldots$. Approximating U as Gaussian, we have Hermite polynomials as $\{P_n(u)\}$ as a set of basis polynomials that the decoder $h(u)$ is represented in, since the Hermite polynomials are orthonormal under the Gaussian density. Then, we have, by noting that the odd moments vanish due to symmetry,

$$P_0(u) = 1, \tag{44}$$

$$P_1(u) = \frac{1}{\sqrt{2}} u, \tag{45}$$

$$P_2(u) = \frac{\mu_2 u^3 - \mu_4 u}{\sqrt{\mu_2(\mu_2\mu_6 - \mu_4^2)}} u, \tag{46}$$

where $\mu_i = \mathbb{E}\{U^i\}$ for $i = 2, 4, 6$. Computing these moments explicitly, we have

$$\mu_2 = 2, \tag{47}$$

$$\mu_4 = 12 + \epsilon + \eta_4, \tag{48}$$

$$\mu_6 = \eta_6 + 30\mu_4 - 240. \tag{49}$$

[4] This condition on ϵ is needed to ensure that $f_X(x) \ge 0$.

The polynomial coefficients c_n are

$$c_1 = 0, \tag{50}$$

$$c_2 = \frac{1}{\sqrt{2}}, \tag{51}$$

$$c_3 = \frac{\epsilon - \eta_4}{\sqrt{\mu_2(\mu_2\mu_6 - \mu_4^2)}}. \tag{52}$$

The MSE cost can be then computed using (34) as

$$J = \frac{1}{2} - \frac{1}{2}\frac{(\epsilon - \eta_4)^2}{(2\mu_6 - \mu_4^2)}. \tag{53}$$

We first note that $J \le J_L = \frac{1}{2}$. The jammer is trying to maximize this cost (achieve $J = J_L$), hence optimal jammer sets $\eta_4 = \epsilon$ and achieves $J = J_L = \frac{1}{2}$ and sets all higher order moments to zero ($\eta_m = 0, \forall m \ge 6$) to minimize the jamming power. Hence, the jammer is essentially rendering the jamming noise density identical to that of the source. Noting also that $P_T \approx P_A \approx 1$, this result is intuitively expected from the symmetry (by similar reasoning as done in the proof of Corollary 2).

Next, let us consider the more practical case of noisy channel. We again assume a 3^{rd} order approximation ($M = 3$) of the decoder and total channel noise variance is approximately equal to transmit power and source variance, i.e., $P_A + \sigma_N^2 \approx P_T \approx 1$. In this case, if the channel noise density is such that the jammer cannot make the effective channel noise density (the density of $N + Z$) identical to the source density, i.e., $\eta_4 \neq \epsilon$ (where η_4 is the perturbation associated with the effective channel noise $(N+Z)$), from (53) we conclude that it chooses η_4 and η_6 to minimize $\frac{(\epsilon-\eta_4)^2}{(2\mu_6-\mu_4^2)}$.

Further numerical analysis, given the channel noise density, can be done by following approach presented in this section and routine procedures (see [26] for details).

VII. Vector Extensions

In this section, we extend our results obtained for the scalar setting to higher dimensional source and channel spaces. We first need two auxiliary lemmas in matrix analysis and majorization theory, and some prior results on optimal linear coding.

A. Background

Throughout this section, we assume that the effective dimension[5] of the source is identical to that of the channel noise (i.e., no bandwidth compression or expansion). This assumption is essential in the sense that when source and channel dimensions do not match, the jammer cannot ensure linearity of encoding and decoding mappings. A well-known example is 2:1 bandwidth compression where the optimal mappings are highly nonlinear, even in the case of Gaussian source and channel (see [20], [27] and the references therein for details). Let us assume that the source and the channel are m-dimensional vectors with respective covariances R_X and R_N, where R_X and R_N allow the diagonalization

$$R_X = Q_X \Lambda_X Q_X^T, \text{ and } R_N = Q_N \Lambda_N Q_N^T \tag{54}$$

where $Q_X Q_X^T = Q_N Q_N^T = I$, and Λ_X and Λ_N are diagonal matrices, having ordered eigenvalues as entries, i.e., $\Lambda_X = \text{diag}\{\boldsymbol{\lambda}_X\}$ and $\Lambda_N = \text{diag}\{\boldsymbol{\lambda}_N\}$ where $\boldsymbol{\lambda}_X$ and $\boldsymbol{\lambda}_X$ are ordered (descending) eigenvalues and Q_X^T and Q_N^T are the eigenmatrices of the source and the channel, respectively. We will make use of the following auxiliary lemma; see [28] for a proof.

Lemma 3: Let $\boldsymbol{\lambda}_X$ and $\boldsymbol{\lambda}_N$ be two ordered vectors in $\mathbb{R}_+^m$ with descending entries $\lambda_X(1) \ge \lambda_X(2), \ldots, \lambda_X(m)$, and $\lambda_N(1) \ge \lambda_N(2), \ldots, \lambda_N(m)$; and Π denote any permutation of the indices $\{1, 2, \ldots, m\}$. Then,

$$\min_{\Pi} \sum_{i=1}^{m} \lambda_X(\Pi(i))\lambda_N(i) = \sum_{i=1}^{m} \lambda_X(i)\lambda_N(m-i) \tag{55}$$

and

$$\max_{\Pi} \sum_{i=1}^{m} \lambda_X(i)\lambda_N(\Pi(i)) = \sum_{i=1}^{m} \lambda_X(i)\lambda_N(i) \tag{56}$$

Toward deriving the vector extension, we need the optimal encoding and decoding transforms for a general communication problem with source and channel noise covariances R_X and R_N, and total encoding power limit P_T. Here, we first state the classical result due to [29] (see also [30]–[32] for alternative derivations of this result).

Theorem 6 [29]–[32]: The encoding transform that minimizes the MSE distortion subject to the power constraint P_T is

$$C = Q_N \Sigma Q_X^T \tag{57}$$

where Σ is a diagonal power allocation matrix. Moreover the total MSE distortion as a function of source and channel eigenvalues is

$$J(\boldsymbol{\lambda}_X, \boldsymbol{\lambda}_N) = \frac{\left(\sum\limits_{i=1}^{w} (\sqrt{\lambda_X(i)\lambda_N(m-i)})\right)^2}{P_T + \sum\limits_{i=1}^{w} \lambda_N(m-i)} + \sum_{w+1}^{m} \lambda_X(i) \tag{58}$$

where w is the number of active channels determined by the power P_T.

Remark 8: The distortion expression (58) has an interesting interpretation of power allocation as "reverse water-filling" over the source eigenvalues. As we will show in the next section, the optimal jammer also performs power allocation as water-filling over the channel eigenvalues.

Remark 9: Note that the ordering of the eigenvalues is such that the largest source eigenvalue is multiplied by the smallest noise eigenvalue and so on, which physically means that the encoder uses the best channel for the smallest variance source component. This is a direct consequence of Lemma 3.

Assumption 3: P_T is high enough, so that all source and channel components are active (no channel is allocated

[5] The effective dimension refers to the number of source components actually transmitted over the channel. This number depends on the source and channel dimensions, and the encoding power P_T.

zero transmit power), i.e., $w = m$, and hence (58) can be rewritten as

$$J(\boldsymbol{\lambda}_X, \boldsymbol{\lambda}_N) = \frac{\left(\sum_{i=1}^{m} \sqrt{\lambda_X(i)\lambda_N(m-i)}\right)^2}{P_T + \sum_{i=1}^{m} \lambda_N(i)} \tag{59}$$

Assumption 3 is not necessary but it leads to substantial simplification in the results.

Assumption 4: There exists a matching vector noise, i.e., the problem parameters are such that optimal encoding and decoding mappings can be made linear.

The precise condition for the existence of the matching specified in Assumption 4 is presented in Theorem 8.

Next, we state two results in matrix analysis, which we will prove using majorization theory [33]. These results admit shorter proofs by contradiction, i.e., by showing that any solution other than those stated in the lemmas will incur strictly higher cost. Nevertheless, we proceed with constructive proofs using majorization. The needed background on majorization theory is provided in Appendix A.

Lemma 4: Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be vectors in $\mathbb{R}_+^n$, with $0 < a(i) \leq a(i+1)$ for $i = 1, 2 \ldots m-1$ and $0 < b(i) \leq b(i+1)$ for $i = 1, 2 \ldots m-1$. The globally optimal solution to the convex optimization problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & f(\boldsymbol{x}) = \sum_{i=1}^{n} a(i)\,(x(i)+b(i)) \\ \text{subject to} \quad & x(i)+b(i) \geq x(i+1)+b(i+1), \\ & \sum_{i=1}^{n} x(i) = P \text{ and } x(i) \geq 0, \quad \forall i \end{aligned}$$

is given by

$$x(i) = (\theta - b(i))^+ \tag{60}$$

where

$$\sum_{i=1}^{n} (\theta - b(i))^+ = P \tag{61}$$

Proof: From an intuitive viewpoint, the choice of $\boldsymbol{x}$ will aim to make $\boldsymbol{x}+\boldsymbol{b}$ uniform, e.g., if $\boldsymbol{b}$ is uniform, the uniform selection of $\boldsymbol{x}$ i.e, $x(i) = P/n$ will be optimal. This is a direct consequence of well-known results on the uniformity of solutions to such optimization problems, see [34]. The optimal solution aims to make $\boldsymbol{x}+\boldsymbol{b}$ as close to uniform as possible, which implies that it maximizes the smallest element of $\boldsymbol{x}+\boldsymbol{b}$: the well-known water-filling solution.

In the following, we formally prove this lemma using majorization theory [33]. First, we rewrite the objective as $\sum f_i(x(i))$ where $f_i(x(i)) = a(i)\,(x(i)+b(i))$. Noting that $f_i'(\alpha) \leq f_{i+1}'(\beta)$ whenever $\alpha \geq \beta$, we note that $f(x)$ is Schur concave (see [33, 3.H.2]). Uniform $x(i) = P/n, \forall i$ is majored by any other $\boldsymbol{x}$ satisfying $\sum_{i=1}^{n} x(i) = P$. Hence, if this choice satisfies the constraint, $x(i) + b(i) \geq x(i+1) + b(i+1)$, it would be the solution to the optimization problem above. In general, it is straightforward to show that $x(i) = (\theta - b(i))^+$ is majored by any other $\boldsymbol{x}$ satisfying the constraints and hence it maximizes the objective function. This result can also be proved using convex optimization theory [35]. □

Corollary 6: Let the channel noise covariance R_N be fixed. Let $\boldsymbol{\lambda}_{N+Z}$ denote the eigenvalues of the matrix $R_N + R_Z$, $\boldsymbol{\lambda}_N$ denote the ordered (ascending) eigenvalues of R_N, and $\boldsymbol{\lambda}_Z$ denote the inverse ordered eigenvalues of R_Z. Also assume that $\boldsymbol{\lambda}_Z$ is chosen so that it performs water filling over $\boldsymbol{\lambda}_N$, i.e., $\lambda_N(i) = (\theta - \lambda_Z(i))^+$. Then, $\boldsymbol{\lambda}_{N+Z}$ majorizes $\boldsymbol{\lambda}_N + \boldsymbol{\lambda}_Z$ for any R_Z that satisfies a power constraint $tr(R_Z) \leq P$.

Proof: The proof follows directly from Lemma 4 and Lemma 6 in Appendix A. □

B. An Upper Bound on Distortion

In this section, we extend Lemma 1 to the non-scalar case. Note that a variation of Lemma 5 appeared in [18, Th. 3.1] where optimal jamming, encoding and decoding strategies are derived for vector Gaussian settings. In [18], the policies are restricted to (randomized) linear strategies, while in this paper, we do not *a priori* restrict encoding and decoding mappings to be linear[6] and show that linearity is a natural consequence of the problem formulation.

Lemma 5: Consider the problem setting in Figure 1. For any given jammer $\mathbf{Z}$, the distortion achievable by the transmitter-receiver, D, is upper bounded by D_L, i.e., $D_L \geq D$ where

$$D_L = \frac{\left(\sum_{i=1}^{m} \sqrt{\lambda_X(i)\max(\theta, \lambda_N(m-i))}\right)^2}{P_T + \sum_{i=1}^{m} \max(\theta, \lambda_N(i))} \tag{62}$$

and θ satisfies the water-filling condition:

$$\sum_{i=1}^{m} (\theta - \lambda_N(i))^+ = P_A. \tag{63}$$

Proof: Similar to the scalar case (Lemma 1), the key idea of the proof is to make use of the fact that the transmitter and the receiver can always use randomized linear mappings that satisfy the power constraints, hence the MSE resulting from linear mappings will constitute an upper bound on the actual distortion. In the rest of the proof, we show that optimal linear solution yields (62), specifically i) the jammer performs water-filling power allocation, and ii) the jammer aligns its eigenvalues in the channel noise subspace.

Before proceeding further, we give an intuitive explanation of why the jammer will perform water-filling power allocation. Due to the symmetry of the problem, if the jammer does not use a water-filling solution (trying to make λ_{Z+N} as close to uniform as possible), the transmitter will perform an inverse water-filling, i.e., it will redistribute source eigenvalues $\boldsymbol{\lambda}_X$ so that the largest source eigenvalue is aligned with the smallest effective λ_{Z+N} (see Lemma 3). Therefore, maximizing the smallest effective channel eigenvalue $\boldsymbol{\lambda}_{N+Z}$ intuitively seems to be appealing for the jammer. Indeed, this approach precisely

[6]By "linearity", we mean here and throughout the rest of this section, linear maps for each value of the randomization parameter.

describes the water-filling power allocation for jammer over the noise eigenvalues $\boldsymbol{\lambda}_N$.

The formal proof directly follows from Lemma 4 and Corollary 6. Lemma 6 guarantees that the jammer sets $R_Z = Q_N \Lambda_Z Q_N^T$, i.e., the eigenvectors of the noise and the jammer must be aligned. This result can be viewed as a consequence of the well-known optimality of the diagonalizing structure [34], [36]. Lemma 4 ensures that the quantity

$$\min_{\Pi} J(\boldsymbol{\lambda}_X(\Pi), \boldsymbol{\lambda}_N + \boldsymbol{\lambda}_Z) \tag{64}$$

is maximized when $\boldsymbol{\lambda}_Z$ performs water filling over $\boldsymbol{\lambda}_N$. □

Remark 10: The bound, presented in Lemma 5 is determined by second-order statistics, and, otherwise, does not depend on the shape of the densities.

Remark 11: The optimal jammer performs water-filling over the channel eigenvalues while the encoder allocates power according to reverse water-filling over the source eigenvalues. This observation parallels the information theoretic (at asymptotically long delay) water-filling duality, where the rate distortion optimal vector encoding scheme allocates the rate by reverse water-filling over the source eigenvalues, and vector channel capacity achieving scheme allocates power over the channel eigenvalues by water-filling (see [21]).

Remark 12: Lemma 5 is the key result that connects the recent results on linearity of optimal estimation and communication mappings to the jamming problem in the vector case. Lemma 5 implies that the optimal strategy for a jammer which can only control the density of the additive noise channel, is to force the transmitter and receiver to use (randomized) linear mappings.

C. Conditions for Linearity of Communication Mappings in Vector Spaces

In order to make use of Lemma 5, we need the conditions for linearity of optimal encoder and decoder mappings in non-scalar settings. We make use of the following result that appeared in [20].

Theorem 7 [20]: Let the characteristic functions of the transformed source and noise ($\Sigma Q_X^T \boldsymbol{X}$ and $Q_N^T \boldsymbol{N}$) be $F_{\Sigma Q_X^T X}(\boldsymbol{\omega})$ and $F_{Q_N^T N}(\boldsymbol{\omega})$, respectively. The necessary and sufficient condition for linearity of optimal mappings is:

$$\frac{\partial \log F_{\Sigma Q_X^T X}(\boldsymbol{\omega})}{\partial \omega_i} = S_i \frac{\partial \log F_{Q_N^T N}(\boldsymbol{\omega})}{\partial \omega_i}, \quad 1 \le i \le m \tag{65}$$

where S_i are the elements of the diagonal matrix $S \triangleq \Sigma \Lambda_X \Sigma \Lambda_N^{-1}$.

Further insight into the above necessary and sufficient condition is provided via the following corollaries. The first one states that the scalar matching condition, necessary and sufficient for linearity of optimal mappings, is also a necessary condition for each source and channel component in the transform domain.

Corollary 7 [20]: Let $F_{[\Sigma Q_X^T X]_i}(\omega)$ and $F_{[Q_N^T N]_i}(\omega)$ be the marginal characteristic functions of the transform coefficients $[\Sigma Q_X^T \boldsymbol{X}]_i$ and $[Q_N^T \boldsymbol{N}]_i$, respectively. Then, a necessary condition for linearity of optimal mappings is:

$$F_{[\Sigma Q_X^T X]_i}(\omega) = F^{S_i}_{[Q_N^T N]_i}(\omega), \quad 1 \le i \le m \tag{66}$$

Another set of necessary conditions is presented in the following corollary.

Corollary 8 [20]: A necessary condition for linearity of optimal mappings is that one of the following holds for every pair i, j, $1 \le i, j \le m$:

- *i) $S_i = S_j$*
- *ii) $[Q_X^T \boldsymbol{X}]_i$ is independent of $[Q_X^T \boldsymbol{X}]_j$ and $[Q_N^T \boldsymbol{N}]_i$ is independent of $[Q_N^T \boldsymbol{N}]_j$.*

Note that the above corollaries focus on the necessary conditions. In the following, we present a sufficient condition.

Corollary 9 [20]: If the necessary condition of Corollary 7 is satisfied, then a sufficient condition for linearity of optimal estimation is that the transform coefficients $Q_X^T \boldsymbol{X}$ and $Q_N^T \boldsymbol{N}$ are both independent.

Remark 13: While the condition in Corollary 9 requires independence of transform coefficients, the weaker property of uncorrelatedness is already guaranteed by the use of eigen transformations.

Corollary 10 [20]: For a vector Gaussian source and channel, linear mappings are optimal, irrespective of the covariance matrices and allowed power.

Remark 14: Linear mappings are optimal for a Gaussian vector source and channel pair, in the zero-delay setting, but they are not, in general, optimal from an information theoretic point of view (asymptotically high delay), see [21]. This observation highlights the difference between the problem setting considered here and that considered in prior work where mutual information is the objective function [1]–[8].

D. Main Result-Vector Setting

The following theorem presents the optimal strategy for the transmitter, the adversary and the receiver shown in Figure 1, as extended to vector spaces (for the non-scalar case).

Theorem 8: For the jamming problem, the optimal encoding function for the transmitter is:

$$\boldsymbol{g}(\boldsymbol{X}) = C\boldsymbol{X}, \tag{67}$$

where $C = Q_N \Gamma \Sigma Q_X^T$ and $\{\Gamma\}$ is a zero-mean $m \times m$ diagonal matrix, distributed independently of X and with a symmetric density that satisfies $\Gamma(i) \sim Bern(\frac{1}{2})$ over the alphabet $\{-1, 1\}$. The optimal jamming function is to generate i.i.d. output $\{\boldsymbol{Z}\}$, independent of $\boldsymbol{X}$, that satisfies:

$$\frac{\partial \log F_{\Sigma Q_X^T X}(\boldsymbol{\omega})}{\partial \omega_i} = S_i \frac{\partial \log F_{Q_N^T (N+Z)}(\boldsymbol{\omega})}{\partial \omega_i}, \quad 1 \le i \le m \tag{68}$$

for

$$S = \Sigma \Lambda_X \Sigma \Lambda_Z^{-1} \tag{69}$$

and

$$R_Z = Q_N \Lambda_Z Q_N^T \tag{70}$$

where Λ_Z is a diagonal $m \times m$ matrix with diagonal elements

$$\lambda_Z(i) = (\theta - \lambda_N(i))^+ \tag{71}$$

and θ satisfies the water-filling condition:

$$\sum_{i=1}^{m} (\theta - \lambda_N(i))^+ = P_A \tag{72}$$

The optimal receiver is

$$\boldsymbol{h}(\boldsymbol{U}) = R_X C^T (C R_X C^T + R_N + R_Z)^{-1} \boldsymbol{U}, \tag{73}$$

and total cost is

$$J = \frac{\left(\sum_{i=1}^{m} \sqrt{\lambda_X(i) \max(\theta, \lambda_N(m-i))}\right)^2}{P_T + \sum_{i=1}^{m} \max(\theta, \lambda_N(i))} \tag{74}$$

Moreover, this saddle-point solution is (almost surely) unique.

Proof: The proof follows from verification of the saddle-point inequalities given in (5), and is an extension of the scalar result in Theorem 5. It is in the spirit of the proof given in [18] for the Gaussian case, extended here to the non-Gaussian case.

RHS of (5): Suppose the policy of the jammer is given as specified in the theorem statement. The communication system at hand becomes identical to the one in Section VII-C. Using Theorem 7, we conclude that if (68) is satisfied, the linear encoder $Y = Q_N \Gamma \Sigma Q_X^T \boldsymbol{X}$ is optimal (see Theorem 6) for any diagonal randomization matrix that satisfy the power constraint. Since elements of Γ are Bernoulli over the alphabet $\{-1, 1\}$, $\boldsymbol{Y} = Q_N \Gamma \Sigma Q_X^T \boldsymbol{X}$ (irrespective of the joint density of $\Gamma(i)$) yields the same MSE and power cost as $\boldsymbol{Y} = Q_N \Sigma Q_X^T \boldsymbol{X}$ (in conjunction the optimal decoder for this encoder) and hence is optimal.

LHS of (5): Let us derive the overall cost conditioned on the randomization matrix realizations (Γ) used in conjunction with the decoder given in (73). Conditioned on a realization of Γ,

$$J(\Gamma) = \frac{\left(\sum_{i=1}^{m} \sqrt{\lambda_X(i) \max(\theta, \lambda_N(m-i))}\right)^2}{P_T + \sum_{i=1}^{m} \max(\theta, \lambda_N(i))} + \xi(\Gamma)\,\mathrm{tr}(\Gamma^T \mathbb{E}\{\boldsymbol{ZX}\}) + \phi(\Gamma)\,\mathrm{tr}(\Gamma^T \mathbb{E}\{\boldsymbol{NX}\}) \tag{75}$$

for some $\xi(\Gamma), \phi(\Gamma) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ that are even functions of Γ, i.e., $\xi(\Gamma) = \xi(-\Gamma)$ and $\phi(\Gamma) = \phi(-\Gamma)$, $\forall \Gamma$. Taking an expectation over Γ, interchanging trace and the expectation operators, and noting that for a symmetric distribution over Γ, all terms except the first in (75) vanish, and hence we have (74). Noting that J is only a function of the second-order statistics of the adversarial outputs, irrespective of the higher order moments of $\boldsymbol{Z}$, we conclude that the presented solution is a saddle-point. Similar to the scalar case, we need the following analysis for the essential uniqueness of the saddle-point solution:

i) Characteristic Function of $\boldsymbol{Z}$ and Independence of $\boldsymbol{Z}$ of $\boldsymbol{X}$ and $\boldsymbol{N}$: The condition in (68) renders the transmitter and receiver mappings (randomized) linear in conjunction with independence of $\boldsymbol{Z}$ and $\boldsymbol{X}$ and $\boldsymbol{N}$ which maximizes MSE (see Lemma 5). Note that correlating $\boldsymbol{Z}$ with $\boldsymbol{X}$ and $\boldsymbol{N}$ cannot increase overall cost since terms involving $\mathbb{E}\{\boldsymbol{ZN}\}$ and $\mathbb{E}\{\boldsymbol{ZX}\}$ in (75) cancel out due to the choice of randomization Γ.

ii) Choice of Randomization: The choice of Γ must cancel out the cross terms in (75). These cross terms can be exploited by the adversary to increase the cost, and hence an optimal strategy for transmitter is to set $\Gamma(i) \sim Bern(1/2)$ with a symmetric joint distribution over Γ, i.e., $\mathbb{P}(\Gamma) = \mathbb{P}(-\Gamma), \forall \Gamma$. □

Remark 15: The randomization choice for Γ includes two extreme randomization methods as special cases. The first one is using a scalar randomization (replacing Γ with γ in the solution of the scalar problem in Theorem 5) which corresponds to fully correlated $\Gamma(i)$, and hence randomizing the entire source vector at once. The second one is using independent $\Gamma(i)$, which corresponds to randomizing each component independently. Both of these solutions render useless the source information available to the jammer and yield the same saddle-point solution.

In this section, we assumed that there exists a matching jamming density that satisfies (68). An approximation to the matching jamming density can be numerically obtained, extending the approach in Section VI to vector spaces. For brevity, and to avoid repetition, we omit the details.

VIII. Discussion

In this paper, we have studied the problem of optimal zero-delay jamming over an additive noise channel. We first studied the scalar case and showed that linearity is essential to the jamming problem, in the sense that the optimal jamming strategy is to effectively force both optimal encoder and decoder to employ (randomized) linear mappings. We analyzed conditions and general settings where such a strategy can indeed be implemented by the jammer, and provided a "matching condition" which, if satisfied, guarantees linearity of optimal encoder and decoder mappings. Moreover, we provided a procedure to approximate optimal jamming strategy in the cases where the jammer cannot impose linearity on the encoder and the decoder. Intuitively, the jammer approximates the matching solution in terms of an expansion in polynomials that are orthogonal under the measure of the channel output.

Next, we extended the analysis to (higher dimensional) vector settings. Similar to the scalar setting, linearity conditions for encoding and decoding mappings play a key role in the vector jamming problem. We showed that the optimal strategy is randomized linear encoding and decoding for the transmitter and the receiver, and independent noise for the jammer. The eigenvalues of the optimal jamming noise are allocated according to water-filling over the eigenvalues of the channel noise, and the density of the jamming noise is matched to the source and the channel so as to render the optimal mappings linear. We derived the matching condition to be satisfied by the jamming noise. The power allocation solutions in the zero-delay problems (water-filling for jammer and reverse water-filling for the transmitter) parallel the resource allocation strategies in asymptotically high delay (Shannon type) problems, such as rate allocation in rate-distortion (reverse water-filling) and power allocation in channel capacity (water-filling).

Some directions for future work include extensions to network settings (see [37], [38] for preliminary results in this direction), and to sources and channels with memory, and also a detailed numerical study on the optimal approximation of encoding, decoding and jamming strategies in the non-matching case analyzed in Section VI.

Appendix A
Majorization Theory: Basic Concepts

This appendix summarizes a few basic concepts in majorization which are useful in this paper (see [33] for complete reference on majorization). Let $\boldsymbol{x} \in \mathbb{R}^m$ be such that $x(1) \geq x(2) \geq \ldots \geq x(m)$.

Majorization: $\boldsymbol{y} \in \mathbb{R}^m$ majorizes $\boldsymbol{x}$ if and only if

$$\sum_{i=1}^{k} y(i) \geq \sum_{i=1}^{k} x(i), \ 1 \leq k \leq m \tag{76}$$

$$\sum_{i=1}^{m} y(i) = \sum_{i=1}^{m} x(i), \tag{77}$$

which is denoted as $\boldsymbol{y} \succ \boldsymbol{x}$.

Schur-Concave Functions: A real-valued function f is said to be Schur-concave if and only if

$$\boldsymbol{x} \succ \boldsymbol{y} \Rightarrow f(\boldsymbol{x}) \leq f(\boldsymbol{y}) \tag{78}$$

A multivariate function is said to be symmetric if any two of its arguments can be interchanged without modifying the value of the function. Symmetry is a necessary condition for a function to be Schur-concave. If a function is symmetric and concave, then it is a Schur-concave function. For example, $\min(\boldsymbol{x})$ is a Schur-concave function since it is concave and symmetric.

Lemma 6 [33]: Let A be a Hermitian matrix with ordered diagonal elements denoted by the vector $\boldsymbol{a}$ *and ordered eigenvalues denoted by the vector* $\boldsymbol{\lambda}$*. Then* $\boldsymbol{\lambda} \succ \boldsymbol{a}$.

References

[1] N. M. Blachman, "Communication as a game," in *Proc. IRE WESCON Conf.*, 1957, pp. 61–66.

[2] J. M. Borden, D. M. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM J. Control Optim.*, vol. 23, no. 1, pp. 129–143, 1985.

[3] W. E. Stark and R. J. McEliece, "On the capacity of channels with block memory," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 322–324, Mar. 1988.

[4] K. Yanagi, "Optimal mutual information for coders and jammers in mismatched communication channels," *SIAM J. Control Optim.*, vol. 31, no. 1, pp. 41–51, 1993.

[5] A. Kashyap, T. Başar, and R. Srikant, "Correlated jamming on MIMO Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2119–2123, Sep. 2004.

[6] S. Shafiee and S. Ulukus, "Mutual information games in multiuser channels with correlated jamming," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4598–4607, Oct. 2009.

[7] R. H. Gohary, Y. Huang, Z.-Q. Luo, and J.-S. Pang, "A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2660–2674, Jul. 2009.

[8] M. H. Brady, M. Mohseni, and J. M. Cioffi, "Spatially-correlated jamming in Gaussian multiple access and broadcast channels," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, 2006, pp. 1635–1639.

[9] J. Kiefer and J. Wolfowitz, "Channels with arbitrarily varying channel probability functions," *Inf. Control*, vol. 5, no. 1, pp. 44–54, 1962.

[10] R. Ahlswede, "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 1027–1033, 1970.

[11] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inf. Theory*, vol. 32, no. 5, pp. 621–629, Sep. 1986.

[12] I. Csiszár and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 18–26, Jan. 1991.

[13] B. Hughes and P. Narayan, "The capacity of a vector Gaussian arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 995–1003, Sep. 1988.

[14] E. Akyol, K. Rose, and T. Başar, "On optimal jamming over an additive noise channel," in *Proc. IEEE 52nd Annu. Conf. Decision Control*, Dec. 2013, pp. 3079–3084.

[15] E. Akyol and K. Rose, "Optimal jamming over additive noise: Vector source-channel case," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput.*, 2013, pp. 1329–1335.

[16] T. Başar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. Inf. Theory*, vol. 29, no. 1, pp. 152–157, Jan. 1983.

[17] T. Başar and Y.-W. Wu, "A complete characterization of minimax and maximin encoder- decoder policies for communication channels with incomplete statistical description," *IEEE Trans. Inf. Theory*, vol. 31, no. 4, pp. 482–489, Jul. 1985.

[18] T. Başar and Y. W. Wu, "Solutions to a class of minimax decision problems arising in communication systems," *J. Optim. Theory Appl.*, vol. 51, no. 3, pp. 375–404, 1986.

[19] E. Akyol, K. B. Viswanatha, and K. Rose, "On conditions for linearity of optimal estimation," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3497–3508, Jun. 2012.

[20] E. Akyol, K. B. Viswanatha, K. Rose, and T. A. Ramstad, "On zero-delay source-channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7473–7489, Dec. 2014.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[22] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[23] P. Billingsley, *Probability and Measure*. New York, NY, USA: Wiley, 2008.

[24] E. Lukacs, *Characteristics Functions*. London, U.K.: Griffin, 1960.

[25] F. W. Steutel and K. van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*. Boca Raton, FL, USA: CRC Press, 2003.

[26] H. Cramér, *Mathematical Methods of Statistics*, vol. 9. Princeton, NJ, USA: Princeton Univ. Press, 1999.

[27] F. Hekland, P. A. Floor, and T. A. Ramstad, "Shannon–Kotel'nikov mappings in joint source-channel coding," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, Jan. 2009.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[29] K.-H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.

[30] T. Başar, "A trace minimization problem with applications in joint estimation and control under nonclassical information," *J. Optim. Theory Appl.*, vol. 31, no. 3, pp. 343–359, 1980.

[31] T. Ü. Başar, B. Sankur, and H. Abut, "Performance bounds and optimal linear coding for discrete-time multichannel communication systems (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 212–217, Mar. 1980.

[32] E. Akyol and K. Rose, "On linear transforms in zero-delay Gaussian source channel coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1543–1547.

[33] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Academic, 1979.

[34] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, Jul. 2003.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[36] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, “Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization,” *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[37] E. Akyol, K. Rose, and T. Başar, “Gaussian sensor networks with adversarial nodes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 539–543.

[38] E. Akyol, K. Rose, and T. Başar, “On communication over Gaussian sensor networks with adversaries: Further results,” in *Decision and Game Theory for Security*. New York, NY, USA: Springer-Verlag, 2013, pp. 1–9.

Emrah Akyol (S’03–M’12) received the Ph.D. degree in 2011 from the University of California at Santa Barbara. From 2006 to 2007, he held positions at Hewlett-Packard Laboratories and NTT Docomo Laboratories, both in Palo Alto, where he worked on topics in video compression. From 2013 to 2014, Dr. Akyol was a postdoctoral researcher in the Electrical Engineering Department, University of Southern California. Currently, Dr. Akyol is a postdoctoral research associate in the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. His current research is on the interplay of networked information theory, game theory, communications and control. Dr. Akyol received the 2010 UCSB Dissertation Fellowship and the 2014 USC Postdoctoral Training Award.

Kenneth Rose (S’85–M’91–SM’01–F’03) received the Ph.D. degree in 1991 from the California Institute of Technology, Pasadena.

He then joined the Department of Electrical and Computer Engineering, University of California at Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose was co-recipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.

Tamer Başar (S’71–M’73–SM’79–F’83–LF’13) is with the University of Illinois at Urbana-Champaign, where he holds the academic positions of Swanlund Endowed Chair; Center for Advanced Study Professor of Electrical and Computer Engineering; Research Professor at the Coordinated Science Laboratory; and Research Professor at the Information Trust Institute. He is also the Director of the Center for Advanced Study. He received B.S.E.E. from Robert College, Istanbul, and M.S., M.Phil, and Ph.D. from Yale University. He is a member of the US National Academy of Engineering, member of the European Academy of Sciences, and Fellow of IEEE, IFAC (International Federation of Automatic Control) and SIAM (Society for Industrial and Applied Mathematics), and has served as president of IEEE CSS (Control Systems Society), ISDG (International Society of Dynamic Games), and AACC (American Automatic Control Council). He has received several awards and recognitions over the years, including the highest awards of IEEE CSS, IFAC, AACC, and ISDG, the IEEE Control Systems Award, and a number of international honorary doctorates and professorships. He has over 700 publications in systems, control, communications, and dynamic games, including books on non-cooperative dynamic game theory, robust control, network security, wireless and communication networks, and stochastic networked control. He was the Editor-in-Chief of *Automatica* between 2004 and 2014, and is currently editor of several book series. His current research interests include stochastic teams, games, and networks; security; and cyberphysical systems.