Joint communication and target detection with multiple antennas

Hamdi Joudeh
Information and Communication Theory Lab
Eindhoven University of Technology, The Netherlands
h.joudeh@tue.nl

***Abstract*—We study a basic joint communication and sensing setting comprising a base station, an active user and a potential passive target. The base station wishes to send a message to the user and at the same time detect the presence or absence of the target. Both base station and user are equipped with multiple antennas. We characterize the fundamental performance trade-off between the communication and sensing tasks, captured in terms of the rate of reliable communication against the detection error exponent. In doing so, we identify optimal schemes for this multi-antenna joint communication and sensing setting.**

I. Introduction

We consider a joint communication and sensing (JCAS) setting where a base station (BS) wishes to communicate with an active user, and at the same time detect the presence of a passive target. The BS is equipped with multiple transmit antennas used to send signals for both communication and probing purposes; and multiple receive antennas used to collect backscattered echo signals for sensing, as in multiple-input multiple-output (MIMO) radar [1]. Naturally, there is a performance trade-off between message communication and target detection as a signal optimized one task is often not optimal for the other, e.g. consider a line-of-sight channel where the user and target occupy distinct angular directions. We are interested in characterizing the fundamental performance trade-off between the two tasks, and designing information-theoretically optimal schemes that strike the best trade-off.

Problems related to signal and waveform design for JCAS have received increased interest in recent years, see, e.g., [2], [3]. Most relevant to the present work are [4]–[8], which consider a variety of MIMO JCAS settings with multiple users and targets. These works, however, adopt schemes with predetermined signal structures based on, e.g., linear precoding (or beamforming) and the superposition of radar and communication waveforms. Without an information-theoretic treatment, there are no guarantees that these predetermined signal structures are optimal amongst all feasible signal designs; and it remains unclear whether the performance trade-offs identified in these works are fundamental. In contrast to these works, in the current paper we focus on the elemental MIMO setting with a single active user and a basic sensing task of target detection. Our aim in studying this elemental setting is to gain initial theoretical insights, which will hopefully serve as a first step towards more involved generalizations.

In the setting of interest, we characterize the fundamental performance trade-off in terms of the rate of reliable communication against the detection error exponent. This fundamental rate-exponent trade-off was proposed in our recent work [9], where we studied binary symmetric channels and single-antenna Gaussian channels; and was recently extended to general discrete memoryless channels in [10], [11]. In the current paper, we generalize the result on single-antenna Gaussian channels in [9] to vector Gaussian channels, where both the BS and the user are equipped with multiple antennas.

Other information-theoretic formulations for JSAC have been considered in [12], [13], which are however less relevant to the current work as they focus on sensing a channel state sequences that varies in an i.i.d. fashion across channel uses. In contrast, and similar to [9]–[11], in the current work we are concerned with sensing a variable (or state) that remains fixed over channel uses, i.e. presence of absence of a target.

II. System model and performance measures

In the setting of interest, the BS has K transmit antennas and J receive antennas (i.e. co-located transmitter and sensor). In the n-th channel use (i.e. baseband symbol period), the BS's transmitter sends $\boldsymbol{x}_n \in \mathbb{C}^K$ through the K antennas. The L-antenna user receives the vector $\boldsymbol{Y}_n \in \mathbb{C}^L$ given by

$$\boldsymbol{Y}_n = \mathbf{H}\boldsymbol{x}_n + \boldsymbol{N}_{\boldsymbol{Y},n} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{L \times K}$ is a fixed channel response matrix between the BS and the user. $\boldsymbol{N}_{\boldsymbol{Y},n} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{Y}}^2 \mathbf{I})$ is a zero-mean complex-Gaussian noise vector independent over channel uses. Through the J receive antennas, the BS's sensor obtains a noisy backscattered echo $\boldsymbol{Z}_n \in \mathbb{C}^J$ given by

$$\boldsymbol{Z}_n = \Theta\mathbf{G}\boldsymbol{x}_n + \boldsymbol{N}_{\boldsymbol{Z},n} \tag{2}$$

where $\Theta \in \{0, 1\}$ is a binary variable that is 0 when the target is absent and 1 when the target is present, $\mathbf{G} \in \mathbb{C}^{J \times K}$ is the target channel response matrix, and $\boldsymbol{N}_{\boldsymbol{Z},n} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\boldsymbol{Z}}^2 \mathbf{I})$ is the corresponding noise vector. In what follows, we assume that $\sigma_{\boldsymbol{Y}}^2 = \sigma_{\boldsymbol{Z}}^2 = 1$. This incurs no loss of generality, as different signal-to-noise-ratio (SNR) levels for the user and sensor can be attained by scaling $\mathbf{H}$ and $\mathbf{G}$.

Transmission occurs over a block of N channel uses, where a signal (i.e. a sequence of vector channel symbols) given by $\boldsymbol{x}^N \triangleq \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ is sent, serving the dual purpose of communicating with the user and probing the target. Such signal is subject to an average power constraint

$$\frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{x}_n\|^2 \leq \mathsf{P}. \tag{3}$$

A. Encoding, decoding, and detection

In a block of N channels uses, the BS wishes to send a message W to the user, where W is drawn uniformly at random from $[M_N] \triangleq \{1, 2, \ldots, M_N\}$; and at the same time detect the presence of the target by estimating the binary variable Θ. To this end, an encoded signal (i.e. a codeword) $\boldsymbol{x}^N(W) \in \mathbb{C}^{K\times N}$ is sent. The set of all codewords is known as a codebook, and is given by $\boldsymbol{\mathcal{C}}_N \triangleq \{\boldsymbol{x}^N(1), \boldsymbol{x}^N(2), \ldots, \boldsymbol{x}^N(M_N)\}$. Each codeword must satisfy the power constraint in (3).

The user receives the signal $\boldsymbol{Y}^N \triangleq \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_N$, from which it decodes the message and produces a message estimate $\hat{W} = \varphi_N(\boldsymbol{Y}^N)$, taking values in the message set $[M_N]$. A *decoding error* occurs if $\varphi_N(\boldsymbol{Y}^N) \neq W$, and a maximum *decoding error probability* ϵ_N is achievable if

$$\max_{w\in[M_N]} \mathbb{P}\left[\varphi(\boldsymbol{Y}^N) \neq W \mid W = w\right] \leq \epsilon_N. \tag{4}$$

The maximum over messages in (4) reflects that fact that all messages are equally important, and hence uniform reliability is desired. The sensor receives the signal $\boldsymbol{Z}^N \triangleq \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ and, with knowledge of the codeword $\boldsymbol{x}^N(W)$, estimates the parameter as $\hat{\Theta} = \psi\big(\boldsymbol{Z}^N, \boldsymbol{x}^N(W)\big)$. A *detection error* occurs if $\psi\big(\boldsymbol{Z}^N, \boldsymbol{x}^N(W)\big) \neq \Theta$, and a maximum *detection error probability* δ_N is achievable if

$$\max_{w\in[M_N]}\max_{\theta\in\{0,1\}} \mathbb{P}\left[\psi\big(\boldsymbol{Z}^N, \boldsymbol{x}^N(W)\big) \neq \Theta \,|\, W = w, \Theta = \theta\right] \leq \delta_N. \tag{5}$$

The maximum over messages $w \in [M_N]$ in (5) reflects the natural operational requirement that a sensing performance, in terms of the target detection error probability in this case, must be guaranteed regardless of which message is being communicated to the user. On the other hand, the maximum over the parameter $\theta \in \{0, 1\}$ leads to a minimax detection error formulation, i.e. a worst-case criterion, where both states are equally important. Alternative criteria include Bayesian and Neyman-Pearson detection errors [10].

A scheme as defined above is an $(N, M_N, \epsilon_N, \delta_N)$-scheme if it has a codebook $\boldsymbol{\mathcal{C}}_N$ of M_N length-N codewords, and achieves a decoding error probability of ϵ_N and a detection error probability of δ_N, as specified above.

B. Rate-exponent trade-off

We are interested in the fundamental trade-off between the rate of reliable communication and the detection error probability in the asymptotic regime (i.e. $N \to \infty$). This trade-off is formalized through the achievable communication rate against the detection error exponent.

Definition 1. The rate-exponent tuple (R, E) is achievable if there exists a sequence of $(N, M_N, \epsilon_N, \delta_N)$-schemes with $\lim_{N\to\infty} \epsilon_N = 0$ (vanishing decoding error probability) and

$$\lim_{N\to\infty}\frac{1}{N}\log M_N = R \tag{6}$$

$$\lim_{N\to\infty}\frac{1}{N}\log\frac{1}{\delta_N} = E. \tag{7}$$

The rate-exponent region $\mathcal{R}$ is the closure of the set of all achievable pairs (R, E), that is

$$\mathcal{R} \triangleq \mathrm{cl}\big\{(R, E) : (R, E) \text{ is achievable}\big\}. \tag{8}$$

To illustrate the operational significance of the rate-exponent trade-off, suppose that a pair (R, E) is achievable. The detection error probability is approximately equal to $\delta_N \approx e^{-NE}$ for large N, and hence achieving $\delta_N = \delta^\star$ requires roughly $N^\star \approx \frac{1}{E}\log\frac{1}{\delta^\star}$ channel uses. At the same time, $N^\star R$ bits of information can be communicated to the user with a small decoding error probability, provided that $N^\star$ is large enough.

III. Main result

Here we present the main result of this paper.

Theorem 1. *The rate-exponent region $\mathcal{R}$ is given by all non-negative tuples (R, E) that satisfy*

$$R \leq \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right) \tag{9}$$

$$E \leq \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathsf{H}}\right) \tag{10}$$

for some matrix $\mathbf{Q} \in \mathbb{C}^{K\times K}$ satisfying $\mathbf{Q} \succeq 0$ and $\mathrm{tr}(\mathbf{Q}) \leq \mathsf{P}$.

The proof of Theorem 1 is presented in Section IV. As seen through the achievability part of the proof, the boundary points of the rate-exponent region are achieved through sequences of schemes in which all codewords have approximately the same sample covariance matrix, i.e. we have $\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}} \approx \mathbf{Q}$ for all codewords $\boldsymbol{x}^N \in \boldsymbol{\mathcal{C}}_N$ and all blocklenths N. This restriction guarantees a uniform detection error performance across all codewords (or messages); and at the same time yields a rate that coincides with the one achieved through complex-Gaussian random coding with covariance $\mathbf{Q}$. Next, we will discuss optimal communication-centric and sensing-centric strategies and optimal trade-offs.

The optimal covariance matrix for communication is

$$\mathbf{Q}_{\mathrm{c}}^\star = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathsf{H}} \tag{11}$$

where $\mathbf{V}$ comprises the eigenvectors of $\mathbf{H}^{\mathsf{H}}\mathbf{H}$, and $\mathbf{D}$ is a non-negative diagonal matrix with $\mathrm{tr}(\mathbf{D}) \leq \mathsf{P}$ obtained through water-filling over the corresponding eigenvalues.[1] The communication-optimal rate-exponent pair is denoted by $(R_{\mathrm{c}}^\star, E_{\mathrm{c}}^\star)$. On the other hand, the optimal covariance matrix for sensing is given by

$$\mathbf{Q}_{\mathrm{s}}^\star = \mathsf{P}\mathbf{u}_1\mathbf{u}_1^{\mathsf{H}} \tag{12}$$

where $\mathbf{u}_1$ is the eigenvector associated with the largest eigenvalue of $\mathbf{G}^{\mathsf{H}}\mathbf{G}$. The corresponding sensing-optimal rate-exponent pair is denoted by $(R_{\mathrm{s}}^\star, E_{\mathrm{s}}^\star)$.

Both $(R_{\mathrm{c}}^\star, E_{\mathrm{c}}^\star)$ and $(R_{\mathrm{s}}^\star, E_{\mathrm{s}}^\star)$ are corner points of $\mathcal{R}$ (illustrated in an example further on). All points in $\mathcal{R}$ that satisfy $(R, E) \leq (R_{\mathrm{c}}^\star, E_{\mathrm{c}}^\star)$ or $(R, E) \leq (R_{\mathrm{s}}^\star, E_{\mathrm{s}}^\star)$, where

[1] This is equivalent to the famous singular value decomposition (SVD) with water-filling strategy [14].

the inequalities are coordinate-wise, are sub-optimal in the Pareto sense. Points on the boundary connecting $(R_{\rm c}^\star, E_{\rm c}^\star)$ and $(R_{\rm s}^\star, E_{\rm s}^\star)$ are Pareto optimal, and each such point is achieved by a $\mathbf{Q}$ that maximizes $\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right)$ subject to $\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathsf{H}}\right) \geq E$ for some $E \in (E_{\rm c}^\star, E_{\rm s}^\star)$.

Example. Consider a BS with co-located transmit and receive uniform linear arrays, a single-antenna user and a possible point target. For an azimuth angle of $\phi \in [0, \pi]$, the steering vectors of the BS's transmit and receive arrays are denoted by $\mathbf{a}(\phi) \in \mathbb{C}^K$ and $\mathbf{b}(\phi) \in \mathbb{C}^J$, respectively. The user has a line-of-sight channel $\mathbf{h} \in \mathbb{C}^{1\times K}$ given by $\alpha\mathbf{a}(\phi_Y)^{\mathsf{H}}$, where $\alpha \in \mathbb{C}$ is a scalar channel coefficient and ϕ_Y is the user's angle. On the other hand, the target's channel $\mathbf{G} \in \mathbb{C}^{J\times K}$ is given by $\beta\mathbf{b}(\phi_Z)\mathbf{a}(\phi_Z)^{\mathsf{H}}$, where $\beta \in \mathbb{C}$ is the corresponding scalar channel coefficient and ϕ_Z is the target's angle.

The rate-exponent region described in Theorem 1 is illustrated for two instances of the above setting in Fig. 1. The angular directions occupied by the user and target are closer in the first instance (solid blue line), resulting in higher synergies between the communication and detection tasks. The more orthogonal $\mathbf{a}(\phi_Y)$ and $\mathbf{a}(\phi_Z)$, the weaker the synergies. In the extreme case of orthogonal $\mathbf{a}(\phi_Y)$ and $\mathbf{a}(\phi_Z)$, the communication-centric and sensing-centric corner points are given by $(R_{\rm c}^\star, 0)$ and $(0, E_{\rm s}^\star)$, respectively.

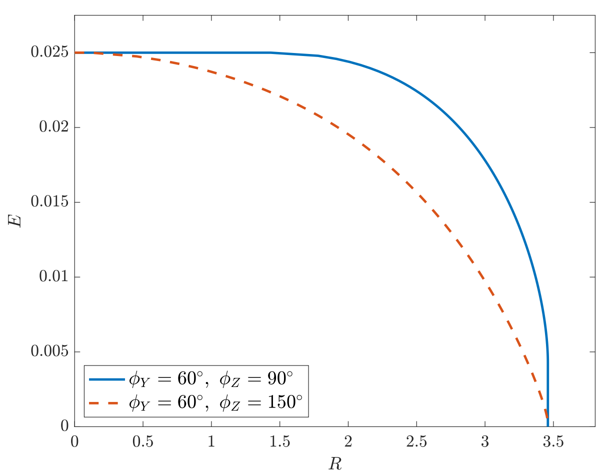

Fig. 1. Rate-exponent regions for two instances of the above-described setting. In both instances, the BS has $K = 10$ transmit and $J = 10$ receive antennas; and the user and target SNRs are equal to 10 dB and -10 dB respectively.

IV. Proof of Main Result

This section is dedicated to proving Theorem 1. We start by presenting essential preliminaries on the target detection problem, which is an active form of binary hypothesis testing.

Let $\boldsymbol{Z} \in \mathbb{C}^K$ be a random vector and let $f_{\boldsymbol{Z}}^0(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}^1(\boldsymbol{z})$ be two possible probability densities for $\boldsymbol{Z}$. The Bhattacharyya distance [15] between the two densities is defined as

$$D_{\rm B}\left(f_{\boldsymbol{Z}}^0, f_{\boldsymbol{Z}}^1\right) \triangleq -\log\int_{\mathbb{C}^K}\sqrt{f_{\boldsymbol{Z}}^0(\boldsymbol{z})f_{\boldsymbol{Z}}^1(\boldsymbol{z})}\mathrm{d}\boldsymbol{z}. \tag{13}$$

A particular case of interest to us here is when $f_{\boldsymbol{Z}}^0$ and $f_{\boldsymbol{Z}}^0$ are the densities of the complex-Gaussian distributions $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{I})$, respectively. In this case, it can be verified that

$$D_{\rm B}\left(f_{\boldsymbol{Z}}^0, f_{\boldsymbol{Z}}^1\right) = \frac{1}{4}\|\boldsymbol{\mu}\|^2. \tag{14}$$

Let us now return to our target detection problem. Given that $\boldsymbol{x}^N$ has been sent, the signal $\boldsymbol{Z}^N$ received by the sensor is complex-Gaussian with independent entries, where the n-th entry $\boldsymbol{Z}_n$ is drawn from $\mathcal{CN}\big(0, \mathbf{I}\big)$ under $\Theta = 0$ and $\mathcal{CN}\big(\mathbf{G}\boldsymbol{x}_n, \mathbf{I}\big)$ under $\Theta = 1$. Let $f_{\boldsymbol{Z}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}^0(\boldsymbol{z}^N)$ and $f_{\boldsymbol{Z}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}^1(\boldsymbol{z}^N)$ be the corresponding densities under $\Theta = 0$ and $\Theta = 1$ respectively. In this case we have

$$\begin{aligned}
&D_{\rm B}\left(f_{\boldsymbol{Z}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}^0, f_{\boldsymbol{Z}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}^1\right)\\
&\quad= \sum_{n=1}^{N} D_{\rm B}\left(f_{\boldsymbol{Z}_n|\boldsymbol{X}_n=\boldsymbol{x}_n}^0, f_{\boldsymbol{Z}_n|\boldsymbol{X}_n=\boldsymbol{x}_n}^1\right) && (15)\\
&\quad= \frac{1}{4}\sum_{n=1}^{N}\|\mathbf{G}\boldsymbol{x}_n\|^2 && (16)\\
&\quad= \frac{N}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) && (17)
\end{aligned}$$

where $\mathbf{Q}(\boldsymbol{x}^N) \triangleq \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}}$ is the sample (empirical) covariance matrix of the codeword $\boldsymbol{x}^N$. The Bhattacharyya distance plays a central role in characterizing the detection error probability as seen next. In particular, define the minimax detection error probability as

$$\delta^\star(\boldsymbol{x}^N) \triangleq \min_{\psi}\max_{\theta\in\{0,1\}}\mathbb{P}\left[\psi\big(\boldsymbol{Z}^N, \boldsymbol{x}^N\big) \neq \theta \mid \Theta = \theta\right]. \tag{18}$$

which is the best (i.e. smallest) possible detection error probability when $\boldsymbol{x}^N$ is used.

Lemma 1. *Given that $\boldsymbol{x}^N$ is sent, the minimax detection error is bounded above and below as:*

$$e^{-\left(\frac{N}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right)+o(N)\right)} \leq \delta^\star(\boldsymbol{x}^N) \leq e^{-\frac{N}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right)}.$$

The above result is given without proof as it follows directly from classical results in the literature, see [16], [17, Ch. 2.3]. From the bounds in Lemma 1, it follows that under optimal target detection, an $(N, M_N, \epsilon_N, \delta_N)$-scheme with an associated codebook $\boldsymbol{\mathcal{C}}_N$ will have

$$\begin{aligned}
\min_{\boldsymbol{x}^N\in\boldsymbol{\mathcal{C}}_N}\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) &\leq \frac{1}{N}\log\frac{1}{\delta_N}\\
&\leq \min_{\boldsymbol{x}^N\in\boldsymbol{\mathcal{C}}_N}\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) + \frac{o(N)}{N}. && (19)
\end{aligned}$$

A. Achievability of Theorem 1

To show that the region in Theorem 1 is achievable, it is sufficient to show that for any covariance matrix $\mathbf{Q} \succeq 0$ that satisfies the power constraint $\mathrm{tr}(\mathbf{Q}) \leq \mathsf{P}$, there exists a sequence of $(N, M_N, \epsilon_N, \delta_N)$-schemes with

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\log M_N \geq \log\det\left(\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right)\\
&\quad\text{and}\ \lim_{N\to\infty}\frac{1}{N}\log\frac{1}{\delta_N} \geq \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathsf{H}}\right). && (20)
\end{aligned}$$

1) Exponent: For a given blocklength N and codebook $\mathcal{C}_N$, the detection error performances is dictated by the *worst* codeword, as clearly seen through the bounds in (20). To guarantee a somewhat uniform detection error performance regardless of which codeword is sent, we impose the constraint that codewords of $\mathcal{C}_N$ must be drawn from the set $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$, defined for some $\varepsilon > 0$ as

$$\mathcal{B}_{N,\varepsilon}(\mathbf{Q}) \triangleq \left\{ \boldsymbol{x}^N \in \mathbb{C}^N : \mathbf{Q} - \varepsilon\mathbf{I} \preceq \mathbf{Q}(\boldsymbol{x}^N) \preceq \mathbf{Q} \right\}. \quad (21)$$

All sequences in $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$ have an empirical covariance matrix $\mathbf{Q}(\boldsymbol{x}^N)$ which is sufficiently close to $\mathbf{Q}$ (in the semidefinite sense). with the assumption that $\mathcal{C}_N \subset \mathcal{B}_{N,\varepsilon}(\mathbf{Q})$, from the lower bound in (19) we get

$$\begin{aligned} \frac{1}{N}\log\frac{1}{\delta_N} &\geq \min_{\boldsymbol{x}^N\in\mathcal{C}_N} \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) && (22)\\ &\geq \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\big(\mathbf{Q}-\varepsilon\mathbf{I}\big)\mathbf{G}^{\mathsf{H}}\right) && (23)\\ &= \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathsf{H}}\right) - \varepsilon' && (24)\end{aligned}$$

where $\varepsilon' = \frac{\varepsilon}{4}\mathrm{tr}(\mathbf{G}\mathbf{G}^{\mathsf{H}})$. Taking $N \to \infty$ and $\varepsilon \to 0$, the exponent in (20) is achieved.

2) Rate: The most common way to prove achievability of the rate in (20) is by using a random coding argument, where each vector channel symbol $\boldsymbol{x}_n$ in every codeword $\boldsymbol{x}^N \in \mathcal{C}_N$ is independently drawn from a complex-Gaussian distribution $\mathcal{CN}(\mathbf{0},\mathbf{Q})$, see, e.g., [18, Ch.9] and [14]. This construction, however, does not guarantee that all codewords are in $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$, and will occasionally produce codewords that are not favourable for the detection task. Instead, we impose the constraint that codewords are drawn from $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$ for all $N \in \mathbb{N}$, and we show that the desired rate in (20) is still achievable with this constraint in place. To this end, we base our achievability proof on Feinstein's lemma [19] (see also [20, Theorem 1]), which makes it easier to deal with the constrain $\mathcal{C}_N \subset \mathcal{B}_{N,\varepsilon}(\mathbf{Q})$ as seen next.

let $\boldsymbol{X}^N = \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ be a random channel input sequence with a probability distribution supported on $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$. Given $\boldsymbol{X}^N = \boldsymbol{x}^N$ is used as an input to the channel, the distribution of the corresponding output sequence $\boldsymbol{Y}^N$ has a Gaussian density denoted by $f_{\boldsymbol{Y}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}(\boldsymbol{y}^N)$. Let $f_{\boldsymbol{Y}^N}(\boldsymbol{y}^N)$ be the output distribution induced by $\boldsymbol{X}^N$. The information density [20] is defined as

$$\imath(\boldsymbol{x}^N;\boldsymbol{y}^N) \triangleq \log\frac{f_{\boldsymbol{Y}^N|\boldsymbol{X}^N=\boldsymbol{x}^N}(\boldsymbol{y}^N)}{f_{\boldsymbol{Y}^N}(\boldsymbol{y}^N)} \quad (25)$$

which can be seen as an *instantaneous* mutual information. Note that $\mathbb{E}\left[\imath(\boldsymbol{X}^N;\boldsymbol{Y}^N)\right] = I(\boldsymbol{X}^N;\boldsymbol{Y}^N)$.

Lemma 2. ***(Feinstein [19])*** *For any $N \in \mathbb{N}$ and $\gamma > 0$, there exists a code with codebook $\mathcal{C}_N \subset \mathcal{B}_{N,\varepsilon}(\mathbf{Q})$ of M_N codewords and a maximum decoding error probability ϵ_N satisfying*

$$\epsilon_N \leq \mathbb{P}\left[\imath(\boldsymbol{X}^N;\boldsymbol{Y}^N) \leq N\gamma\right] + M_N\exp(-N\gamma). \quad (26)$$

Now consider a complex-Gaussian vector $\boldsymbol{X}_g \sim \mathcal{CN}(\mathbf{0},\mathbf{Q})$, and let $\boldsymbol{X}_g^N = \boldsymbol{X}_{g,1}, \boldsymbol{X}_{g,2}, \ldots, \boldsymbol{X}_{g,N}$ be a sequence of N i.i.d. copies of $\boldsymbol{X}_g$. We take $\boldsymbol{X}^N$ in Lemma 2 to be distributed as $\boldsymbol{X}_g^N$ conditioned on the event $\boldsymbol{X}_g^N \in \mathcal{B}_{N,\varepsilon}(\mathbf{Q})$. In this case, the distribution of $\boldsymbol{X}^N$ has a density $f_{\boldsymbol{X}^N}(\boldsymbol{x}^N)$ defined as

$$f_{\boldsymbol{X}^N}(\boldsymbol{x}^N) = \frac{g(\boldsymbol{x}^N)}{g\left(\mathcal{B}_{N,\varepsilon}(\mathbf{Q})\right)}\mathbb{1}\left[\boldsymbol{x}^N \in \mathcal{B}_{N,\varepsilon}(\mathbf{Q})\right] \quad (27)$$

where $g(\boldsymbol{x}^N) = \prod_{n=1}^N g_{\boldsymbol{X}_g}(\boldsymbol{x}_n)$ is the Gaussian probability density of $\boldsymbol{X}_g^N$, and

$$g\left(\mathcal{B}_{N,\varepsilon}(\mathbf{Q})\right) \triangleq \int_{\mathcal{B}_{N,\varepsilon}(\mathbf{Q})} g(\boldsymbol{x}^N)\mathrm{d}\boldsymbol{x}^N. \quad (28)$$

The input distribution in (27) is supported on $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$, which guarantees $\mathcal{C}_N \subset \mathcal{B}_{N,\varepsilon}(\mathbf{Q})$. Moreover, with such an input distribution, the term $\mathbb{P}\left[\imath(\boldsymbol{X}^N;\boldsymbol{Y}^N) \leq N\gamma\right]$ in Lemma 2 is bounded as

$$\begin{aligned} &\mathbb{P}\left[\imath(\boldsymbol{X}^N;\boldsymbol{Y}^N) \leq N\gamma\right] \\ &\quad= \mathbb{P}\left[\imath(\boldsymbol{X}_g^N;\boldsymbol{Y}_g^N) \leq N\gamma \mid \boldsymbol{X}_g^N \in \mathcal{B}_{N,\varepsilon}(\mathbf{Q})\right] && (29)\\ &\quad\leq \frac{1}{\mathbb{P}\left[\boldsymbol{X}_g^N \in \mathcal{B}_{N,\varepsilon}(\mathbf{Q})\right]}\mathbb{P}\left[\imath(\boldsymbol{X}_g^N;\boldsymbol{Y}_g^N) \leq N\gamma\right] && (30)\\ &\quad= \frac{1}{g\big(\mathcal{B}_{N,\varepsilon}(\mathbf{Q})\big)}\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^N \imath(\boldsymbol{X}_{g,n};\boldsymbol{Y}_{g,n}) \leq \gamma\right] && (31)\end{aligned}$$

where $\boldsymbol{Y}_g^N = \boldsymbol{Y}_{g,1}, \boldsymbol{Y}_{g,2}, \ldots, \boldsymbol{Y}_{g,N}$ is the output sequence induced by the input sequence $\boldsymbol{X}_g^N$.

Since $\boldsymbol{X}_g^N$ and $\boldsymbol{Y}_g^N$ are both i.i.d. and

$$\mathbb{E}\left[\imath(\boldsymbol{X}_{g,n};\boldsymbol{Y}_{g,n})\right] = I(\boldsymbol{X}_g;\boldsymbol{Y}_g) = \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right),$$

by setting $\gamma = \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right) - \varepsilon/2$ we get

$$\lim_{N\to\infty}\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^N \imath(\boldsymbol{X}_{g,n};\boldsymbol{Y}_{g,n}) \leq \gamma\right] = 0$$

by the the weak law of large numbers (WLLN). Base on the WLLN, we can also show that

$$\lim_{N\to\infty} g\big(\mathcal{B}_{N,\varepsilon}(\mathbf{Q})\big) = 0.$$

Combining these observations with (31), we obtain

$$\lim_{N\to\infty}\mathbb{P}\left[\frac{1}{N}\imath(\boldsymbol{X}^N;\boldsymbol{Y}^N) \leq \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right) - \frac{\varepsilon}{2}\right] = 0. \quad (32)$$

From (32) and the result in Lemma 2, it follows that we guarantee a vanishing decoding error probability, i.e. $\epsilon_N \to 0$ as $N \to \infty$, by further setting

$$\frac{1}{N}\log M_N = \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right) - \varepsilon. \quad (33)$$

Taking $N \to \infty$ and $\varepsilon \to 0$ in (33), we achieve the desired rate specified in (20).

Remark 1. In the above proof, we constructed a multi-letter input distribution $f_{\boldsymbol{X}^N}(\boldsymbol{x}^N)$ supported on the desired set of sequences $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$ from an i.i.d. input distribution $g(\boldsymbol{x}^N) = \prod_{n=1}^N g_{\boldsymbol{X}_g}(\boldsymbol{x}_n)$ by conditioning. We then showed that the rate achieved by using $f_{\boldsymbol{X}^N}(\boldsymbol{x}^N)$ coincides with the rate achieved by the i.i.d. input distribution $g(\boldsymbol{x}^N)$, as the support of the former (i.e. $\mathcal{B}_{N,\varepsilon}(\mathbf{Q})$) is a high-probability set for the latter. This construction method has been used in the literature to prove coding theorems in memoryless channels under cost constraints, see, e.g., [21, Ch. 7.3].

Remark 2. The codes emerging from the above achievability proof, in which all codewords have almost the same empirical covariance matrix, can be seen as a vector generalization of Shannon's spherical codes for the scalar Gaussian channel, in which all codewords have the same empirical power [22].

B. Proof of converse

Consider an arbitrary sequence of $(N, M_N, \epsilon_N, \delta_N)$-schemes achieving the rate-exponent tuple (R, E). To prove the converse part of Theorem 1, we show that we must have

$$R \leq \log\det\left(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right) \text{ and } E \leq \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathsf{H}}\right). \quad (34)$$

for some $\mathbf{Q} \succeq 0$ satisfying $\mathrm{tr}(\mathbf{Q}) \leq \mathsf{P}$. To this end, let $\{\mathcal{C}_N\}_{N\in\mathbb{N}}$ be the associated sequence of codebooks. Recall that $\mathbf{Q}(\boldsymbol{x}^N)$ is the sample covariance matrix of codeword $\boldsymbol{x}^N$, and define

$$\bar{\mathbf{Q}}(\mathcal{C}_N) \triangleq \frac{1}{M_N}\sum_{\boldsymbol{x}^N\in\mathcal{C}_N}\mathbf{Q}(\boldsymbol{x}^N) = \frac{1}{M_N}\sum_{\boldsymbol{x}^N\in\mathcal{C}_N}\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}}$$

which is the sample covariance matrix averaged over all codewords in $\mathcal{C}_N$. We make the technical assumption that for any considered sequence of codebooks, the corresponding sequence $\left\{\bar{\mathbf{Q}}(\mathcal{C}_N)\right\}_{N\in\mathbb{M}}$ has a well-defined limit,[2] i.e. $\bar{\mathbf{Q}}(\mathcal{C}_N) \to \bar{\mathbf{Q}}$ as $N \to \infty$.

Since $\mathrm{tr}\big(\mathbf{Q}(\boldsymbol{x}^N)\big) \leq \mathsf{P}$ for all $\boldsymbol{x}^N \in \mathcal{C}_N$ and $N \in \mathbb{N}$, then we must also have $\mathrm{tr}\big(\bar{\mathbf{Q}}(\mathcal{C}_N)\big) \leq \mathsf{P}$ for all $N \in \mathbb{N}$. The latter implies that $\mathrm{tr}\big(\bar{\mathbf{Q}}\big) \leq \mathsf{P}$, due to the compactness of the set of trace-constrained positive semidefinite matrices.

1) Exponent: Starting from the upper bound in (19) and taking the limit $N \to \infty$, we obtain

$$\begin{aligned} E &\leq \lim_{N\to\infty}\min_{\boldsymbol{x}^N\in\mathcal{C}_N}\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) && (35)\\ &\leq \lim_{N\to\infty}\frac{1}{M_N}\sum_{\boldsymbol{x}^N\in\mathcal{C}_N}\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\mathbf{Q}(\boldsymbol{x}^N)\mathbf{G}^{\mathsf{H}}\right) && (36)\\ &= \lim_{N\to\infty}\frac{1}{4}\mathrm{tr}\left(\mathbf{G}\bar{\mathbf{Q}}(\mathcal{C}_N)\mathbf{G}^{\mathsf{H}}\right) && (37)\\ &= \frac{1}{4}\mathrm{tr}\left(\mathbf{G}\bar{\mathbf{Q}}\mathbf{G}^{\mathsf{H}}\right) && (38)\end{aligned}$$

2) Rate: To find an upper bound on R, we start from Fano's inequality. Note that $\boldsymbol{x}^N(W)$ is a random sequence uniformly distributed on the codebook $\mathcal{C}_N$. Therefore, we have

$$\begin{aligned} &\log M_N - N\bar{\epsilon}_N\\ &\leq I(\boldsymbol{x}^N(W);\boldsymbol{Y}^N) && (39)\\ &\leq \sum_{n=1}^{N} I(\boldsymbol{x}_n(W);\boldsymbol{Y}_n) && (40)\\ &\leq \sum_{n=1}^{N}\log\det\left(\mathbf{I} + \mathbf{H}\mathbb{E}\left[\boldsymbol{x}_n(W)\boldsymbol{x}_n(W)^{\mathsf{H}}\right]\mathbf{H}^{\mathsf{H}}\right) && (41)\\ &\leq N\log\det\left(\mathbf{I} + \frac{1}{N}\sum_{n=1}^{N}\mathbf{H}\mathbb{E}\left[\boldsymbol{x}_n(W)\boldsymbol{x}_n(W)^{\mathsf{H}}\right]\mathbf{H}^{\mathsf{H}}\right) && (42)\\ &= N\log\det\left(\mathbf{I} + \mathbf{H}\bar{\mathbf{Q}}\left(\mathcal{C}_N\right)\mathbf{H}^{\mathsf{H}}\right) && (43)\end{aligned}$$

where $\bar{\epsilon}_N \to 0$ as $N \to \infty$; (40) holds since the channel is memoryless; (41) is due to the extremal property of Gaussian inputs under a covariance constraint; and (42) is due to concavity and Jensen's inequality. It follows that

$$\begin{aligned} R &\leq \lim_{N\to\infty}\log\det\left(\mathbf{I} + \mathbf{H}\bar{\mathbf{Q}}\left(\mathcal{C}_N\right)\mathbf{H}^{\mathsf{H}}\right) && (44)\\ &= \log\det\left(\mathbf{I} + \mathbf{H}\bar{\mathbf{Q}}\mathbf{H}^{\mathsf{H}}\right). && (45)\end{aligned}$$

The completes the converse proof.

[2]This assumption is not essential, but helps simplify the steps and avoid the cumbersome use of $\lim\inf$ and $\lim\sup$.

Acknowledgement

The author gratefully acknowledges insightful discussions with Frans M. J. Willems and Han Wu.

References

[1] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[2] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1295–1315, 2021.

[3] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, 2022.

[4] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, 2018.

[5] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, 2018.

[6] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, 2020.

[7] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2021.

[8] L. Chen, F. Liu, W. Wang, and C. Masouros, "Joint radar-communication transmission: A generalized Pareto optimization framework," *IEEE Trans. Signal Process.*, vol. 69, pp. 2752–2765, 2021.

[9] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 1, pp. 113–124, 2022.

[10] H. Wu and H. Joudeh, "Joint communication and channel discrimination," *arXiv:2208.07450*, 2022.

[11] M.-C. Chang, S.-Y. Wang, T. Erdoğan, and M. R. Bloch, "Rate and detection-error exponent tradeoff for joint communication and sensing of fixed channel states," *arXiv:2210.07963*, 2022.

[12] M. Kobayashi, G. Caire, and G. Kramer, "Joint state sensing and communication: Optimal tradeoff for a memoryless case," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018, pp. 111–115.

[13] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Trans. Inf. Theory*, pp. 1–1, 2022.

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

[15] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 1967.

[16] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[17] A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*. New York: McGraw-Hill, Inc., 1979.

[18] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York: John Wiley & Sons, Inc., 2006.

[19] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[20] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[21] R. G. Gallager, *Information theory and reliable communication*. New York: John Wiley & Sons, Inc., 1968.

[22] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.