OPTIMAL SIGNALING SCHEMES AND CAPACITIES OF NON-COHERENT

CORRELATED MISO CHANNELS UNDER PER-ANTENNA POWER

CONSTRAINTS

A Thesis

Presented to

The Graduate Faculty of The University of Akron

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Minh Nhat Vu

July, 2018

OPTIMAL SIGNALING SCHEMES AND CAPACITIES OF NON-COHERENT

CORRELATED MISO CHANNELS UNDER PER-ANTENNA POWER

CONSTRAINTS

Minh Nhat Vu

Thesis

Approved: Accepted:

_____   _____
Advisor   Interim Chair
Dr. Nghi H. Tran   Dr. Robert J. Veillette

_____   _____
Committee Member   Dean of the College
Dr. Hamid Bahrami   Dr. Donald P. Visco Jr.

_____   _____
Committee Member   Dean of the Graduate School
Dr. Shivakumar Sastry   Dr. Chand Midha

                                         _____
                                         Date

ABSTRACT

This thesis studies the optimal inputs and capacities of non-coherent correlated multiple-input single-output (MISO) channels in fast Rayleigh fading. We consider two scenarios: channels under per-antenna power constraints and channels under joint per-antenna and sum power constraints. For per-antenna power constraints, we establish the convexity and compactness of the feasible sets, and demonstrate the existence of optimal input distribution. By exploiting the solutions of a quadratic optimization problem, we show that the Kuhn-Tucker condition (KTC) on the optimal inputs can be simplified to a single dimension and prove the discreteness and finiteness of the optimal effective magnitude distribution. Then, we are able to construct a finite and discrete optimal input vector and determine the capacity gain of MISO over SISO. We also extend the results to MISO channels subject to the joint per-antenna and sum power constraints. For this case, the optimal phases and the optimal power allocation among the transmit antennas need to be determined simultaneously via a quadratic optimization subject to inequality constraints. Based on our results, the capacity of considered channels can be obtained and exploited as an upper bound for the operational transmission rate. Further researches can also rely on our analysis of the optimal inputs to construct reliable coding schemes for MISO fading channels.

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AWGN   Additive White Gaussian Noise

CDF   Cumulative Distribution Function

CSI   Channel State Information

KTC   Kuhn-Tucker Condition

LOS   Line of Sight

MISO   Multiple Input Single Output

MIMO   Multiple Input Multiple Output

PDF   Probability Distribution Function

SISO   Single Input Single Output

SNR   Signal to Noise Ratio

CHAPTER I

INTRODUCTION

1.1 Motivation and Literature Review

The use of multiple antennas has been widely acknowledged as an effective solution to enhance the robustness and performance of current and future wireless networks. In multi-antenna wireless systems, while the knowledge of channel state information (CSI) plays an essential role for performance gains, acquiring accurate CSI at the receiver and/or the transmitter is a challenging task. This particularly holds true in fast fading environments or in massive antenna systems [1, 2]. In fact, the estimation of CSI can also be confronted by pilot-contamination, which may degrade the system performance significantly [3]. As such, non-coherent antenna systems where neither the transmitter nor receiver has the CSI have attracted significant attention in the literature [4–9].

Over the last two decades, considerable efforts have been dedicated to information theoretic studies of non-coherent multi-antenna systems under the sum power constraint, leading to several interesting characterizations. For example, in uncorrelated fading, it is well-known that unitary space-time codes are capacity-achieving [4, 5]. Perera *et al.* in [10] further showed that for uncorrelated multiple-input

1

multiple-output (MIMO) Rayleigh fading channels under, the capacity-achieving distribution of the input vector's magnitude is discrete and finite, with a mass points located at the origin. Such results can be considered as an extension of the well-established results obtained earlier in [11, 12] for non-coherent single-antenna channels. However, it is worth mentioning that for non-coherent uncorrelated fading, capacity does not increase when the number of transmit antenna increases beyond the coherence interval [4, 5]. It means that in fast fading environments, there is no benefit of using multiple antennas at the transmitter.

In contrast to the above information-theoretic drawback in uncorrelated fading, Jafar and Goldsmith in [13] demonstrated optimistic results regarding the capacity advantage of using multiple transmit antennas in non-coherent correlated fading channels. In particular, while the unitary structure of the optimal inputs still remains, it was shown in [13] that the capacity is Schur-convex in the eigenvalues of the correlation matrix. As such, spatial correlation enhances capacity. However, without a detailed characterization of the capacity-achieving signal, the capacity gain of multiple antenna systems over a single-antenna system can only be determined precisely in [13] for the special case of fully correlated channels. Sommerfeld in [14] attempted to generalize the results in [10] to more realistic correlated multi-antenna channels. Since the Identity Theorem is not valid for holomorphic functions of multiple complex variables [14, 15], only the boundedness of the optimal magnitude distribution was shown in [14] for correlated channels. Thus, under the sum power constraint, the question whether the optimal input amplitude for a general non-coherent correlated

multi-antenna channel is discrete or not still remains unanswered.

In addition to the sum power constraint that reflects the total power budget at the transmitter, a per-antenna constraint represents the power limitation on each individual RF chain of each antenna [16]. Such a constraint is also suitable for distributed multiple inputs system, in which the transmitters are placed at different locations, and the power cannot be shared among antennas [17]. The per-antenna constraint is, therefore, more realistic in many practical systems [18–21]. While information-theoretic aspects of multi-antenna systems under per-antenna constraints are less well understood due to the complexity of related optimization problems, several interesting results have been obtained recently, but only for coherent channels [22–24]. Under this line of work, since Gaussian signaling schemes are still optimal, the main focus is on the power allocation to maximize the capacity. For example, the capacity of multiple-input single-output (MISO) coherent Rayleigh fading channel under per-antenna constraints was established in closed-form in [22]. The consideration of joint per-antenna and sum power constraints has recently been considered in [18, 25] for coherent multi-antenna channels. However, for non-coherent multi-antenna channels, the problem of characterizing the optimal inputs and determining the capacity under per-antenna power constraints is challenging, even in uncorrelated fading. It is because the use of per-antenna power constraints constitutes different feasible sets and makes it more difficult to examine the conditions for which an input is optimal.

1.2 Contributions

In this thesis, we provide a comprehensive analysis of the optimal signaling schemes and capacities of general non-coherent correlated MISO channels in fast Rayleigh fading. We consider both scenarios where channels are subjected to per-antenna power constraints, and joint per-antenna and sum power constraints. The latter includes channels under the sum power constraint as a special case. We prove the discrete nature of the optimal effective input amplitude $Z = \sqrt{\boldsymbol{X}^{\mathrm{H}}\Sigma\boldsymbol{X}}$, where $\boldsymbol{X}$ is the transmit vector and Σ is the covariance matrix of the channel. We also show the existence of a finite and discrete capacity-achieving input vector for practical implementation. The optimal amplitude and phase distributions of this input allow us to calculate and compare MISO capacity directly from SISO capacity. Our contributions in this research can be summarized as follows.

- In the first part of the thesis, the focus is on MISO channels under per-antenna power constraints. We first establish the convex and compact properties of the feasible sets, and demonstrate the existence of optimal input distribution and the uniqueness of optimal effective magnitude input distribution. Then by formulating the Kuhn-Tucker condition (KTC) on the optimal input and examining it via a quadratic optimization problem, we show that the KTC can be simplified to one dimension. As a consequence, the Identity Theorem [26] can be applied. This result allows us to prove the discrete and finite nature of the optimal effective magnitude distribution, with a mass point located at

4

the origin. We then exploit this optimal distribution to construct a finite and discrete optimal input vector. More interestingly, we show that the phases of the optimal transmitted signals are the phase solutions of a constrained non-convex quadractic optimization problem on a sphere, which can be calculated effectively via a penalized optimization algorithm. The use of this optimal signaling scheme helps to determine precisely the capacity gain of MISO over SISO channels.

- In the second part of this work, we extend the results to MISO channels under the joint per-antenna and sum power constraints. The findings therefore include the MISO channel under the sum power constraint only as a special case, which has not been considered before. Different from channels under per-antenna constraints, we show that at least of one of the per-antenna power constraints must be inactive. As such, the KTC coefficients need to be examined more carefully. While the finiteness and discreteness of the optimal effective magnitude distribution and the optimal input vector distribution still hold, the optimal phase solutions and the optimal power allocation among the transmit antennas are determined simultaneously via a quadratic optimization problem under inequality constraints. These solutions can be used to find the MISO capacity gain.

It should be noted that several novel methods have been proposed recently to overcome the problem of having multiple complex variables in the KTC in multi-antenna systems. For example, the author in [15] studies the optimal inputs and

capacity of multi-antenna channels under peak power constraints via lower and upper bounds on the capacity. The work in [27] addresses the discreteness of the optimal input and capacity of multi-antenna channels under both peak and power constraints. However, the channel models considered in [15, 27] are deterministic, where the channel gains are assumed to be constant, and they are known at both the receiver and transmitter. On the other hand, we consider a general non-coherent correlated fading channel. The key aspects that make our contributions stand out from the previously used approaches and ideas, especially those in [11, 15, 27], can be summarized as follows:

- While we follow a similar methodology in [11] to prove the existence of optimal input distribution and the uniqueness of optimal effective magnitude input distribution, the presence of multi-dimensional input vectors in non-coherent MISO channels requires new derivations.

- Under the assumption of deterministic channels in [15], it is straightforward to convert the problem of finding the optimal inputs to one-dimensional space for the case of MISO. The discreteness of the optimal input then comes directly from the SISO channel because of the imposed peak constraint in [15]. By also considering deterministic channels, [27] exploits the independence of optimal phases and amplitude in the spherical domain to simplify the problem of maximizing the channel mutual information to one dimension under both peak and power constraints. Then the peak constraint plays an essential role in proving

6

showing the discreetness and finiteness of the optimal amplitude in [27]. For our considered non-coherent channels, the optimal magnitudes and phases are not independent, and the dimension reduction techniques used in [15, 27] cannot be applied. In fact, as we demonstrate, the optimal phases depend strictly on the signal's magnitude of each signal vector component. Our approach is to exploit the solutions of a quadratic optimization problem to simplify the KTC before showing the finiteness and discreteness of the optimal effective magnitude distribution without the need of peak power constraints. Note that the proof of finiteness and discreteness is not the same with [11]. It is because in our case, we do not have the strictly inequality in the KTC for non-optimal mass points.

- Reference [15] further evaluates the capacity of a deterministic multi-antenna channel under peak constraints via lower and upper bounds. The main idea is based on the transformation of the considered coupled multi-antenna channel into decoupled channels with coupled inputs for which the capacities are known. Unfortunately, this technique can only be used for deterministic channels. For our considered non-coherent fading channels, we have exploited the dependence between the phases and amplitudes of the optimal inputs to find the capacity gain of non-coherent MISO over SISO. Specifically, we first carefully select a specific optimal input vector based on the discrete optimal effective magnitude distribution. It is then shown that the phase solutions of the optimal signals depend on the power constraints via non-convex quadratic optimization problems, and they can be obtained effectively using our proposed penalized optimization

7

algorithm. From the phase and power solutions, the capacity gain of MISO over SISO can be determined precisely. These interesting results only hold for non-coherent channels, and they certainly make our work different from the results in [15, 27].

- Our study in this work for non-coherent correlated MISO is comprehensive, and it includes both channels under per-antenna power constraints as well as channels under joint per-antenna and sum power constraints. The later includes MISO channels under the sum power constraint as a special case for which the detailed characterization of the optimal inputs and the determination of the capacity are not yet understood.

The rest of the thesis is organized as follows. Chapter 2 introduces the considered non-coherent MISO fading channel and formulate the input-output mutual information and channel capacities. In Chapter 3, we study the detailed characterizations of the capacity-achieving signals and the capacity under per-antenna power constraints. The investigation is extended to the non-coherent multiple input channel under joint per-antenna and sum power constraints in Chapter 4. Finally, conclusions are drawn in Chapter 5.

CHAPTER II

NON-COHERENT MISO: CHANNEL MODEL, MUTUAL INFORMATION, AND

CHANNEL CAPACITY

2.1 Channel Model and Power Constraints

We are interested in a correlated MISO system having M transmit antennas and one

receive antenna, e.g., downlink systems. The input-output model of the system can

be described as:

$$Y = \sum_{k=1}^{M} H_k X_k + N. \tag{2.1}$$

In (2.1), X_k, $1 \le k \le M$, is the complex transmitted signal from antenna k, Y is the

received signal, and N is the complex circularly symmetric Gaussian noise, denoted

as $\mathbb{CN}(0, \sigma_N^2)$. For convenience, we define $\boldsymbol{X} = [X_1, ..., X_M]$ as the channel input

vector. The vector $\boldsymbol{H} = [H_1, ..., H_M]$ consists of M channel gains, each of them

being a complex circularly symmetric Gaussian with zero mean. For a non-coherent

fast fading channel, it is assumed that $\boldsymbol{H}$ is neither known at the transmitter nor at

the receiver, and it changes independently over time, i.e., the fading process is i.i.d.

Furthermore, it is assumed that the covariance matrix of the channel $\Sigma = \mathbf{E}(\boldsymbol{H} \otimes \boldsymbol{H}^{\mathrm{H}})$

of size $M \times M$, with $\mathbf{E}$ being the expectation operation, is a constant matrix. The

matrix Σ reflects the correlation between channel gains as well as the average power

gain on each transmit-receive antennas. The elements of Σ can be estimated using observed channel samples, and they are made available at both the transmitter and receiver. Over the years, various realistic physical and analytical models have been proposed for the design and analysis of multi-antenna systems [28]. In this work, we simply adopt a distance-based correlated multi-antenna channel model [28], which has been widely used in the literature.

In this work, we consider two types of average power constraints, per-antenna and sum power constraints. For per-antenna power constraints, the transmitted signal at each antenna is subject to a constraint as follows:

$$\mathbf{E}\left[|X_k|^2\right] \leq P_k, \ k = 1, 2, ..., M. \tag{2.2}$$

In case of sum power constraint, the total average transmit power from all M antennas is constrained by

$$\mathbf{E}\left[\sum_{k=1}^{M}|X_k|^2\right] \leq P_{sum}. \tag{2.3}$$

A system is under only the condition in (2.2) can be referred to as the per-antenna power constraint system. On the other hand, if both (2.2) and (2.3) are imposed, we have a system under joint per-antenna and sum power constraints. Without loss of generality, we can assume that $P_{sum} < \sum_{k=1}^{M} P_k$. Note that when $P_k \geq P_{sum}$, $\forall \ k = 1, 2, \cdots, M$, the system becomes a traditional system under the sum power constraint only.

2.2 Mutual Information and Channel Capacity

For the channel in (2.1), the conditional output probability density function (PDF) for a given input vector $\boldsymbol{x}$ can be written as [4]:

$$p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) = \frac{1}{\pi(\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x})}\exp\left(-\frac{|y|^2}{\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}}\right). \tag{2.4}$$

Since the phases of the fading coefficients are uniform, the conditional PDF in (2.4) involves only the squared amplitude of Y. Let Y_R and Y_I be the real and imaginary components of Y, and $V = |Y|^2$. By considering the polar transformation $Y_R = \sqrt{V}\cos(\Theta)$ and $Y_I = \sqrt{V}\sin(\Theta)$, we obtain the joint PDF $p_{V,\Theta|\boldsymbol{X}}(v,\theta|\boldsymbol{x})$ as

$$p_{V,\Theta|\boldsymbol{X}}(v,\theta|\boldsymbol{x}) = p_{Y_R,Y_I|\boldsymbol{X}}(y_R(v,\theta), y_I(v,\theta)|\boldsymbol{x})|J(v,\theta)|$$

$$= \frac{1}{2\pi(\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x})}\exp\left(-\frac{v}{\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}}\right). \tag{2.5}$$

Here, $J(v,\theta)$ is the determinant of the Jacobian matrix, which is

$$J(v,\theta) = \det\begin{pmatrix} \frac{\partial y_R}{\partial v} & \frac{\partial y_R}{\partial \theta} \\ \frac{\partial y_I}{\partial v} & \frac{\partial y_I}{\partial \theta} \end{pmatrix} = \det\begin{pmatrix} \frac{\cos(\theta)}{2\sqrt{v}} & -\sqrt{v}\sin(\theta) \\ \frac{\sin(\theta)}{2\sqrt{v}} & \sqrt{v}\cos(\theta) \end{pmatrix} = \frac{1}{2}. \tag{2.6}$$

By integrating the PDF in (2.5) over the phase, we obtain an equivalent conditional PDF of (2.4) as

$$p_{V|\boldsymbol{X}}(v|\boldsymbol{x}) = \frac{1}{\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}}\exp\left(-\frac{v}{\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}}\right). \tag{2.7}$$

For a given input distribution $F_{\boldsymbol{X}}(\boldsymbol{x})$, which is denoted as $F_{\boldsymbol{X}}$ hereafter, the mutual information (MI) between the input and output of the channel can then be expressed as [4]:

$$I(\boldsymbol{X};Y) \triangleq I(F_{\boldsymbol{X}}) = \int\int p_{V|\boldsymbol{X}}(v|\boldsymbol{x})\ln\left(\frac{p_{V|\boldsymbol{X}}(v|\boldsymbol{x})}{p_V(v;F_{\boldsymbol{X}})}\right)\mathrm{d}v\mathrm{d}F_{\boldsymbol{X}}, \tag{2.8}$$

where

$$p_V(v; F_{\boldsymbol{X}}) = \int p_{V|\boldsymbol{X}}(v|\boldsymbol{x}) \mathrm{d}F_{\boldsymbol{X}}, \tag{2.9}$$

is the marginal density of V induced by $F_{\boldsymbol{X}}$. It can be easily observed that the conditional PDF $p_{V|\boldsymbol{X}}(v|\boldsymbol{x})$ is a function of $Z = \sqrt{\boldsymbol{X}^{\mathrm{H}}\Sigma\boldsymbol{X}}$. Z is referred to as the effective magnitude of the input signal $\boldsymbol{X}$, and it follows a distribution F_Z. Therefore, $p_{V|\boldsymbol{X}}(v|\boldsymbol{x})$ can be simplified to the transition probability between a non-negative input Z and a non-negative output V as follows:

$$p_{V|Z}(v|z) \triangleq p(v|z) = \frac{1}{\sigma_N^2 + z^2}\exp\left(-\frac{v}{\sigma_N^2 + z^2}\right), \quad \forall z \geq 0. \tag{2.10}$$

Furthermore, the MI $I(\boldsymbol{X};Y)$ can also be re-written as

$$I(\boldsymbol{X};Y) \triangleq I(F_Z) = \int\int p(v|z)\ln\left(\frac{p(v|z)}{p(v; F_Z)}\right)\mathrm{d}v\mathrm{d}F_Z, \tag{2.11}$$

where

$$p(v; F_Z) = \int p(v|z)\mathrm{d}F_Z. \tag{2.12}$$

As shown in [4, 29, 30], for a non-coherent multi-antenna channel with i.i.d. fading, channel coding can be performed over many independent coherent intervals so that favorable fading gains can be exploited to compensate for unfavorable ones. As a result, information can be transmitted reliably at any rate below the maximum MI between the transmitted and received signals under long-term average power constraints [4, 29, 30]. Therefore, the channel capacity is the supremum of the input-output MI over the set of all input distributions satisfying the average power constraints, either sum or per-antenna constraints. Let denote $\Omega_{\boldsymbol{X}}^{per}$ as the feasible set

of distribution functions $F_{\boldsymbol{X}}$ of $\boldsymbol{X}$ satisfying the per-antenna constraints in (2.2). In a similar manner, $\Omega_{\boldsymbol{X}}^{joint}$ is defined as the feasible set of distributions $F_{\boldsymbol{X}}$ satisfying both (2.2) and (2.3). The channel capacity under per antenna power constraints C^{per} and the channel capacity under joint per-antenna and sum power constraints C^{joint} can be written respectively as follows

$$C^{per} = \sup_{F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}} I(\boldsymbol{X}; Y), \tag{2.13}$$

and

$$C^{joint} = \sup_{F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{joint}} I(\boldsymbol{X}; Y). \tag{2.14}$$

Different from coherent channels, the optimal input distribution, which is denoted as $F_{\boldsymbol{X}}^{*}$, is no longer Gaussian. In the next chapter, we shall characterize the optimal input and study the capacity for the channel under per-antenna power constraints before extending the results to the channel under joint per-antenna and sum power constraints in Chapter 4.

CHAPTER III

NON-COHERENT MISO CHANNEL UNDER PER-ANTENNA POWER

CONSTRAINTS: OPTIMAL INPUTS AND CHANNEL CAPACITY

In this chapter, the focus is on the non-coherent MISO channel under per-antenna power constraints. We first show the existence of the optimal $F_{\boldsymbol{X}}^*$ and the existence and uniqueness of the optimal F_Z^*. We then formulate the KTC and examine the detailed characteristics of F_Z^*. Finally, we construct a highly-structure optimal input $F_{\boldsymbol{X}}^*$, which is helpful for the calculation of the channel capacity.

3.1 Existence of Optimal $F_{\boldsymbol{X}}^*$ and Existence and Uniqueness of Optimal F_Z^*

In [11], Abou-Faycal et. al. proved the existence and uniqueness of the optimal amplitude distribution for non-coherent SISO channels. In the following, we shall extend the result to the MISO channels taking into account the presence of multiple per-antenna constraints. Because of such multiple variables and constrains in the considered MISO channels, a new derivation is needed. The key step is to establish the convexity and compactness of the feasible set $\Omega_{\boldsymbol{X}}^{per}$ and the set Ω_Z^{per}, which is the feasible set of all distribution function F_Z of a non-negative random variable $Z = \sqrt{\boldsymbol{X}^{\mathrm{H}} \Sigma \boldsymbol{X}}$ with $F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}$. The results are given in the following two lemmas.

Lemma 1. $\Omega_{\boldsymbol{X}}^{per}$ *is convex and compact with respect to the weak convergence.*

Proof. For any $F_{\boldsymbol{X}_A}$, $F_{\boldsymbol{X}_B}$ in $\Omega_{\boldsymbol{X}}^{per}$, and $\lambda \in [0,1]$, the convex combination $F_{\boldsymbol{X}} = \lambda F_{\boldsymbol{X}_A} + (1-\lambda)F_{\boldsymbol{X}_B}$ is a distribution function. Furthermore, for each component X_k of vector $\boldsymbol{X}$, we have:

$$\int_{\mathbb{C}} |x_k|^2 dF_{X_k}(x_k) = \int_{\mathbb{C}} |x_k|^2 d\left(\lambda F_{X_{A_k}}(x_k) + (1-\lambda)F_{X_{B_k}}(x_k)\right)$$

$$= \lambda \int_{\mathbb{C}} |x_k|^2 dF_{X_{A_k}}(x_k) + (1-\lambda)\int_{\mathbb{C}} |x_k|^2 dF_{X_{B_k}}(x_k) \leq \lambda P_k + (1-\lambda)P_k = P_k, \quad (3.1)$$

for all $k = 1, 2, ..., M$. Therefore, F_X is in $\Omega_{\boldsymbol{X}}^{per}$ and the convexity of $\Omega_{\boldsymbol{X}}^{per}$ follows.

Regarding the compactness, since the weak* topology on distribution functions is metrizable, it is sufficient to show that the set of distribution functions subject to per-antenna power constraints $\Omega_{\boldsymbol{X}}^{per}$ is *tight* [11, Appendix I.A, p. 1297]. To this end, let define a ball $B(K)$ as $B(K) = \left\{ \boldsymbol{x} \in \mathbb{C}^M : \sqrt{\sum_{i=1}^{M} |x_k|^2} \leq K \right\}$. It then follows that

$$\sum_{k=1}^{M} P_k \geq \sum_{k=1}^{M} \int_{\mathbb{C}} |x_k|^2 dF_{X_k}(x_k) \geq \sum_{k=1}^{M} \int_{|x_k|>K\sqrt{M}} |x_k|^2 dF_{X_k}(x_k)$$

$$\geq \sum_{k=1}^{M} \int_{|x_k|>K/\sqrt{M}} \frac{K^2}{M} dF_{X_k}(x_k) = \frac{K^2}{M} \sum_{k=1}^{M} \int_{|x_k|>K/\sqrt{M}} dF_{X_k}(x_k)$$

$$\geq \frac{K^2}{M} \int_{\boldsymbol{x} \notin B(K)} dF_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{K^2}{M} F_{\boldsymbol{X}}\left\{\boldsymbol{x} \notin B(K)\right\}, \quad \forall K > 0. \quad (3.2)$$

Thus, for every $\epsilon > 0$, there is a ball $B(K)$ such that

$$\sup_{F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}} F_{\boldsymbol{X}}\left\{\boldsymbol{x} \notin B(K)\right\} \leq \frac{M \sum_{k=1}^{M} P_k}{K^2} < \epsilon. \quad (3.3)$$

Thus, $\Omega_{\boldsymbol{X}}^{per}$ is *tight*. It then follows that $\Omega_{\boldsymbol{X}}^{per}$ is compact with respect to the weak convergence. $\square$

15

Using Lemma 1, we have the existence of an optimal input distribution $F_{\boldsymbol{X}}^*$ achieving the supremum (2.13). It is because the MI is a continuous and concave function of the input distribution [10, 31]. In the next lemma, we will establish the convexity and compactness of the feasible set Ω_Z^{per} of F_Z induced by $F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}$.

Lemma 2. Ω_Z^{per} *is convex and compact with respect to the weak convergence.*

Proof. For any $\lambda \in [0, 1]$, we consider $F_Z = \lambda F_{Z_1} + (1 - \lambda)F_{Z_2}$, where F_{Z_1} and F_{Z_2} are induced from two distributions $F_{\boldsymbol{X}_1}$ and $F_{\boldsymbol{X}_2} \in \Omega_{\boldsymbol{X}}^{per}$, respectively. F_Z can then be expressed as:

$$
\begin{aligned}
F_Z(z) &= \lambda \int_{\sqrt{\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}} \leq z} \mathrm{d}F_{\boldsymbol{X}_1}(\boldsymbol{x}) + (1 - \lambda) \int_{\sqrt{\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}} \leq z} \mathrm{d}F_{\boldsymbol{X}_2}(\boldsymbol{x}) \\
&= \int_{\sqrt{\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}} \leq z} \mathrm{d}\left(\lambda F_{\boldsymbol{X}_1} + (1 - \lambda)F_{\boldsymbol{X}_2}\right)(\boldsymbol{x}).
\end{aligned}
\tag{3.4}
$$

Therefore, F_Z is in Ω_Z^{per}, since $\Omega_{\boldsymbol{X}}^{per}$ is convex. Consequently, Ω_Z^{per} is convex.

We now define the ball $B'(K)$ as $B'(K) = \left\{ \boldsymbol{x} \in \mathbb{C}^M : \sqrt{\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}} \leq K \right\}$. Let λ_{min} be the minimum eigenvalue of Σ. Note that because Σ is positive definite, λ_{min} is positive. We then have $\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x} \geq \lambda_{min} \sum_{i=1}^{M} |x_k|^2$. As a result, $B'(K) \subset B(K/\sqrt{\lambda_{min}})$. Combining this with (3.3) leads to

$$
F_Z\{z > K\} = F_{\boldsymbol{X}}\left\{\boldsymbol{x} \notin B'(K)\right\} \leq F_{\boldsymbol{X}}\left\{\boldsymbol{x} \notin B(K/\sqrt{\lambda_{min}})\right\} \leq \frac{\lambda_{min} M \sum_{k=1}^{M} P_k}{K^2},
\tag{3.5}
$$

for all $K > 0$. As a result, we can choose a sufficiently large K such that Ω_Z^{per} is tight. Since Ω_Z^{per} is tight, it is compact with respect to weak convergence. □

By using the MI formula in (2.11) as a function of a single distribution F_Z, the analysis in [11, Appendix I.B, p. 1297] can be applied directly to show that the MI is a continuous and strictly concave function of F_Z [11, Appendix I.B, p. 1297]. Then given the compactness and tightness of Ω_Z^{per} from Lemma 2, it can be concluded that F_Z^* exists, and it is unique.

On the other hand, the uniqueness of optimal $F_{\boldsymbol{X}}^*$ does not hold in general. Consider two distributions $F_{\boldsymbol{X}_A}$ and $F_{\boldsymbol{X}_B}$ with the convex combination $F_{\boldsymbol{X}} = \lambda F_{\boldsymbol{X}_A} + (1-\lambda)F_{\boldsymbol{X}_B}$. Let F_{Z_A}, F_{Z_B} and F_Z be the distributions induced by $F_{\boldsymbol{X}_A}$, $F_{\boldsymbol{X}_B}$ and $F_{\boldsymbol{X}}$ respectively. We then have:

$$I(F_{\boldsymbol{X}}) = I(F_Z) \geq \lambda I(F_{Z_A}) + (1-\lambda)I(F_{Z_B}) = \lambda I(F_{\boldsymbol{X}_A}) + (1-\lambda)I(F_{\boldsymbol{X}_B}). \quad (3.6)$$

In (3.6), the equality can be achieved by using different $F_{\boldsymbol{X}_A}$ and $F_{\boldsymbol{X}_B}$ that produce the same effective magnitude distribution, i.e., $F_{Z_A} = F_{Z_B}$. Therefore, the MI is only a concave function of $F_{\boldsymbol{X}}$. Thus, there might exist multiple solutions of optimal $F_{\boldsymbol{X}}^*$ in $\Omega_{\boldsymbol{X}}^{per}$, each induces the same and unique F_Z^*. In fact, as we shall demonstrate, there exists an infinite number of optimal $F_{\boldsymbol{X}}^*$.

3.2 Kuhn-Tucker Condition (KTC) and Detailed Characteristics of Optimal F_Z^*

Given the uniqueness of F_Z^*, this section shall shed more light on the detailed properties of F_Z^*. To this end, the following proposition states a necessary and sufficient condition for a distribution $F_{\boldsymbol{X}}^*$ to be optimal. This condition is referred to as the KTC.

Proposition 1 (The KTC). *A distribution $F_{\boldsymbol{X}}^*$ is optimal if and only if there exists a set of non-negative $\{\gamma_k\}$, $1 \leq k \leq M$ such that*

$$\Phi_{F_{\boldsymbol{X}}^*}^{per}(\boldsymbol{x}) = \sum_{k=1}^{M} \gamma_k(|x_k|^2 - P_k) + C^{per} - \int p(v|\boldsymbol{x})\ln\left(\frac{p(v|\boldsymbol{x})}{p(v; F_{\boldsymbol{X}}^*)}\right) dv \geq 0, \qquad (3.7)$$

for all $\boldsymbol{x} = [x_1, ..., x_M]$, with the equality being achieved when $\boldsymbol{x}$ belongs to the support of $F_{\boldsymbol{X}}^*$.

Proof. See Appendix A. □

It can be verified that $-\int p(v|\boldsymbol{x})\ln(p(v|\boldsymbol{x}))dv = 1 + \ln\left(\sigma_N^2 + \boldsymbol{x}^H\Sigma\boldsymbol{x}\right)$. Also, observe from (2.7) and (2.9) that $p(v|\boldsymbol{x})\ln(p(v; F_{\boldsymbol{X}}^*))$ depends on $\boldsymbol{x}$ only through $\boldsymbol{x}^H\Sigma\boldsymbol{x}$. Let Γ be a diagonal matrix of size $M \times M$ having γ_k as its diagonal elements. We can then re-write the KTC in (3.7) as:

$$\Phi_{F_{\boldsymbol{X}}^*}^{per}(\boldsymbol{x}) = \boldsymbol{x}^H\Gamma\boldsymbol{x} + K_{F_{\boldsymbol{X}}^*}\left(\boldsymbol{x}^H\Sigma\boldsymbol{x}\right), \qquad (3.8)$$

where $K_{F_{\boldsymbol{X}}^*}\left(\boldsymbol{x}^H\Sigma\boldsymbol{x}\right)$ is a function of $\boldsymbol{x}^H\Sigma\boldsymbol{x}$ for a given $F_{\boldsymbol{X}}^*$, and it is expressed as

$$K_{F_{\boldsymbol{X}}^*}\left(\boldsymbol{x}^H\Sigma\boldsymbol{x}\right) = \int p(v|\boldsymbol{x})\ln(p(v; F_{\boldsymbol{X}}^*))dv + \ln\left(\sigma_N^2 + \boldsymbol{x}^H\Sigma\boldsymbol{x}\right) + C^{per} + 1 - \sum_{k=1}^{M} \gamma_k P_k.$$

$$(3.9)$$

The presence of both $\boldsymbol{x}^H\Gamma\boldsymbol{x}$ and $\boldsymbol{x}^H\Sigma\boldsymbol{x}$ in (3.8) makes it more difficult to examine the KTC and the characteristics of $F_{\boldsymbol{X}}^*$ and $F_{\boldsymbol{Z}}^*$, and the techniques in [11] for a single variable cannot be used. Furthermore, since our considered channels are non-coherent under the power constraints, the proposed tools in [15, 27] to reduce the dimension cannot be applied. It is because they are applicable only to deterministic

18

multi-antenna channels under peak constraints. In order to overcome this issue, we first have the following lemma regarding the KTC coefficients $\{\gamma_k\}$.

Lemma 3. *The KTC coefficients $\{\gamma_k\}$, $1 \leq k \leq M$, in (3.7) are positive. Equivalently, all per-antenna power constraints are active.*

Proof. See Appendix B. □

Then the next lemma states the relationship between $\boldsymbol{x}^H \Sigma \boldsymbol{x}$ and $\boldsymbol{x}^H \Gamma \boldsymbol{x}$ for any mass point $\boldsymbol{x}$ belonging to the support set of an optimal $F_{\boldsymbol{X}}^*$, which is helpful to simplify and analyze the KTC in the subsequent steps.

Lemma 4. *For any non-zero mass point $\boldsymbol{x}^*$ that belongs to any optimal $F_{\boldsymbol{X}}^*$, the ratio $\alpha = \frac{\boldsymbol{x}^{*H} \Sigma \boldsymbol{x}^*}{\boldsymbol{x}^{*H} \Gamma \boldsymbol{x}^*}$ is a constant, and it is equal to the maximum eigenvalue of $\Gamma^{-1} \Sigma$.*

Proof. Let $z = \sqrt{\boldsymbol{x}^{*H} \Sigma \boldsymbol{x}^*}$. Define a set Q_z as $Q_z = \{\boldsymbol{x} | \boldsymbol{x}^H \Sigma \boldsymbol{x} = z^2\}$. From (3.9), it can be seen that $K_{F_{\boldsymbol{X}}^*} \left(\boldsymbol{x}^H \Sigma \boldsymbol{x} \right)$ is the same for all $\boldsymbol{x} \in Q_z$, which includes $\boldsymbol{x}^*$. Then from the KTC in (3.7) and (3.8), the mass point $\boldsymbol{x}^*$ is one of the minimizers of the following quadratic optimization problem:

$$\arg \min_{\boldsymbol{x}} \boldsymbol{x}^H \Gamma \boldsymbol{x}, \text{ subject to } \boldsymbol{x}^H \Sigma \boldsymbol{x} = z^2. \tag{3.10}$$

Since $\{\gamma_k\}$ are positive, Γ is positive definite. We also have Σ as a positive definite matrix. As such, it has been well-known that the minimum value of $\boldsymbol{x}^H \Gamma \boldsymbol{x}$ in the quadratic optimization problem in (3.10) is the product of z^2 and the minimum eigenvalue of $\Sigma^{-1} \Gamma$ [32]. As $\Gamma^{-1} \Sigma$ is the inverse of $\Sigma^{-1} \Gamma$ and any mass point $\boldsymbol{x}^*$ is a minimizer, the lemma is therefore proved. □

19

Now, by applying Lemma 4 to (3.8), we obtain

$$\frac{1}{\alpha}\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x} + K_{F_{\boldsymbol{X}}^*}\left(\boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}\right) = 0, \ \forall \boldsymbol{x} \in \mathrm{supp}(F_{\boldsymbol{X}}^*). \tag{3.11}$$

As a result, we obtain the following condition for all mass points z of F_Z^*:

$$\Upsilon^{per}(z) = \int p(v|z)\ln(p(v; F_Z^*))\mathrm{d}v + \frac{z^2}{\alpha} + \ln\left(\sigma_N^2 + z^2\right) + C_1 = 0, \tag{3.12}$$

where $p(v|z)$ and $p(v; F_Z^*)$ are given in (2.10) and (2.12), respectively, when F_Z^* is used, and $C_1 = C^{per} + 1 - \sum_{k=1}^M \gamma_k P_k$.

Observe that the condition in (3.12) is similar to the KTC in [11] for non-coherent SISO channels. The difference is that we do not know yet whether the strictly inequality can be achieved for all non-negative z. However, similar complex analysis as in [12] can be used to show the discreteness and finiteness of F_Z^*, which is stated in the following theorem.

Theorem 1. *The optimal distribution F_Z^* is discrete and finite.*

Proof. First, assume that F_Z^* has an infinite number of mass points on a bounded interval. Extending $\Upsilon^{per}(z)$ in (3.12) to the complex domain, we obtain $\Upsilon^{per}(\mathrm{z})$ for a complex z as

$$\Upsilon^{per}(\mathrm{z}) = \int p(v|\mathrm{z})\ln(p(v; F_Z^*))\mathrm{d}v + \frac{\mathrm{z}^2}{\alpha} + \ln\left(\sigma_N^2 + \mathrm{z}^2\right) + C_1. \tag{3.13}$$

Using similar arguments as in [12], it is easy to see that $\Upsilon^{per}(\mathrm{z})$ is analytic in the region $\mathcal{D}$ where $\mathrm{Re}\{\sigma_N^2 + \mathrm{z}^2\} > 0$. Then following Bolzano-Weierstrass Theorem [33] and Identity Theorem [26], we have $\Upsilon^{per}(\mathrm{z}) = 0$ for all z satisfying $\mathrm{Re}\{\sigma_N^2 + \mathrm{z}^2\} > 0$.

By replacing $s = \frac{1}{\sigma_N^2 + z^2}$, $s \in \mathcal{D}$, where the region $\mathcal{D}$ is expressed in term of s as $\text{Re}\{1/s\} > 0$, we have (3.13) is equivalent to

$$\int e^{-vs} \ln(p(v; F_Z^*)) \mathrm{d}v = -\frac{1}{s} \left[\frac{1}{\alpha} \left(\frac{1}{s} - \sigma_N^2 \right) - \ln(s) + C_1 \right], \tag{3.14}$$

for all $s \in \mathcal{D}$. The LHS of (3.14) is the unilateral Laplace transform of $\ln(p(v; F_Z^*))$, while the RHS is the Laplace transform of $-\frac{v}{\alpha} + \left[\frac{\sigma_N^2}{\alpha} - C_1 - C_E \right] - \ln(v)$, with C_E being the Euler constant. Given the uniqueness of the Laplace transform for continuous function of bounded variations, we obtain:

$$p(v; F_Z^*) = K \frac{e^{-v/\alpha}}{v}, \tag{3.15}$$

where $K = \exp \left[\frac{\sigma_N^2}{\alpha} - C_1 - C_E \right]$. However, observe that, for any value of K and α, the integral over $(0, \infty)$ of $p(v; F_Z^*)$ in (3.15) is infinite. Therefore, $p(v; F_Z^*)$ in (3.15) is not a valid PDF. As a result, F_Z^* can only have a finite number of mass points on any bounded interval, i.e., F_Z^* is discrete.

Given that, we now assume that the support set of F_Z^* is unbounded. Then for any positive and finite L, we have:

$$p(v; F_Z^*) = \int_0^\infty \frac{1}{\sigma_N^2 + z^2} \exp \left(-\frac{v}{\sigma_N^2 + z^2} \right) \mathrm{d}F_Z^*(z) \tag{3.16}$$

$$\geq \int_L^\infty \frac{1}{\sigma_N^2 + z^2} \exp \left(-\frac{v}{\sigma_N^2 + z^2} \right) \mathrm{d}F_Z^*(z) \geq \exp \left(-\frac{v}{\sigma_N^2 + L^2} \right) \int_L^\infty \frac{1}{\sigma_N^2 + z^2} \mathrm{d}F_Z^*(z)$$

$$\tag{3.17}$$

$$\geq \exp \left(-\frac{v}{\sigma_N^2 + L^2} \right) D_{[F_Z^*, L]}, \quad \forall v \geq 0, \forall \, 0 \leq L < \infty, \tag{3.18}$$

where $D_{[F_Z^*,L]} > 0$. Applying this inequality to (3.12), we obtain the following bound on $\Upsilon^{per}(z)$:

$$\Upsilon^{per}(z) \geq \ln\left(D_{[F_Z^*,L]}\right) - \frac{\sigma_N^2 + z^2}{\sigma_N^2 + L^2} + \frac{z^2}{\alpha} + \ln\left(\sigma_N^2 + z^2\right) + C_1, \forall\, 0 \leq L < \infty. \quad (3.19)$$

It is clear that we can choose L sufficiently large such that $1/\alpha > 1/(\sigma_N^2 + L^2)$, which makes the RHS of (3.19) go to infinity as $z \to \infty$ However, if the support of F_Z^* is unbounded, $\Upsilon(z)$ is equal to 0 infinitely often, which constitutes a contradiction. As a consequence, the optimal F_Z^* must be discrete with a finite number of mass points. $\qquad\square$

It is also worth noting that by following a similar analysis as in [11], we can also show that F_Z^* has a mass point at the origin. However, the proof is omitted here for brevity.

3.3   A Finite and Discrete $F_{\boldsymbol{X}}^*$ and MISO Capacity

As we demonstrate shortly, given the discreteness of F_Z^*, efficient numerical methods such as gradient descent-based methods or Arimoto-Blahut algorithms can be applied to locate the mass points of F_Z^*. However, it is also of importance to further understand the detailed structure of the optimal $F_{\boldsymbol{X}}^*$ and obtain an insight on the behavior of the capacity. Therefore, in this section, we shall construct a highly structured optimal distribution of $\boldsymbol{X}$, denoted as $F_{\boldsymbol{X}}^{D*}$. Then by further exploiting certain amplitude and phase characteristics of the mass points of $F_{\boldsymbol{X}}^{D*}$, we establish a direct connection between MISO and SISO capacities. Note that different from

SISO work in [11], there exist an infinite number of optimal input vectors in MISO. It is then clear that for the considered non-coherent correlated MISO channel, the policy of optimally distributing power among different antennas, and the calculation of MISO capacity as the sum of the capacities of a set of SISO channels do not hold as in deterministic multi-antenna channels. Thus, our approaches and results below are radically different from those in [15, 27] that only considered constant channels.

As the support of Z^* is finite and discrete, we denote its mass points as $\{z_1, \cdots, z_N\}$, with the corresponding probabilities $\{p_1, \cdots, p_N\}$. Without loss of generality, assume that $z_1 < z_2 < \cdots < z_N$. It is then clear that $z_1 = 0$ and z_N is positive. Now, consider an arbitrary optimal distribution $\bar{F}_{\boldsymbol{X}}$. Since $\bar{F}_{\boldsymbol{X}}$ induces F_Z^*, we can choose a mass point $\boldsymbol{x}_N^*$ of $\bar{F}_{\boldsymbol{X}}$ of size $M \times 1$ such that $\boldsymbol{x}_N^{*\mathrm{H}} \Sigma \boldsymbol{x}_N = z_N^2$. For convenience, let denote

$$\boldsymbol{x}_N^* = [r_1^* e^{j\theta_1^*}, r_2^* e^{j\theta_2^*}, \cdots, r_M^* e^{j\theta_M^*}]. \tag{3.20}$$

From $\boldsymbol{x}_N^*$, we obtain $(N-1)$ vectors $\{\boldsymbol{x}_i^*\}$, $1 \leq i \leq N-1$ as follows:

$$\boldsymbol{x}_i^* = \frac{z_i}{z_N} \boldsymbol{x}_N^*. \tag{3.21}$$

We then have the following proposition:

Proposition 2. *Consider a finite and discrete distribution $F_{\boldsymbol{X}}^{D*}$ having $\{\boldsymbol{x}_i^*\}$, $1 \leq i \leq N$, as its mass points as follows:*

$$F_{\boldsymbol{X}}^{D*} = \sum_{i=1}^{N} p_i \delta(\boldsymbol{x} - \boldsymbol{x}_i^*). \tag{3.22}$$

Then $F_{\boldsymbol{X}}^{D*}$ is an optimal distribution of $\boldsymbol{X}$.

Proof. See Appendix C. □

Given the optimality of $F_{\boldsymbol{X}}^{D*}$, and the fact that all of its mass points are linearly scaled versions of $\boldsymbol{x}_N^*$, we can shed further light on the amplitude r_k^* and phase θ_k^*, $1 \leq k \leq M$, of each element of $\boldsymbol{x}_N^*$ in (3.20). In particular, from Lemma 3, we know that all per-antenna power constraints are active. Therefore, when $F_{\boldsymbol{X}}^{D*}$ is used, the power consumed by antenna k^{th} is P_k, and it can be expressed as

$$P_k = \sum_{i=1}^{N} p_i \left| \frac{z_i}{z_N} r_k^* e^{j\theta_k^*} \right|^2 = \frac{r_k^{*2}}{z_N^2} P_Z^{per}, \qquad (3.23)$$

where

$$P_Z^{per} = \sum_{i=1}^{N} p_i z_i^2 \qquad (3.24)$$

is an expected value of Z^2. It then follows that:

$$r_k^* = z_N \sqrt{\frac{P_k}{P_Z^{per}}}, \quad 1 \leq k \leq M. \qquad (3.25)$$

Now, regarding the optimal phases $\boldsymbol{\theta}_{per}^* = [\theta_1^*, \cdots, \theta_M^*]$, from Lemma 4, it is clear that $\boldsymbol{x}_N^*$ is a solution of of the Rayleigh quotient:

$$\boldsymbol{x}_N^* \in \arg \max_{\boldsymbol{x}} \frac{\boldsymbol{x} \Sigma \boldsymbol{x}}{\boldsymbol{x} \Gamma \boldsymbol{x}}. \qquad (3.26)$$

As a result, we have $\boldsymbol{x}_N^*$ is the optimal solution of the problem

$$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{H}} \Sigma \boldsymbol{x}, \quad \text{subject to} \ \boldsymbol{x}^{\mathrm{H}} \Gamma \boldsymbol{x} = \frac{z_N^2}{\alpha}. \qquad (3.27)$$

Now, by applying Lemma 4 for $\boldsymbol{x}_N^*$ and using (3.25), we have:

$$\frac{z_N^2}{\alpha} = \frac{z_N^2 \boldsymbol{x}_N^{*H} \Gamma \boldsymbol{x}_N^*}{\boldsymbol{x}_N^{*H} \Sigma \boldsymbol{x}_N^*} = \frac{z_N^2 \sum_{k=1}^{M} \gamma_k r_k^{*2}}{z_N^2} = \sum_{k=1}^{M} \gamma_k r_k^{*2}. \qquad (3.28)$$

24

Thus, the feasible set of (3.27) contains the set of M amplitude constraints $\{|x_k| = r_k^*\}$. Therefore, $\boldsymbol{x}_N^*$ is the solution of

$$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{H}} \Sigma \boldsymbol{x}, \text{ subject to } \left\{ |x_k| = z_N \sqrt{\frac{P_k}{P_Z^{per}}} \right\}. \tag{3.29}$$

In fact, it can be verified that any of the solutions of (3.29) can be chosen as $\boldsymbol{x}_N^*$. It is because by using any of the solutions of (3.29), we can construct a distribution as in (3.22) that is optimal. Since $z_N/\sqrt{P_Z^{per}}$ is a common factor in all constraints in (3.29), the optimal phase vector $\boldsymbol{\theta}_{per}^*$ is simply a phase of the optimal solution of the following problem:

$$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{H}} \Sigma \boldsymbol{x}, \text{ subject to } \{|x_k| = \sqrt{P_k}\}. \tag{3.30}$$

Equivalently, the optimal phase $\boldsymbol{\theta}_{per}^*$ is the solution of

$$\max_{\boldsymbol{\theta}} \sum_{k=1}^{M} \sum_{l=k+1}^{M} \sigma_{kl} \sqrt{P_k P_l} \cos(\theta_k - \theta_l), \text{ subject to } \boldsymbol{\theta} \in [0, 2\pi]^M. \tag{3.31}$$

It is clear that the optimal $\boldsymbol{\theta}_{per}^*$ involves only the channel covariance matrix Σ and the per-antenna constrains $\{P_k\}$ but not the KTC coefficients $\{\gamma_k\}$. Problem (3.30) is a nonconvex optimization problem because its objective is nonconcave and constraints are nonconvex. By using a new variable $\mathbf{X} = \boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}$ we obtain the following equivalent problem:

$$\max_{\mathbf{X}} \langle \mathbf{X}, \Sigma \rangle, \quad \text{subject to} \tag{3.32a}$$

$$\mathbf{X} \succeq 0, \mathbf{X}(k, k) = P_k, \ k = 1, \dots, M, \tag{3.32b}$$

$$\mathrm{rank}(\mathbf{X}) = 1, \tag{3.32c}$$

where all difficulties are concentrated at the matrix-rank-one constraint (3.32c). The semi-definite programming relaxation based approach (see e.g. [34]) is to drop this matrix-rank-one constraint to treat (3.32) as a semi-definite program. It has been shown in [35] that this approach is only efficient when the solution of the semi-definite program (3.32a)-(3.32b) is of rank-one that is not the case. Following our previous works (see e.g. [35]) we address (3.32) via the following penalized optimization problem

$$\max_{\mathbf{X}} \; \langle \mathbf{X}, \Sigma \rangle + \mu(\lambda_{\max}(\mathbf{X}) - \mathrm{Tr}(\mathbf{X})), \quad \text{subject to} \quad (3.32b), \tag{3.33}$$

where $\lambda_{\max}(\mathbf{X})$ is the maximal eigenvalue of $\mathbf{X}$, and $\mu > 0$ is a penalty parameter. Take feasible point $X^{(0)}$ for (3.32b) by solving

$$\max_{\mathbf{X}} \; \langle \mathbf{X}, \Sigma \rangle, \quad \text{subject to} \quad (3.32b), \tag{3.34}$$

and choose

$$\mu = \frac{\langle X^{(0)}, \Sigma \rangle}{\mathrm{Tr}(X^{(0)}) - \lambda_{\max}(X^{(0)})}.$$

At the κth iteration we solve the following semi-definite optimization problem to generate $X^{(\kappa+1)}$

$$\max_{\mathbf{X}} \; \langle \mathbf{X}, \Sigma \rangle + \mu \left(\lambda_{\max}(X^{(\kappa)}) + (x^{(\kappa)})^H (\mathbf{X} - X^{(\kappa)}) x^{(\kappa)} - \mathrm{Tr}(\mathbf{X}) \right) \quad \text{subject to} \quad (3.32b), \tag{3.35}$$

where $x^{(\kappa)}$ is the normalized ($||x^{(\kappa)}||^2 = 1$) eigenvector corresponding to $\lambda_{\max}(X^{(\kappa)})$. By doing so, we will have

$$\mathrm{Tr}(X^{(\kappa)}) - \lambda_{\max}(X^{(\kappa)}) \to 0+,$$

26

so rank$(X^{(\kappa)}) \to 1$, under which $\sqrt{\lambda_{\max}(X^{(\kappa)})}x^{(\kappa)}$ is the optimal solution of (3.30).

It is worth mentioning that for some special cases of the covariance matrix Σ, $\boldsymbol{\theta}^*_{per}$ can be obtained in closed-form. For example, when all cross channel coefficients are non-negative, i.e. $\sigma_{kl} \geq 0$ for all $k, l \in \{1, ..., M\}$, the phase solution is $\theta^*_1 = \theta^*_2 = ... = \theta^*_M = \theta$ for any θ. Closed-form solutions of $\boldsymbol{\theta}^*_{per}$ can also be derived for any covariance matrix Σ when $M = 2$ and $M = 3$. The results are given in Appendix D.

Given the solutions in (3.25) and (3.31), we are now ready to study the behavior of MISO capacity as compared to SISO capacity. First, it can be verified that P_Z^{per}, the expected value of Z^2, can be calculated as

$$P_Z^{per} = \frac{P_k}{r_k^{*2}}z_N^2 = \frac{P_k}{r_k^{*2}}\boldsymbol{x}_N^{*H}\Sigma\boldsymbol{x}_N^* = \left(\sqrt{\frac{P_k}{r_k^{*2}}}\boldsymbol{x}_N^{*H}\right)\Sigma\left(\sqrt{\frac{P_k}{r_k^{*2}}}\boldsymbol{x}_N^*\right)$$

$$= \left[\sqrt{P_1}e^{j\theta^*_1}, \cdots, \sqrt{P_M}e^{j\theta^*_M}\right]^H \Sigma \left[\sqrt{P_1}e^{j\theta^*_1}, \cdots, \sqrt{P_M}e^{j\theta^*_M}\right], \qquad (3.36)$$

where we have used the fact that $\sqrt{\frac{P_k}{r_k^{*2}}} = \sqrt{\frac{P_i}{r_i^{*2}}}$ for all $k, i \in \{1, ..., M\}$. It then turns out that P_Z^{per} is the maximum value of $\boldsymbol{x}^H\Sigma\boldsymbol{x}$ in (3.30), which is equivalently computed as:

$$P_Z^{per} = \sum_{k=1}^{M} \sigma_{kk}P_k + \sum_{k=1}^{M}\sum_{l=k+1}^{M} 2\sigma_{kl}\sqrt{P_kP_l}\cos(\theta^*_k - \theta^*_l). \qquad (3.37)$$

The next theorem states the relationship between MISO capacity under per-antenna power constraint and SISO capacity.

Theorem 2. *The capacity C^{per} of a non-coherent MISO channel under per-antenna power constraints $[P_1, \cdots, P_M]$ is equal to the capacity of non-coherent SISO channel under power constraint P_Z^{per}, denoted as $C_{SISO}(P_Z^{per})$.*

27

Proof. Consider the non-coherent Rayleigh SISO channel with complex input U and complex output Y under power constraint $\mathbf{E}\left[Z^2\right] \leq P_Z^{per}$, where $Z = |U|$. It can be verified that the mutual information of SISO channel [11] is in the same form as (2.11). Let $F_{Z,SISO}^*$ be the optimal distribution of Z for this channel. Since $F_{Z,SISO}^*$ is finite and discrete [11], let T be the number of mass points of $F_{Z,SISO}^*$ with the corresponding mass points $\{s_t\}$ and probabilities $\{q_t\}$, $1 \leq t \leq T$. Because $\mathbf{E}[Z^2]$ also equals P_Z^{per}, $C_{SISO}(P_Z^{per}) = I(F_{Z,SISO}^*) \geq I(F_Z^*) = I(F_{\mathbf{X}}^*) = C^{per}$. From $F_{Z,SISO}^*$ and the optimal mass point $\mathbf{x}_N^*$ of $F_{\mathbf{X}}^*$ in (3.22), construct a discrete distribution $F_{X,SISO}^*$ for the MISO channel as follows:

$$F_{\mathbf{X},SISO}^* = \sum_{t=1}^{T} q_t \delta(\mathbf{x} - \mathbf{x}_t), \tag{3.38}$$

with $\mathbf{x}_t = \frac{s_t}{z_N}\mathbf{x}_N^*$, $1 \leq t \leq T$. Is is straightforward to show that $F_{\mathbf{X},SISO}^*$ induces $F_{Z,SISO}^*$, and we have $I(F_{\mathbf{X},SISO}^*) = C_{SISO}(P_Z^{per})$. On the other hand, when $F_{\mathbf{X},SISO}^*$ is used for the MISO channel, it can be verified from (3.25) and (3.38) that the power consumed at antenna k is P_k. Thus, $I(F_{\mathbf{X},SISO}^*) \leq I(F_{\mathbf{X}}^*)$. As a result, we have $C_{SISO}(P_Z^{per}) = C^{per}$. $\qquad\square$

The result from Theorem 2 helps us to quantify precisely the SNR gain, i.e., the gain in the ratio of powers, of MISO over SISO. In particular, for the same normalized total transmit power $P_{total} = \sum_{k=1}^{M} \sigma_{kk}P_k$, the improvement in the capacity is:

$$\frac{P_Z^{per}}{P_{total}} = 1 + \frac{\sum_{k=1}^{M}\sum_{l=k+1}^{M} 2\sigma_{kl}\sqrt{P_k P_l}\cos(\theta_k^* - \theta_l^*)}{\sum_{k=1}^{M} \sigma_{kk}P_k}. \tag{3.39}$$

28

Note that when $\{\sigma_{kl}\}$ are non-negative, the gain is simplified to $1+\frac{\sum_{k=1}^{M}\sum_{l=k+1}^{M}2\sigma_{kl}\sqrt{P_k P_l}}{\sum_{k=1}^{M}\sigma_{kk}P_k}$.

Similar to the result in [13] for systems under the average power constraint, the gain is M when $P_k = P$ for all k over fully correlated channels.

3.4 Numerical Results

In this section, numerical examples on non-coherent channel under per-antenna power constraint are provided to illustrate the theoretical results obtained in the previous sections. Unless otherwise stated, we adopt a realistic distance-based correlated multi-antenna channel model as in [28]. In all the results, the signal-to-noise ratio (SNR) is defined as the ratio between the total transmit power and noise power.

3.4.1 Optimal Inputs and KTC

Let first consider a 2-antenna system where the distance d_T between the two antennas is 0.5 of signal's wavelength. The corresponding covariance matrix of this 2×1 channel is given as [28]:

$$\Sigma^{2\times1} = \begin{bmatrix} 1 & -0.3042 \\ -0.3042 & 1 \end{bmatrix}. \tag{3.40}$$

Furthermore, assume that the per-antenna constraints are $P_1 = 2P_2$. To find the optimal input F_Z^*, we simply apply the gradient descent-based method [11,31] and use the KTC to find the globally optimal solution. The amplitudes and the corresponding probabilities of F_Z^* are shown in Fig. 3.1 as a function of SNR. As an illustrative
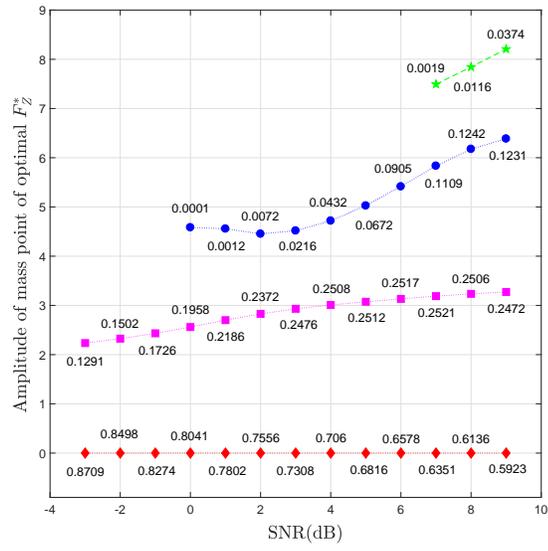
Figure 3.1: The amplitudes and corresponding probabilities of F_Z^* for the per-antenna power constraint channel.
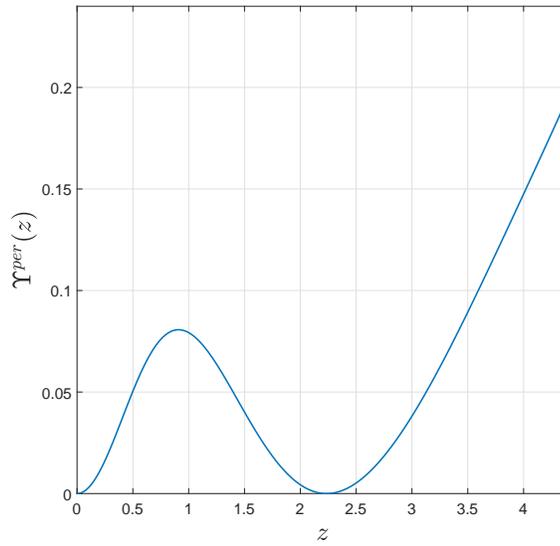
Figure 3.2: The KTC of the effective amplitude at SNR=-3dB for the per-antenna power constraint channel.

example, Fig. 3.2 also plot the KTC of z at SNR$=-3$dB. It is clear that at the two optimal mass points, the KTC equals zero.

It can be observed from Fig. 3.1 that there is always a mass point at zero. At sufficiently low SNRs, F_Z^* has only two mass points. As SNR increases, the number of mass points also increases. Via our extensive numerical trials, we observe that the adopted gradient decent-based method is not stable at higher SNR ranges where the optimal input consists of more mass points, some of which having low probabilities. It is because this method involves the calculation of MI in the form of a definite integral with one limit being infinite, and the sensitivity of this value to the number mass points as well as their locations and probabilities used in each iteration is small. It therefore makes it difficult locate the optimal mass points and the corresponding probabilities with high accuracy at higher SNRs. This drawback has been observed in the literature [11,36]. It is certainly of interest to investigate more effective numerical methods to find the optimal inputs, especially for multi-antenna channels. The topic, however, is beyond the scope of our work, and it deserves further studies.

The optimal distribution $F_{\boldsymbol{X}}^*$ obtained from (3.22) at different SNRs is shown in Table 3.1. Note that the probabilities of the mass points are the same with that of F_Z^*. Furthermore, we do not include the zero mass point in Table 3.1. Over this 2×1 channel, because $\sigma_{12} < 0$, the optimal phases are any pair (θ_1^*, θ_2^*) such that $|\theta_1^* - \theta_2^*| = \pi$. As such, only the optimal amplitude distributions are given. A 2-dimensional (2-D) KTC for real mass points in (3.7) of the input vector at SNR$=-3$dB is also plotted in Fig. 3.3. It can be seen that at the two optimal mass points $(0,0)$

Table 3.1: Table of mass points' amplitude of $F_{\boldsymbol{X}}^*$. Note that the zero mass point is excluded.

	SNR(dB)	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
$\boldsymbol{x}_1$	x_1^1	1.61	1.67	1.75	1.84	1.95	2.04	2.11	2.17	2.21	2.26	2.30	2.33	2.36
	x_1^2	1.14	1.18	1.24	1.30	1.38	1.44	1.49	1.53	1.57	1.60	1.62	1.65	1.67
$\boldsymbol{x}_2$	x_2^1				3.30	3.28	3.21	3.25	3.40	3.62	3.90	4.20	4.45	4.60
	x_2^2				2.33	2.32	2.27	2.30	2.40	2.56	2.76	2.97	3.15	3.25
$\boldsymbol{x}_3$	x_3^1											5.40	5.65	5.91
	x_3^2											3.82	3.99	4.18

and $(1.61, -1.14)$, the KTC is equal to zero. Because an optimal distribution of the input vector $\boldsymbol{X}$ is not unique, there also exists another point that makes the KTC being zero. While this point is not an optimal mass point of $F_{\boldsymbol{X}}^*$, it belongs to another optimal distribution.

To demonstrate the effect of the phases, we consider two 3×1 channels where the distances among the antennas are $d_{T_1} = 0.2$ and $d_{T_2} = 0.25$, respectively. The two corresponding covariance matrices are given as

$$
\Sigma_1^{3 \times 1} = \begin{bmatrix} 1 & 0.6425 & -0.0550 \\ 0.6425 & 1 & 0.6425 \\ -0.0550 & 0.6425 & 1 \end{bmatrix} \text{ and } \Sigma_2^{3 \times 1} = \begin{bmatrix} 1 & 0.4720 & -0.3042 \\ 0.4720 & 1 & 0.4720 \\ -0.3042 & 0.4720 & 1 \end{bmatrix}.
$$

(3.41)

For simplicity, it is assumed that the per-antenna constraints are the same for all antennas. Fig. 3.4 shows the optimal mass points at SNR$=-3$dB for the two
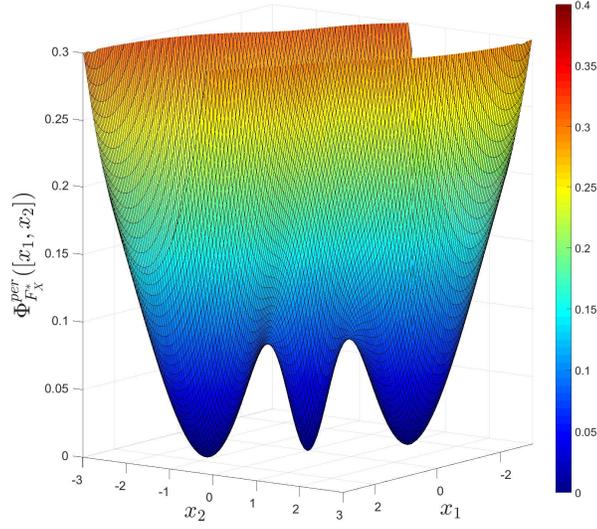
Figure 3.3: The 2-D KTC of the input vector at SNR $= -3$dB for the per-antenna power constraint channel.
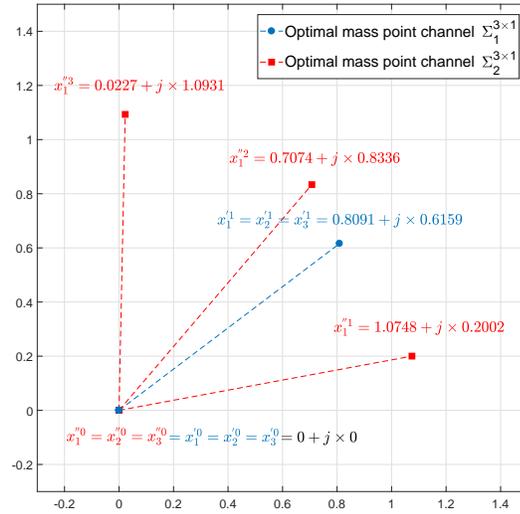
Figure 3.4: Mass points locations of an optimal distribution for channels with covariance matrices $\Sigma_1^{3\times1}$ and $\Sigma_2^{3\times1}$ at SNR $= -3$dB.

channels. It can be observed that both optimal distributions consist of two mass points. While the optimal phases associated with $\Sigma_1^{3\times1}$ are the same across the three antennas, the optimal phase solutions corresponding to $\Sigma_2^{3\times1}$ vary.

3.4.2 Capacity Gains over SISO

It is also of interest to examine the capacity gain of MISO over SISO. Toward this end, besides the antenna distance $d_T = 0.5$, we also consider other correlated channels having $d_T = 0.05$, $d_T = 0.1$, $d_T = 0.15$ and $d_T = 0.20$. Furthermore, it is assumed that the same power constraint is imposed on each transmit antenna. With this condition, it can be verified from (3.39) the capacity gain does not depend on a given operating SNR. Table 3.2 shows the gains (in dB) over different correlated MISO systems having different number of transmit antennas. For each system, the capacity gain is calculated from (3.39) using the optimal phase solutions in (3.30) obtained by the proposed penalized algorithm. It is clear that for a fixed antenna distance d_T, the capacity gain increases when the number of transmit antenna increase. It can also be seen that antenna correlation improves the channel capacity in non-coherent channels.

Table 3.2: Capacity gain of MISO over SISO for correlated channels under per power constraints with different number of transmit antennas M.

M	2	3	4	5	6	7	8	9	10
$d_T = 0.05$	2.96	4.62	5.74	6.57	7.16	7.61	7.94	8.17	8.32
$d_T = 0.10$	2.79	4.20	4.98	5.34	5.47	5.39	5.56	5.84	6.06
$d_T = 0.15$	2.53	3.52	3.77	3.62	4.12	4.42	4.55	4.76	5.00
$d_T = 0.20$	2.15	2.60	2.62	3.14	3.30	3.62	3.89	4.05	4.29
$d_T = 0.50$	1.15	1.91	2.47	2.92	3.29	3.61	3.89	4.13	4.35

CHAPTER IV

EXTENSION TO MISO CHANNELS UNDER JOINT PER-ANTENNA AND

SUM POWER CONSTRAINTS

This chapter shall extend the results developed in Chapter 3 to MISO channels under

joint per-antenna and sum power constraints. In the following, we will first investigate

the KTC and the characteristics of optimal inputs. The connection between MISO

and SISO capacities are then established.

4.1   KTC and Characteristics of Optimal Inputs

It has been well-known that $\Omega_{\boldsymbol{X}}^{sum}$, the set of all distribution functions of $\boldsymbol{X}$ satisfying

the sum power constraint (2.3), is convex and compact [14]. It is then clear that $\Omega_{\boldsymbol{X}}^{joint}$,

the set of input distributions under both the per-antenna and sum power constraints,

is the joint of two convex and compact sets $\Omega_{\boldsymbol{X}}^{per}$ and $\Omega_{\boldsymbol{X}}^{sum}$, which must also be convex

and compact. The result also holds for Ω_{Z}^{joint}, the feasible set of all distributions F_Z

of $Z = \sqrt{\boldsymbol{X}^{\mathrm{H}}\Sigma\boldsymbol{X}}$. The by using a similar analysis as in chapter 3, we can show the

existence of the optimal $F_{\boldsymbol{X}}^*$ and the existence and uniqueness of F_Z^*. Furthermore,

the KTC can be formulated in the following proposition.

Proposition 3 (The KTC). *Under joint per-antenna and sum power constraints,*

a distribution $F_{\boldsymbol{X}}^*$ is optimal if and only if there exists a non-negative β and set of

non-negative $\{\beta_k\}$, $1 \le k \le M$ *such that*

$$\Phi_{F_{\boldsymbol{X}}^*}^{joint}(\boldsymbol{x}) = \beta \left(||\boldsymbol{x}||^2 - P_{sum} \right) + \sum_{k=1}^{M} \beta_k (|x_k|^2 - P_k) + C^{joint} + \int p(v|\boldsymbol{x}) \ln(p(v; F_{\boldsymbol{X}}^*)) \mathrm{d}v \ge 0,$$

(4.1)

for all $\boldsymbol{x} = [x_1, ..., x_M]$, *with the equality being achieved when* $\boldsymbol{x}$ *belongs to the support of* $F_{\boldsymbol{X}}^*$.

Proof. The proof of this proposition is straightforward, and it is omitted here for brevity of the presentation. $\square$

Given the above KTC, the next lemma addresses the positivity of the KTC coefficient β associated with P_{sum}.

Lemma 5. *The KTC coefficient* β *must be positive.*

Proof. Assume $\beta = 0$, and consider the capacity as a function of the total power P for a fixed set of per-antenna constraints $\{P_k\}$. The Lagrange multiplier β is the slope of a line tangent to $C^{joint}(.)$ at P_{sum}. Thus, by concavity and monotonicity, at P_{sum}, $C^{joint}(P) = C^{joint}(P_{sum})$ for all $P \ge P_{sum}$. When P reaches $\sum_{k=1}^{M} P_k$, the system is simply imposed by the per-antenna power constraints only. As a result, we have $C^{joint}(P_{sum}) = C^{per}(P_1, \cdots, P_M)$. On the other hand, assume $[P_{1^{sum}}, \cdots, P_{M^{sum}}]$ be the optimal per-antenna power profile for the joint system when $P = P_{sum}$. It then follows that $C^{joint}(P_{sum}) \le C^{per}(P_{1^{sum}}, \cdots, P_{M^{sum}})$. As a result, $C^{per}(P_{1^{sum}}, \cdots, P_{M^{sum}}) = C^{per}(P_1, \cdots, P_M)$. However, since $P_{sum} < \sum_{k=1}^{M} P_k$, there

must exist at least one antenna k such that $P_{k^{sum}} < P_k$. This results in a contradiction with Lemma 3. Because for the system under per-antenna constraints only, all antennas is active. $\qquad\square$

Different from the per-antenna power constraint system,because $P_{sum} < \sum_{k=1}^{M} P_k$, Lemma 5 indicates that at least one of the per-antenna power constraints must be inactive for the joint system. $\Phi_{F_{\boldsymbol{X}}^*}^{joint}(\boldsymbol{x})$ in (4.1) can be re-written as follows:

$$
\begin{aligned}
\Phi_{F_{\boldsymbol{X}}^*}^{joint}(\boldsymbol{x}) = {} & \boldsymbol{x}^{\mathrm{H}}\Lambda\boldsymbol{x} + \int p(v|\boldsymbol{x})\ln(p(v; F_{\boldsymbol{X}}^*))\mathrm{d}v \\
& + \ln\left(\sigma_N^2 + \boldsymbol{x}^{\mathrm{H}}\Sigma\boldsymbol{x}\right) + C^{joint} + 1 - \beta P - \sum_{k=1}^{M} \beta_k P_k,
\end{aligned} \qquad (4.2)
$$

where Λ is an $M \times M$ diagonal matrix with $\lambda_k = \beta + \beta_k$, $1 \leq k \leq M$, as its diagonal elements. Observe that $\Phi_{F_{\boldsymbol{X}}^*}^{joint}(\boldsymbol{x})$ in (4.2) is in the same form as $\Phi_{F_{\boldsymbol{X}}^*}^{per}(\boldsymbol{x})$ in (3.8). Since $\beta > 0$, we have $\lambda_k > 0$ for all k. Therefore, using the same arguments as in Lemma 4, it is obvious that the ratio $\nu = \frac{\boldsymbol{x}^{*H}\Sigma\boldsymbol{x}^*}{\boldsymbol{x}^{*H}\Lambda\boldsymbol{x}^*}$ is a constant for any mass point $\boldsymbol{x}^*$ of an optimal $F_{\boldsymbol{X}}^*$. Therefore, we can obtain the following condition on all mass points z of F_Z^*:

$$
\Upsilon^{joint}(z) = \int p(v|z)\ln(p(v; F_Z^*))\mathrm{d}v + \frac{z^2}{\nu} + \ln\left(\sigma_N^2 + z^2\right) + C_2 = 0, \qquad (4.3)
$$

where $C_2 = C^{joint} + 1 - \beta P - \sum_{k=1}^{M} \beta_k P_k$. As similar to chapter 3, complex analysis can then be applied to show that F_Z^* is discrete with a finite number of mass points.

4.2 MISO Capacity

Given F_Z^*, we can also construct an optimal F_X^{D*} having the same number of mass points with F_Z^* as in (3.21) for the joint system. However, the determination of the optimal phases $\boldsymbol{\theta}_{joint}^*$ and the effective power P_Z^{joint}, the expected value of Z^2, is not the same. It is because under joint per-antenna and sum power constraints, not all per-antenna power constraints are active. By following similar arguments in chapter 3 to end up with (3.30), it can be verified that $\boldsymbol{\theta}_{joint}^*$ is the phase of the solution of

$$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{H}} \Sigma \boldsymbol{x}, \quad \text{subject to } \{|x_k|^2 \le P_k \text{ and } \sum_{k=1}^{M} |x_k|^2 = P\}. \tag{4.4}$$

In addition, it can be seen that the amplitude solution of (4.4) gives us the optimal power allocation P_k^* at each antenna k, $1 \le k \le M$. The effective power P_Z^{joint} is therefore the maximum value of $\boldsymbol{x}^{\mathrm{H}} \Sigma \boldsymbol{x}$ in (4.4), and it can be expressed as:

$$P_Z^{joint} = \sum_{k=1}^{M} \sigma_{kk} P_k^* + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2\sigma_{kl} \sqrt{P_k^* P_l^*} \cos(\theta_k^* - \theta_l^*), \tag{4.5}$$

where θ_k^* is the kth element of $\boldsymbol{\theta}_{joint}^*$. For this non-convex optimization problem, a similar approach for solving (3.30) via a penalized optimization problem as in (3.33) can be applied to find the optimal solutions. Specifically, instead of using the condition in (3.32b), we simply replace it by the following condition

$$\mathbf{X} \succeq 0, \mathbf{X}(k,k) \le P_k, \sum_{k=1}^{M} \mathbf{X}(k,k) = P, \ k = 1, \dots, M. \tag{4.6}$$

Note that in the special case that $\{\sigma_{kl}\}$ are all non-negative, as similar to the per-antenna power constraint system, it is easy to see that the phase solution is also $\theta_1^* =$

$\theta_2^* = ... = \theta_M^* = \theta$ for any θ, and $P_Z^{joint} = \sum_{k=1}^{M} \sigma_{kk} P_k^* + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2\sigma_{kl} \sqrt{P_k^* P_l^*}$.

Furthermore, the optimal power allocation is then obtained from solving the following problem:

$$\max_{\mathbb{P}_k} \sum_{k=1}^{M} \sigma_{kk} \mathbb{P}_k + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2\sigma_{kl} \sqrt{\mathbb{P}_k \mathbb{P}_l}, \quad \text{subject to } \{\mathbb{P}_k \leq P_k, \text{ and } \sum_{k=1}^{M} \mathbb{P}_k = P\}, \tag{4.7}$$

which is a convex optimization problem of (4.4).

With the obtained P_Z^{joint}, the MISO capacity can then be calculated via the SISO capacity as:

$$C_{MISO}^{joint}(P_1, ..., P_M, P) = C_{SISO}(P_Z^{joint}). \tag{4.8}$$

4.3 Numerical Results

In this section, we provides numerical results to support our theoretical analysis on channel under joint per-antenna and sum power constraints. Fig. 4.1 shows the optimal input distributions of F_Z^* over the 2×1 MISO channel with the covariance matrix in (3.40). Here, it is assumed that the total transmit power is P, while the per-antenna power constraints imposed on the two antennas are $P_1 = 0.9\text{P}$ and $P_2 = 0.2\text{P}$. As similar to the per-antenna channel, there is always a mass point located at zero, and the optimal F_Z^* contains more mass points as SNR increases. For verification, we also plot the KTC of z and the 2-D KTC of x at SNR=-3dB in Fig. 4.2 and Fig. 4.3, respectively.

Finally, Table 4.1 shows the capacity gain P_Z^{joint}/P of MISO over SISO for different MISO channels, where P is the total transmit power, and P_Z^{joint} is calculated
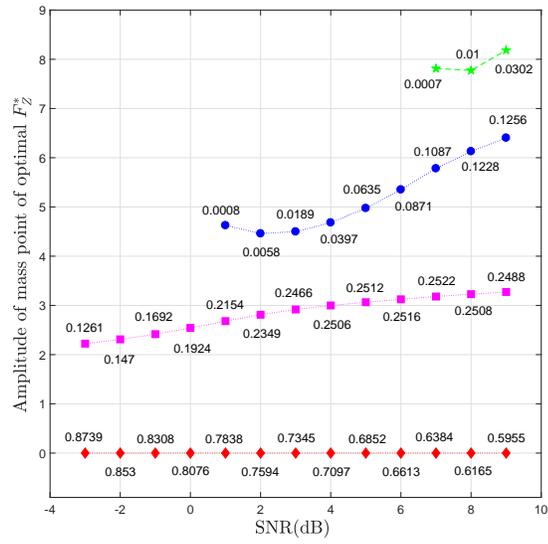
Figure 4.1: The amplitudes and corresponding probabilities of F_Z^* for the 2×1 MISO channel under joint per-antenna and sum power constraints.
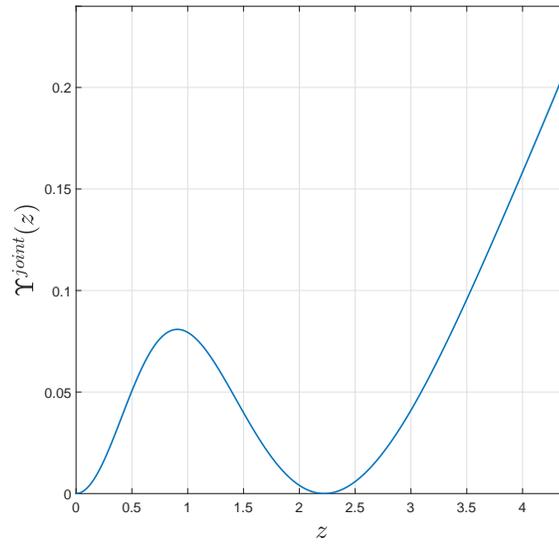
Figure 4.2: The KTC of the effective amplitude at SNR=-3dB for the 2×1 MISO channel under joint per-antenna and sum power constraints.
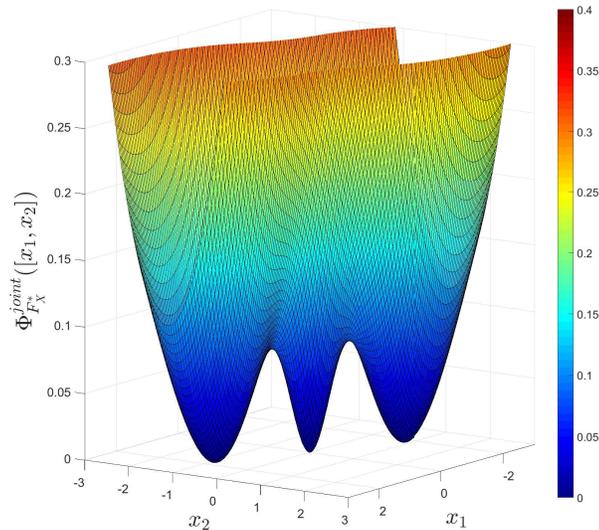
Figure 4.3: The 2-D KTC of the input vector at SNR $= -3$dB for channel under joint constraints.

from (4.5). As similar to the systems under per-antenna power constraints, we also consider five correlated channels having $d_T = 0.05$, $d_T = 0.10$ $d_T = 0.15$, $d_T = 0.20$ and $d_T = 0.50$. For each channel, we assume that each transmit antenna has the same per-antenna constraint $P_i = 2P/M$. It should be noted that P_Z^{joint} in (4.5) is obtained by solving the optimal phases and power allocation scheme simultaneously. As compared to the systems under per-antenna power constraints, we achieve a slightly better gain in the systems under per-antenna and sum power constraints for a given total transmit power P. It is because we have more flexibility in allocating power to each transmit antenna in the systems under per-antenna and sum power constraints .

42

Table 4.1: Capacity gain of MISO over SISO for correlated channels under joint per-antenna and sum power constraints with different number of transmit antennas M.

M	2	3	4	5	6	7	8	9	10
$d_T = 0.05$	2.96	4.63	5.76	6.57	7.18	7.64	7.99	8.25	8.44
$d_T = 0.10$	2.80	4.21	5.02	5.46	5.68	5.78	5.80	5.91	6.26
$d_T = 0.15$	2.53	3.56	3.95	4.04	4.19	4.64	4.79	4.80	5.17
$d_T = 0.20$	2.16	2.75	2.80	3.34	3.56	3.71	4.09	4.14	4.45
$d_T = 0.50$	1.15	1.92	2.48	2.93	3.31	3.63	3.90	4.15	4.37

CHAPTER V

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

5.1   Conclusions

In this thesis, we have comprehensively characterized the optimal signaling schemes and examined the capacity of non-coherent MISO Rayleigh fading channels under per-antenna power constraints and under joint per-antenna and sum power constraints. For both MISO systems, we exploited the solutions of a quadratic optimization problem to simplify the KTC to one dimension. This helps us to prove the uniqueness and discreteness of the optimal effective magnitude input, and construct a finite and discrete optimal input vector distribution. By further exploiting the discreteness and finiteness of the optimal inputs, we have shown that the capacity gains of MISO over SISO for both systems under per-antenna power constraints and joint per-antenna and sum power constraints can be determined precisely via the phase solutions of constrained quadratic optimization problems. Effective algorithms were then proposed to find these phase solutions and calculate the capacity gains. Finally, numerical examples were provided to confirm the analysis and verify the obtained results.

5.2   Future Research Directions

Even though several important information-theoretical aspects of non-coherent MISO Rayleigh fading channels have been obtained, many important research questions remain unanswered for multi-antenna wireless systems. Several potential research directions related to this work are as follows.

- In this work, we adopted the Rayleigh fading channel model. Even though Rayleigh fading has been used extensively in the literature, this wireless fading model is only valid under the assumption that there is no line-of-sight (LOS) component between the transmitter and receiver, and there exist several small objects scattering the signal before it arrives. In many wireless environments, the LOS path does exist, and it can dominate the wireless propagation. Thus, it would be interesting to investigate the fundamental limits of more general multi-antenna channels, such as Rician fading channels.

- The focus of this thesis is only on a MISO channel with a single output. Therefore, the extension of this thesis to the case of a general non-coherent MIMO channel is certainly worth investigating.

BIBLIOGRAPHY

[1] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire. Joint spatial division and multiplexing the large-scale array regime. *IEEE Trans. Inf. Theory*, 59(10):6441–6463, Oct 2013.

[2] H. Yin, D. Gesbert, M. Filippou, and Y. Liu. A coordinated approach to channel estimation in large-scale multiple-antenna systems. *IEEE J. Sel. Areas Commun.*, 31(2):264–273, February 2013.

[3] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya. A comprehensive survey of pilot contamination in massive MIMO 5G system. *IEEE Commun. Surveys Tuts.*, 18(2):905–923, Secondquarter 2016.

[4] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. Inf. Theory*, 45(1):139–157, Jan 1999.

[5] B. M. Hochwald and T. L. Marzetta. Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading. *IEEE Trans. Inf. Theory*, 46(2):543–564, Mar 2000.

[6] D. Xia, J. K. Zhang, S. Dumitrescu, and F. K. Gong. Full diversity non-coherent Alamouti-based Toeplitz space-time block codes. *IEEE Trans. Signal Process.*, 60(10):5241–5253, Oct 2012.

[7] Lizhong Zheng and D. N. C. Tse. Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inf. Theory*, 48(2):359–383, Feb 2002.

[8] Emre Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. on Telecommunications*, 10(6):585–595, 1999.

[9] G. C. Ferrante, T. Q. S. Quek, and M. Z. Win. Revisiting the capacity of noncoherent fading channels in mmWave system. *IEEE Trans. Commun.*, 65(8):3259–3275, August 2017.

[10] R. R. Perera, T. S. Pollock, and T. D. Abhayapala. Non-coherent Rayleigh fading MIMO channels: Capacity and optimal input. In *Proc. IEEE Int. Conf. Commun.*, volume 9, pages 4180–4185, June 2006.

[11] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory*, 47(4):1290–1301, May 2001.

[12] Mustafa Cenk Gursoy, H. Vincent Poor, and Sergio Verdú. The noncoherent Rician fading channel - Part I: Structure of the capacity-achieving input. *IEEE Trans. Wireless Commun.*, 4(5):2193–2206, 2005.

[13] S. A. Jafar and A. Goldsmith. Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information. *IEEE Trans. Wireless Commun.*, 4(3):990–997, May 2005.

[14] J. Sommerfeld, I. Bjelakovic, and H. Boche. On the boundedness of the support of optimal input measures for rayleigh fading channels. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1208–1212, July 2008.

[15] A. ElMoslimany and T. M. Duman. On the capacity of multiple-antenna systems and parallel Gaussian channels with amplitude-limited inputs. *IEEE Trans. Commun.*, 64(7):2888–2899, July 2016.

[16] W. Yu and T. Lan. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Trans. Signal Process.*, 55(6):2646–2660, June 2007.

[17] R. Zhang. Cooperative multi-cell block diagonalization with per-base-station power constraints. *IEEE J. Sel. Areas Commun.*, 28(9):1435–1445, December 2010.

[18] P. L. Cao, T. J. Oechtering, R. F. Schaefer, and M. Skoglund. Optimal transmit strategy for MISO channels with joint sum and per-antenna power constraints. *IEEE Trans. Signal Process.*, 64(16):4296–4306, Aug 2016.

[19] C. E. Chen. MSE-based precoder designs for transmitter-preprocessing-aided spatial modulation under per-antenna power constraints. *IEEE Trans. Veh. Technol.*, 66(3):2879–2883, March 2017.

[20] R. Lopez-Valcarce, N. Gonzalez-Prelcic, C. Rusu, and R. W. Heath. Hybrid precoders and combiners for mmWave MIMO systems with per-antenna power constraints. In *Proc. IEEE Global Commun. Conf.*, pages 1–6, Dec 2016.

[21] H. Tang, W. Chen, and J. Li. Robust joint source-relay-destination design under per-antenna power constraints. *IEEE Trans. Signal Process.*, 63(10):2639–2649, May 2015.

[22] M. Vu. MISO capacity with per-antenna power constraint. *IEEE Trans. Commun.*, 59(5):1268–1274, May 2011.

[23] Zhouyue Pi. Optimal MIMO transmission with per-antenna power constraints. In *Proc. IEEE Global Commun. Conf.*, pages 2493–2498, Dec 2012.

[24] D. Maamari, N. Devroye, and D. Tuninetti. The capacity of the ergodic MISO channel with per-antenna power constraint and an application to the fading cognitive interference channel. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1727–1731, June 2014.

[25] Q. Li, M. Hong, H. T. Wai, W. K. Ma, Y. F. Liu, and Z. Q. Luo. An alternating optimization algorithm for the MIMO secrecy capacity problem under sum power and per-antenna power constraints. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 4359–4363, May 2013.

[26] Serge Lang. *Complex Analysis*. Graduate Texts in Mathematics. Springer, 1999.

[27] B. Rassouli and B. Clerckx. On the capacity of vector Gaussian channels with bounded inputs. *IEEE Trans. Inf. Theory*, 62(12):6884–6903, Dec 2016.

[28] Claude Oestges and Bruno Clerckx. *MIMO Wireless Communications: From Real-World Propagation to Space-Time Code Design*. Academic Press, Inc., Orlando, FL, USA, 2007.

[29] B. M. Hochwald, T. L. Marzetta, and B. Hassibi. Space-time autocoding. *IEEE Trans. Inf. Theory*, 47(7):2761–2781, Nov 2001.

[30] I. Abou-Faycal and B. M. Hochwald. Coding requirement for multiple-antenna channels with unknown rayleigh fading. *Bell Labs. Tech. Memo*, 1999.

[31] M. C. Gursoy, H. V. Poor, and S. Verdu. The capacity of the noncoherent Rician fading channel. Technical Report 70, Princeton Univ, 2002.

[32] Jean Gallier. *Geometric Methods and Applications: For Computer Science and Engineering*. Springer Publishing Company, Incorporated, 2nd edition, 2013.

[33] Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill New York, 3d edition, 1976.

[34] Z. Q. Luo, W. K. Ma, A. M. C. So, Y. Ye, and S. Zhang. Semi-definite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, May 2010.

[35] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo. Non-smooth optimization for efficient beamforming in cognitive radio multicast transmission. *IEEE Trans. on Signal Processing*, 60(6):2941–2951, June 2012.

[36] J. Huang and S. Meyn. Characterization and computation of optimal distributions for channel coding. *IEEE Trans. Inf. Theory*, 51(7):2336–2351, July 2005.

[37] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., 1969.

[38] Joel G. Smith. The information capacity of amplitude and variance-constrained scalar Gaussian channels. *Inf. Control*, 18(3):203 – 219, 1971.

APPENDICES

We know that $I(F_{\boldsymbol{X}})$ is concave in $\Omega_{\boldsymbol{X}}^{per}$. Furthermore, the set of input distributions

forms a convex set, and the per-antenna power constraints are linear functionals of

the input distribution. Therefore, by the theorem of Lagrange Multipliers [37], there

exists a set of non-negative $\{\gamma_k\}$, $1 \leq k \leq M$ such that

$$C^{per} = \sup_{\substack{F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per} \\ \mathbf{E}\left[|X_k|^2\right] \leq P_k \; \forall k}} I(F_{\boldsymbol{X}}) = \sup_{F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}} \left(I(F_{\boldsymbol{X}}) - \sum_{k=1}^{M} \gamma_k \phi_k(F_{\boldsymbol{X}}) \right), \qquad (A.1)$$

where $\phi_k(F_{\boldsymbol{X}}) = \int |x_k|^2 dF_{\boldsymbol{X}} - P_k$. Since $p_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ is a bounded continuous function

of $\boldsymbol{x}$ and $|p(y; F_{\boldsymbol{X}})\mathrm{log}p(y; F_{\boldsymbol{X}})| \leq \min\{1, 1/y^2\}$ for all $F_{\boldsymbol{X}}$, $I(F_{\boldsymbol{X}})$ is weakly differ-

entiable on $\Omega_{\boldsymbol{X}}^{per}$ with weak derivative [11, Appendix I.B, p. 1297]. We also have

weakly differentiable with weak derivative $\phi_k(F_{\boldsymbol{X}})$ [31, Appendix B. Lemma B.1, p.

43]. Hence, by [38], $F_{\boldsymbol{X}}^*$ is optimal if and only if

$$I'_{F_{\boldsymbol{X}}^*}(F_{\boldsymbol{X}}) - \sum_{k=1}^{M} \gamma_k \phi'_{k, F_{\boldsymbol{X}}^*}(F_{\boldsymbol{X}}) \leq 0 \quad \text{for all } F_{\boldsymbol{X}} \text{ in } \Omega_{\boldsymbol{X}}^{per}, \qquad (A.2)$$

This condition can then be explicitly written as

$$-\int \int p(v|\boldsymbol{x})\mathrm{ln}\left(\frac{p(v|\boldsymbol{x})}{p(v; F_{\boldsymbol{X}}^*)}\right) dv dF_{\boldsymbol{X}}(\boldsymbol{x}) + \sum_{k=1}^{M} \gamma_k \int \left(|x_k|^2 - P_k\right) dF_{\boldsymbol{X}}(\boldsymbol{x}) + C^{per} \geq 0,$$

$$(A.3)$$

for all $F_{\boldsymbol{X}} \in \Omega_{\boldsymbol{X}}^{per}$. By using the contradiction arguments in [11], (A.3) is equivalent to

$$-\int p(v|\boldsymbol{x})\ln\left(\frac{p(v|\boldsymbol{x})}{p(v;F_{\boldsymbol{X}}^*)}\right)\mathrm{d}v + \sum_{k=1}^{M}\gamma_k(|x_k|^2 - P_k) + C^{per} \geq 0, \ \forall\boldsymbol{x}, \qquad (A.4)$$

with the equality being achieved at any mass point of any optimal $F_{\boldsymbol{X}}^*$.

APPENDIX B

THE POSITIVITY OF KTC COEFFICIENTS IN (3.7)

Assume that not all KTC coefficients are positive. Without loss of generality, let $\gamma_1 = 0$. Because of the existence of an optimal $F_{\mathbf{X}}^*$ for any positive P_1, we can consider the channel capacity as a function of P_1, i.e., $C^{per}(P_1)$, while keeping all others constraint the same. As a matter of fact, for all $P_1' > P_1$, we have $C^{per}(P_1') \geq C^{per}(P_1)$ since the set of all distribution functions F_X' for the former system contains the set of all distribution functions $F_{\mathbf{X}}$ for the later system. The concavity of mutual information in the input distribution, which is stated in [11], [14], implies that $C^{per}(P_1)$ is also concave. The KTC coefficient γ_1 corresponding to a particular capacity achieving input distribution $F_{\mathbf{X}}^*$ with power $\mathbf{E}[X_1^2] = P_1$ may be interpreted as the slope of a line tangent to $C^{per}(.)$ at P_1. Thus, by convexity and monotonicity, $\gamma_1 = 0$ at P_1 implies that $C^{per}(P_1') = C^{per}(P_1)$ for all $P_1' \geq P_1$. On the other hand, by considering the capacity as $P_1 \to \infty$, we obtain

$$\lim_{P_1 \to \infty} C_{MISO}^{per}(P_1, P_2, ..., P_M) \geq \lim_{P_1 \to \infty} C_{MISO}^{per}(P_1, 0, ..., 0) = \lim_{P_1 \to \infty} C_{SISO}(P_1) = \infty,$$

(B.1)

where the last equality follows the analysis of SISO capacity in [11, Section IV, p. 1293]. This shows the impossibility of having $C^{per}(P_1') = C^{per}(P_1)$ for all $P_1' \geq P_1$.

The result thus leads to the contradiction in the assumption of the inactivity of at least one constraint.

APPENDIX C

PROOF OF PROPOSITION 2

It is clear from (3.21) and (3.22) that $F_{\boldsymbol{X}}^{D*}$ induces F_Z^*. It implies that $p(v; F_{\boldsymbol{X}}^{D*}) = p(v; F_Z^*) = p(v; \bar{F}_{\boldsymbol{X}})$ for all $v \geq 0$. As a result, it can be verified that:

$$\Phi_{F_{\boldsymbol{X}}^{D*}}^{per}(\boldsymbol{x}) = \Phi_{\bar{F}_{\boldsymbol{X}}}^{per}(\boldsymbol{x}) \geq 0, \tag{C.1}$$

for all $\boldsymbol{x} \in \mathbb{C}^{M \times 1}$. In addition, since $\boldsymbol{x}_N^*$ is a mass point of optimal $\bar{F}_{\boldsymbol{X}}$, we then have the KTCs $\Phi_{\bar{F}_{\boldsymbol{X}}}^{per}(\boldsymbol{x}_N) = 0$. Furthermore, because $\boldsymbol{x}_i^*$, $1 \leq i \leq N - 1$, is a scaled version of $\boldsymbol{x}_N^*$ with the scaling factor z_i/z_N, it follows from (3.11) and (3.12) that $\Phi_{\bar{F}_{\boldsymbol{X}}}^{per}(\boldsymbol{x}_i) = \Upsilon^{per}(z_i) = 0$. As a result, $F_{\boldsymbol{X}}^{D*}$ satisfies the necessary condition of the KTC in (3.7), and it is optimal.

APPENDIX D

CLOSED-FORM SOLUTIONS OF $\boldsymbol{\theta}^*$ FOR 2×1 AND 3×1 CHANNELS

For the 2×1 MISO channel, the solutions are trivial. Specifically, when $\sigma_{12} > 0$, we have $\theta_1^* = \theta_2^*$. If $\sigma_{12} < 0$, it is then clear that $\theta_1^* = \pi - \theta_2^*$. For the 3×1 channel, without loss of generality, we can assume that $\theta_3 = 0$. Using the first order necessary Lagrange condition, the optimal phase $\boldsymbol{\theta}^*$ must satisfy the following:

$$
\begin{cases}
-A_3\sin(\theta_1 - \theta_2) + A_2\sin(-\theta_1) = 0, \\
-A_1\sin(\theta_2) + A_3\sin(\theta_1 - \theta_2) = 0,
\end{cases}
\tag{D.1}
$$

where $A_1 = \sigma_{23}\sqrt{P_2 P_3}$, $A_2 = \sigma_{31}\sqrt{P_3 P_1}$, and $A_3 = \sigma_{12}\sqrt{P_1 P_2}$. When either A_1, A_2, or A_3 is 0, the solutions are trivial. Otherwise, we obtain:

$$
\leftrightarrow
\begin{cases}
\frac{-A_2}{A_3}\sin(\theta_1) = \sin(\theta_1)\cos(\theta_2) - \sin(\theta_2)\cos(\theta_1), \\
\frac{A_1}{A_3}\sin(\theta_2) = \sin(\theta_1)\cos(\theta_2) - \sin(\theta_2)\cos(\theta_1).
\end{cases}
\tag{D.2}
$$

After some manipulations, we obtain:

$$
\frac{-A_1 A_2}{A_3} = \pm \sqrt{A_1^2 - A_2^2\sin^2(\theta_1)} \pm A_2\sqrt{1 - \sin^2(\theta_1)},
\tag{D.3}
$$

which gives us several closed-form solutions of $\sin^2(\theta_1)$. These solutions can then be substituted back to the cost function so that the global maximizer can be obtained.