On Capacity of Non-Coherent Rayleigh Fading MIMO Channels

Rasika R. Perera, Kien Nguyen, Tony S. Pollock and Thushara D. Abhayapala

Research School of Information Sciences and Engineering,
The Australian National University, Canberra ACT 0200, Australia
{Rasika.Perera, Thushara.Abhayapala}@anu.edu.au

Abstract— This paper investigates the capacity of discrete time uncorrelated Rayleigh fading multiple input multiple output (MIMO) channels with no channel state information (CSI) at both the transmitter and the receiver. We prove that to achieve the capacity, the amplitude of the multiple input needs to have a discrete distribution with a finite number of mass points with one of them located at the origin. We show how to compute the capacity numerically in multi antenna configuration at any signal to noise ratio (SNR) with the discrete input using the Kuhn-Tucker condition for optimality. Furthermore, we show that at low SNR, the capacity with two mass points is optimal. Since the first mass point is necessarily located at the origin, we argue that at low SNR, on-off keying is optimal for any antenna number. As the number of receiver antennas increases, the maximum SNR at which two mass points are optimal decreases.

I. INTRODUCTION

A. Motivation and Background

The capacity achieving input distribution of non-coherent Rayleigh fading MIMO channels has been an open problem for some time. Early work of [1] using a block fading model gave some insights into the characteristics of the optimal input, with explicit calculations for the special case of single input and single output at high SNR. In [1] it is shown that in a non-coherent Rayleigh fading MIMO channel, no capacity gain is achieved by increasing the number of transmitter antennas beyond the channel coherence time. The general structure of the input signal matrix that achieves the capacity was given, along with the capacity asymptotically in channel coherence time for a single input single output (SISO) system and the signal density that achieves it.

Work in [1] is extended in [2] by taking the channel coherence time into account, and showed that the norm of the transmitted signal on each antenna must be higher than the noise level for high SNR. The asymptotic capacity is computed at high SNR in terms of the channel coherence time, and the number of transmit and receive antennas. The non-coherent channel capacity is compared to the promised capacity increase using MIMO in coherent Rayleigh fading channels [3], [4]. Also it is shown that the non-coherent and coherent capacities are asymptotically equal at low SNR. Hence indicating that in the low SNR regime, to a first order, there is no capacity penalty for not knowing the channel at the receiver, unlike in the high SNR.

In [5], the maximum capacity loss due to lack of receiver CSI for a wideband MIMO channel in Rayleigh fading is considered. The maximum penalty to be paid in terms of capacity not having the CSI at the receiver is shown. Furthermore, it is conjectured that on-off signaling is optimal, however no proof was given. The SISO non-coherent Rayleigh fading channel is extensively studied in [6] showing the optimal input is discrete with a finite number of mass points. Capacity is computed numerically choosing the optimal number of mass points, their probabilities and locations.

B. Problem Definition and Contributions

The previous results do not clearly show neither the optimal input nor the capacity at any SNR. A question arises whether there exists a proper method to compute the optimal input and capacity. In this paper, we answer this with the following contributions:

1) In section II, we develop a proper channel model noting scalar properties in Rayleigh fading due to absence of the phase information.
2) In section III, we prove that the capacity achieving input distribution (i.e. the amplitude of multiple inputs) of the non-coherent Rayleigh fading MIMO channel is discrete with a finite number of mass points, one necessarily located at the origin.
3) In section IV, we show how to compute the capacity with the optimal input, which is an extension to work presented in [6] for a single antenna system. Furthermore, we show that at low SNR, on-off keying is optimal and the input power at which it is optimal decreases with the increase of receiver diversity.

II. CHANNEL MODEL

Consider the following time-varying non-coherent Rayleigh fading MIMO channel model

$$Y = HX + N, \qquad (1)$$

where the output Y is $n_r \times 1$, the channel gain matrix H is $n_r \times n_t$, the input X is $n_t \times 1$ and the noise N is $n_r \times 1$. Each element of H, $h_{ij}, i = 1, ..., n_r, j = 1, ..., n_t$ and N is assumed to be zero mean circular complex Gaussian random variables with a unit variance in each dimension. n_t and n_r denote the number of transmit and receive antennas respectively. $X = |X|$ and $Y = |Y|$ denote scalar random variables whilst $x \in X$ and $y \in Y$ represent the instantaneous realisations of X and Y.

The Euclidean norm is denoted by $|\cdot|$. It is assumed that the input is average power limited with constraint $\int x^2 p_X(x)dx \leq P_{av}$. Furthermore, we use $\Gamma(\cdot)$ and $\Psi(\cdot)$ to indicate the Gamma and Psi functions. Differential entropy of $x \in X$ is denoted by $h(X)$ and the mutual information between x and $y \in Y$ is designated by $I(X;Y)$. All the differential entropies and the mutual information are defined to the base "e", hence the results are expressed in "nats". It is assumed that neither the receiver nor the transmitter has the knowledge of perfect CSI except the fading statistics.

III. CHANNEL CAPACITY

The conditional output probability density function (pdf) of the non-coherent Rayleigh fading MIMO channel with n_r uncorrelated receivers is given by

$$p_{Y|X}(y|x) = \frac{y^{2n_r-1}\exp\left(-\frac{y^2}{2(1+x^2)}\right)}{2^{n_r-1}\Gamma(n_r)(1+x^2)^{n_r}}, \quad (2)$$

and represents the distribution of the magnitudes when Jacobian coordinate transformation is applied on $2n_r$ dimensions [7].

A. Mutual Information

Using (2), we write the mutual information

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= -\int_0^\infty p_Y(y;G)\log p_Y(y;G)dy \\ &\quad - \frac{1}{2}\int_0^\infty \log(1+x^2)dG_X(x) + V(n_r), \quad (3) \end{aligned}$$

between the input and the output of channel (1) where

$$V(n_r) = -\log\left[\frac{\Gamma(n_r)}{\sqrt{2}}\right] + \left(n_r - \frac{1}{2}\right)\Psi(n_r) - n_r, \quad (4)$$

and $G_X(x) \triangleq \int p_X(x)dx$ is the cumulative input distribution function [7]. The channel capacity is the supremum of (3) over the set of all input distributions satisfying the input power constraint $\int x^2 p_X(x)dx \leq P_{av}$, i.e.,

$$C = \sup_{\substack{G_X(.) \\ E\{|x|^2\}\leq P_{av}}} I(G_X) \quad (5)$$

where $I(G_X) \triangleq I(X;Y)$ and $p_Y(y;G) = \int_0^\infty p_{Y|X}(y|x)dG_X(x)$ is the marginal probability density induced by the input distribution G_X. The existence of an optimal amplitude distribution achieving the supremum in (5) can be shown proving i) the mutual information is continuous and concave in the input distribution function, and ii) the set of input distribution functions that meet the constraint is compact [6]. The following lemma gives a necessary and sufficient condition for an amplitude distribution $G_0 \in G_X$ to be optimal.

Lemma 1: For the uncorrelated Rayleigh fading MIMO channel with the input average power constrained P_{av}, G_0

is the capacity achieving input amplitude distribution if and only if there exist λ such that the following is satisfied $\forall x \geq 0$

$$\begin{aligned} \int_0^\infty p_{Y|X}(y|x)\log p_Y(y;G_0)dy \\ + \frac{1}{2}\log(1+x^2) + C - V(n_r) + \lambda(x^2 - P_{av}) \geq 0 \quad (6) \end{aligned}$$

with equality if $x \in E_o$ where E_0 is the set of points of increase of G_0.

The condition (6) is known as the Kuhn-Tucker condition for the optimal input distribution which can be used to characterise its behavior.

Proof: See appendix VI-A. ∎

B. Input Distribution

We adopt the same principle in proving the discrete character of the optimal input X^* given in [6] for the single antenna system. Therefore, X^* should possess one of the following properties:

1) the support set contains an interval,
2) it is discrete, with an infinite number of mass points on some bounded interval,
3) it is discrete and infinite, but with only a finite number of mass points on any bounded interval, or
4) it is discrete with a finite number of mass points.

However, the proof is not a straightforward extension from single antenna to multi antenna systems.

Let's assume both (1) and (2) holds, and define $u = 1/(1+x^2)$, $z = y^2/2$. The support set U_X has infinite number of distinct points and the Kuhn-Tucker condition holds with equality for all real $u \in (0,1]$ [6]. In this case, using the equality in (6) we write

$$\begin{aligned} \frac{u^{n_r}}{2^{n_r}\Gamma(n_r)}\int_0^\infty y^{2n_r-1}\exp\left(-\frac{uy^2}{2}\right)\log[p_Y(y)]dy = \\ -\lambda\left(\frac{1}{u} - 1 - P_{av}\right) - C + \frac{1}{2}\log u + V(n_r). \quad (7) \end{aligned}$$

With the pdf of new variable z, $p_Z(z) = \frac{1}{\sqrt{2z}}p_Y(y)|_{y=\sqrt{2z}}$, we have

$$\begin{aligned} \int_0^\infty e^{-uz}\left\{z^{n_r-1}\log[\sqrt{2z}p_Z(z)]\right\}dz = \frac{\Gamma(n_r)}{u^{n_r}} \\ \times \left\{-\lambda\left(\frac{1}{u} - 1 - P_{av}\right) - C + \frac{1}{2}\log u + V(n_r)\right\}, \quad (8) \end{aligned}$$

where the left hand side (LHS) of (8) is the Laplace transformation of the function $z^{n_r-1}\log[\sqrt{2z}p_Z(z)]$. Taking the inverse Laplace transformations [8, pages 1020-1030] we obtain the output pdf

$$p_Y(y) = \frac{\sqrt{2}}{y}Me^{\frac{-\lambda y^2}{2n_r}}, \quad (9)$$

where

$$M = \exp\left\{\lambda(1+P_{av}) - C + V(n_r) + \frac{\Psi(n_r)}{2}\right\}.$$

However, for any $\lambda/2n_r$, the integral over $(0, \infty)$ is infinite and hence the $p_Y(y)$ in (9) cannot be a valid pdf, negating our original assumptions (1) and (2).

Assume the case (3) holds for X^*, then the support set U^* can be written as a sequence $\{u_i\}$ converging to 0. With $Pr[U = u_i] = p_i \neq 0$, $i = 1, 2, ..., N$ we get

$$
p_Y(y) = \sum_{i=0}^{\infty} p_i \left\{ \frac{y^{2n_r - 1} u_i{}^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r - 1} \Gamma(n_r)} \right\}
$$

$$
> p_i \left\{ \frac{y^{2n_r - 1} u_i{}^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r - 1} \Gamma(n_r)} \right\}, \ \forall \ y \ \geq 0. \quad (10)
$$

Since $\log t_1 > \log t_2$, for all $t_1 > t_2 > 0$, we can pose the following inequality

$$
\int_0^{\infty} p_{Y|U}(y|u) \log[p_Y(y)] > \int_0^{\infty} p_{Y|U}(y|u) \log\left[\frac{y^{2n_r - 1} u_i{}^{n_r}}{2^{n_r - 1} \Gamma(n_r)}\right.
$$

$$
\left. \times \exp\left(-\frac{u_i y^2}{2}\right)\right] dy. \quad (11)
$$

We use the integral solutions [9, Page 260-265]

$$
\int_0^{\infty} x^{2n - 1} e^{-kx^2} dx = \frac{(n - 1)!}{2k^n},
$$

and

$$
\int_0^{\infty} x^t e^{-u_1 x^2} \log[v_1 x^t] dx = \frac{1}{2} u_1{}^{-\frac{1+t}{2}} \log v_1 \Gamma\left(\frac{1+t}{2}\right)
$$

$$
+ \frac{u_1{}^{-\frac{t}{2}} t}{4\sqrt{u_1}} \Gamma\left(\frac{1+t}{2}\right) \left\{\Psi\left(\frac{1+t}{2}\right) - \log u_1\right\},
$$

to simplify (11) into

$$
\int_0^{\infty} p_{Y|U}(y|u) \log[p_Y(y)] > \left(n_r - \frac{1}{2}\right) \Psi(n_r) - n_r
$$

$$
+ \log\left[\frac{\sqrt{2} p_i u_i{}^{n_r}}{u^{n_r - \frac{1}{2}} \Gamma(n_r)}\right] - n_r\left(\frac{u_i}{u}\right). \quad (12)
$$

The result in (12) can be used to derive the following bound on LHS of (6).

$$
\text{LHS} \geq \frac{(\lambda - n_r u_i)}{u} - \lambda(1 + P_{av}) + C - V(n_r)
$$

$$
+ \left(n_r - \frac{1}{2}\right) \Psi(n_r) - n_r + \log\left[\frac{\sqrt{2} p_i}{\Gamma(n_r)} \left(\frac{u_i}{u}\right)^{n_r}\right]
$$

$$
(13)
$$

$$
= \frac{(\lambda - n_r u_i)}{u} + O\left(\frac{1}{u}\right). \quad (14)
$$

This lower bound diverges to ∞ for $\lambda > n_r u_i$ when $u \to 0$, but the LHS of (6) is zero on the support set U^*, hence by contradiction $\lambda \leq n_r u_i$, where $\lambda \leq 0$ when $u \to 0$. Therefore, the only possibility is that $\lambda = 0$ if the input X^* is discrete with infinite mass points. However, the Kuhn-Tucker theorem [10] for convex functions (the mutual information and hence the channel capacity is concave [6]) states that the Lagrangian multiplier $\lambda \geq 0$ on the support set which optimises the objective function, negating the original assumption.

Since X^* does not possess any of the properties 1)-3), the only possibility for the optimal input amplitude distribution that maximises (5) is discrete with a *finite* set of mass points.

Comments:

i) For the first time in the information theory literature, we have proved that to achieve capacity of discrete time uncorrelated Rayleigh fading MIMO channel without CSI, the L_2 norm of the input distribution $X = |\boldsymbol{X}|$ has to be discrete with a finite number of mass points.

ii) However, we still need to find where the mass points are located and their probabilities.

iii) Locating these mass points analytically is very difficult if not impossible.

iv) Tedious numerical methods could be used to locate mass points numerically. In section IV we find mass point locations for specific MIMO systems using numerical techniques.

v) Existence of a mass point at zero is proven for a single antenna system [6]. In next section, we show how to prove this for a MIMO system.

C. Mass Point Locations

We begin with the following lemma:

Lemma 2: The optimal input distribution X^* of a non-coherent uncorrelated Rayleigh fading MIMO channel contains necessarily a mass point located at the origin.

Proof: Since the optimal input X^* is discrete with finite number of mass points, we use the distribution function

$$
G_X^*(x) = \sum_{i=1}^{N} p_i \delta(x - x_i), \quad (15)
$$

where $0 \leq x_0 < x_1 < ... < x_N$. The mutual information for this input distribution is given by

$$
I(\boldsymbol{X}; Y) = \sum_{i=0}^{N} p_i \int_0^{\infty} p_{Y|X}(y|x_i)
$$

$$
\times \log\left[\frac{p_{Y|X}(x|x_i)}{\sum_{j=1}^{N} p_j p_{Y|X}(x|x_j)}\right] dy. \quad (16)
$$

Using $z = y^2/2$, and differentiating with respect to $x_0 \geq 0$, we get

$$
\frac{\partial I(\boldsymbol{X}; \boldsymbol{Z})}{\partial x_0} = p_0 \int_0^{\infty} \frac{\partial}{\partial x_0} p_{Z|X}(z|x_0)
$$

$$
\times \log\left[\frac{p_{Z|X}(z|x_0)}{\sum_{j=1}^{N} p_j p_{Z|X}(z|x_j)}\right] dz \quad (17)
$$

where

$$
\frac{\partial}{\partial x_0} p_{Z|X}(z|x_0) = \frac{2x_0}{(1 + x_0{}^2)^2} [z - n_r(1 + x_0{}^2)] p_{Z|X}(z|x_0). \quad (18)
$$

Let's define

$$
J(z) \triangleq \log\left[\frac{p_{Z|X}(z|x_0)}{\sum_j p_j p_{Z|X}(z|x_j)}\right], \quad (19)
$$

then (17) becomes

$$\frac{\partial I(\boldsymbol{X};\boldsymbol{Z})}{\partial x_0} = \frac{2x_0 p_0}{(1+x_0^2)} \int_0^\infty [z - n_r(1+x_0^2)]$$
$$\times\, p_{Z|X}(z|x_0) J(z)\, dz \tag{20}$$

where $n_r(1+x_0^2)$ is the mean value of $p_{Z|X}(z|x)$.

Corollary 1: The function $J(z)$ in (19) is a decreasing function for $0 \leq x_0 < x_1 < ... < x_N$.

Proof: The ratio

$$\frac{p_Z(z)}{p_{Z|X}(z|x_0)} = p_0 + \sum_{i=1}^{N} \frac{(1+x_0^2)^{n_r}}{(1+x_i^2)^{n_r}}$$
$$\times \exp\left[z \left(\frac{1}{1+x_0^2} - \frac{1}{1+x_i^2} \right) \right], \tag{21}$$

is an increasing function since

$$(1+x_0^2)^{-1} > (1+x_1^2)^{-1} > ... > (1+x_N^2)^{-1}$$

for $0 \leq x_0 < x_1 < ... < x_N$. Therefore, $J(z)$ in (19) is a decreasing function due to logarithm of the reciprocal of ratio given in (21). ∎

Using corollary 1 with the Lemma 1 in [6], we conclude that the derivative is negative with respect to x_0 for $0 \leq x_0 < x_1 < ... < x_N$. Therefore, X^* with $x_0 \geq 0$ can not produce a local maximum, hence the input distribution $G_{X^*}(x)$ necessarily has a mass point located at the origin. ∎

IV. NUMERICAL RESULTS

The capacity is achievable once the optimal number of mass points, their probabilities and locations are found satisfying the Kuhn-Tucker condition stated in Lemma 1.

Fig. 1 depicts the channel capacity as a function of input power for $n_r = \{1, 2, 3, 5\}$. The capacity results obtained for both two and three mass points are shown. It is clear that at low input power, there is no difference in capacity in either case. Also, it is evident that as the number of receivers increases, the maximum input power at which two mass points are inadequate decreases. Fig. 2 shows the difference in simulated capacity in both cases. In this analysis we conclude that at very low SNR, the optimal input distribution has two mass points, one located at the origin. Hence the on-off keying is optimal at low SNR in non-coherent Rayleigh fading MIMO channels. Same input is shown to be optimal in [6] for SISO systems.

The probability distribution which optimises the channel capacity with three mass points is shown in Fig. 3 with five receive antennas. Similar to $n_r = 1$, probability of the third mass point is zero at low input power. Also, at low SNR, zero mass point dominates with a high probability. Fig. 4 depicts the Kuhn-Tucker condition (6) when $P_{av} = 1.4$ for a single receiver. As claimed in Lemma 1, it is above zero except the optimal mass point locations where the Kuhn-Tucker condition equals to zero.
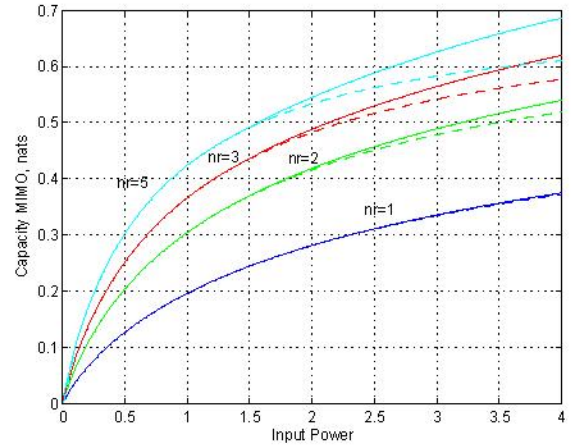

1: Capacity of non-coherent Rayleigh fading MIMO channel vs input power using two and three mass points for different number of receiver antennas $n_r = \{1, 2, 3, 5\}$. The dashed lines show the capacity having two mass points for each receiver configuration.
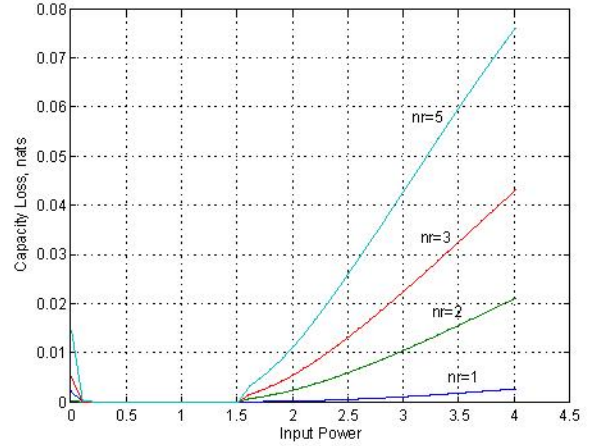
2: Loss in channel capacity with two mass points against three vs input power for $n_r = \{1, 2, 3, 5\}$. High values shown at very low SNR is due to oscillation of optimal mass points and should be ignored.
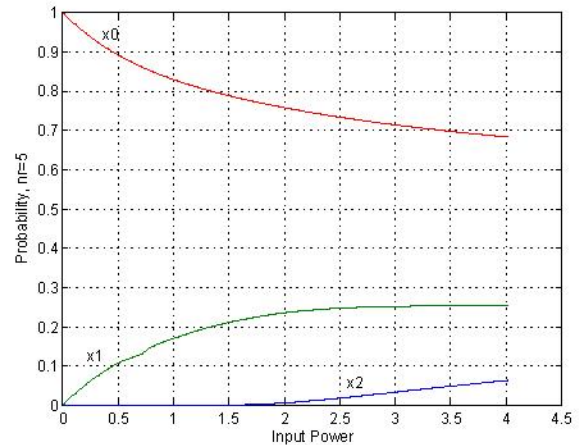
3: Probability distribution of optimal mass points vs input power for $n_r = 5$.
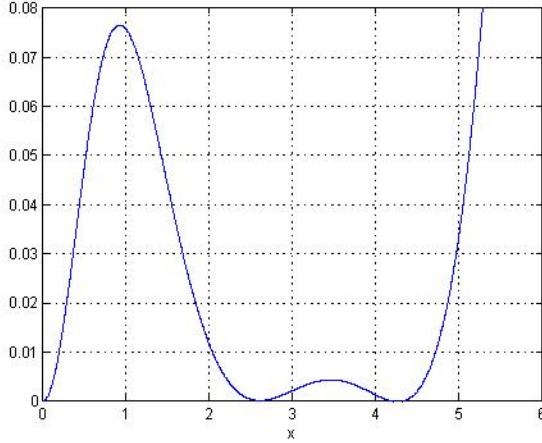
4: The Kuhn-Tucker condition (6) for $P_{av} = 1.4$, $\lambda = 0.08819532394409$ and $p_X(x) = 0.79549244782344\delta(x) + 0.20444300113415\delta(x - 2.6160) + 0.0000645510564\delta(x - 4.1451)$. The channel capacity C=0.23549289482526.

V. CONCLUSIONS

In this paper, we have shown that the capacity achieving input distribution of a non-coherent uncorrelated Rayleigh fading MIMO channel is discrete with a finite number of mass points, one necessary located at the origin. The channel capacity is computed numerically finding the optimal number of mass points, their probabilities and locations. Optimality is guaranteed when the optimal input distribution satisfies the Kuhn-Tucker condition.

The conjectured capacities in [6] are actually achieved. Furthermore, the simulation is extended to MIMO systems and the capacities are presented against the input power for multiple receivers. Although the capacity is shown numerically at any SNR for any antenna configuration, there is a necessity for a simple and easy way to determine the optimal input and hence the capacity. For instance, results in a tabulated form of the optimal mass point properties and the capacity for a given input power constraint is highly desirable. These results are far away from today's research and it is a challenge to look at this problem in a different angle.

VI. APPENDIX

A. Proof of Lemma 1

The necessary and sufficient condition for an input distribution to be optimal is derived in here proving the Lemma 1 stated with Kuhn-Tucker condition. The following definition is given in [11] for the weak differentiability on a convex space.

Definition 1: Let S be a convex space, L a functional from S into the real line $\mathbb{R}$, x_0 a fixed element of S, and θ a number in $[0, 1]$. Suppose there exists a map $L'_{x_0} : \to \mathbb{R}$ such that

$$L'_{x_0}(x) \triangleq \lim_{\theta \to 0} \frac{L[(1-\theta)x_0 + \theta x] - L(x_0)}{\theta} \quad (22)$$

for all x in S. Then L is said to be weakly differentiable in S at x_0 and L'_{x_0} is the weak derivative in S at x_0. If L is

weakly differentiable in S at x_0 for all x_0 in S, L is said to be weakly differentiable in S or simply weakly differentiable.

The following theorem [10, Page 139] shows the necessary and sufficient condition for a weakly differentiable convex function to have an optimum.

Theorem 1: Suppose Q is weakly differentiable, so that for all x_0, y in its domain S,

$$Q(y) \geq Q(x_0) + Q'(x_0)(y - x_0). \quad (23)$$

Let X denote the feasible set, i.e.

$$X = \{x | Q_i(x) \leq 0, i = 1, ..., m, \ h_i(x) = 0, \ i = 1, ..., p\}, \quad (24)$$

then x_0 is optimal if and only if $x_0 \in X$ and

$$Q'(x_0)(y - x_0) \geq 0 \ \text{for all} \ y \in X. \quad (25)$$

Geometrically, if $Q'(x_0) \neq 0$ it means that $-Q'(x_0)$ defines a supporting hyperplane to the feasible set at x_0.

Using Theorem 1 on weakly differentiable concave functional, we get the following.

Corollary 2: Assume L is weakly differentiable, concave functional on a convex set S, If L achieves its maximum on S at x_0, then a necessary and sufficient condition for $L(x_0) = max_{x \in S} L(x)$ is that $L'_{x_0} \leq 0$ for all x in S.

The following shows the Lagrangian theorem [10, Page 215-218] commonly being used to find optimal solutions in both convex and non convex functions. Lagrangian theorem always provides a lower bound, and in most cases the optimal solution in the absence of a duality gap.

Theorem 2: Let X be a linear vector space, Z a normed space, Ω a convex subset of X, and P the positive cone in Z. Assume that P contains an interior point. Let f be a real valued concave functional on Ω and G a convex mapping from Ω into Z. Assume the existence of a point $x_1 \in \Omega$ for which $G(x_1) < 0$. Let

$$\mu_0 = \sup_{\substack{x \in \Omega \\ G(x) < 0}} f(x) \quad (26)$$

and assume μ_0 is finite. Then there is an element $Z_0^* > 0$ in Z (the dual space of Z) such that

$$\mu_0 = \sup_{x \in \Omega} \left\{ f(x) - \langle G(x), Z_0^* \rangle \right\}. \quad (27)$$

Furthermore, if the supremum is achieved in (26) by an $x_0 \in \Omega$, $G(x_0) \leq 0$, it is achieved by x_0 in (27) and $\langle G(x), Z_0^* \rangle = 0$.

Using the theorem 2, we can pose the optimisation problem for channel capacity with $\lambda \geq 0$

$$C = \sup_{\substack{G_X \in \mathcal{G} \\ E\{|x|^2\} \leq P_{av}}} I(G_X) \quad (28)$$

$$= \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \phi(G_X) \quad (29)$$

$$= \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \left(\int_0^\infty x^2 dG_X(x) - P_{av} \right). \quad (30)$$

Note in here that all the conditions of the Lagrangian theorem are satisfied. The set of input distributions of nonnegative random variables forms a convex set, the mutual information is a concave function of the input distribution [6, Appendix I-B], and input power constraint is convex since it is a linear functional of the input distribution. Next we will show that both mutual information $I(\cdot)$ and the input constraint $\phi(\cdot)$ are weakly differentiable functions.

Lemma 3: The mutual information $I(\cdot)$ defined in (3) and $\phi(\cdot)$ defined in (30) are weakly differentiable functionals on $\mathcal{G}$ with weak derivatives

$$I'_{G_0}(G_X) = -\int_0^\infty p_Y(y;G_X)\log p_Y(y;G_0)dy$$
$$-\frac{1}{2}\int_0^\infty \log(1+x^2)dG_X(x) - I(G_0) + V(n_r)$$

$$(31)$$

and

$$\phi'_{G_0}(G_X) = \phi(G_X) - \phi(G_0). \qquad (32)$$

Proof: We define

$$G_\theta = (1-\theta)G_0 + \theta G, \quad \theta \in (0,1) \qquad (33)$$

where G_0 is the optimal input distribution. The difference in mutual information obtained with two distributions G_θ, G_0 is given by

$$I(G_\theta) - I(G_0) = \int_0^\infty p_Y(y;G_0)\log p_Y(y;G_0)dy$$
$$- \int_0^\infty p_Y(y;G_\theta)\log p_Y(y;G_\theta)dy$$
$$+ \frac{1}{2}\left\{ \int_0^\infty \log(1+x^2)dG_0(x) - \int_0^\infty \log(1+x^2)dG_\theta(x) \right\}.$$

$$(34)$$

Using

$$p_Y(y;G_\theta) = (1-\theta)p_Y(y;G_0) + \theta p_Y(y;G)$$

and $dG_\theta = (1-\theta)dG_0 + \theta dG$, we get

$$\lim_{\theta \to 0}\left[\frac{I(G_\theta) - I(G_0)}{\theta}\right] = \int_0^\infty p_Y(y;G_0)\log p_Y(y;G_0)dy$$
$$- \int_0^\infty p_Y(y;G)\log p_Y(y;G_0)dy$$
$$\frac{1}{2}\left\{ \int_0^\infty \log(1+x^2)dG_0(x) \right.$$
$$\left. - \int_0^\infty \log(1+x^2)dG(x) \right\} \qquad (35)$$

Also note that

$$I(G_0) = -\int_0^\infty p_Y(y;G_0)\log p_Y(y;G_0)dy$$
$$-\frac{1}{2}\int_0^\infty \log(1+x^2)dG_0(x) + V(n_r). \qquad (36)$$

Substituting $I(G_0)$ in (36) into (35), we get (31). Similarly we can write the first derivative of $\phi(\cdot)$,

$$\lim_{\theta \to 0}\left[\frac{\phi(G_\theta) - \phi(G_0)}{\theta}\right] = \int_0^\infty x^2 dG_X(x) - \int_0^\infty x^2 dG_0(x)$$
$$= \phi(G_X) - \phi(G_0), \qquad (37)$$

proving (32). ∎

Therefore $I(G_X)$ and $\phi(G)_X$ are weakly differentiable functions on $\mathcal{G}$. Using Corollary 2, and the weak differentiability of $I(G_x)$ and $\phi(G)_X$, (29) achieves its maximum if and only if

$$I'_{G_0}(G_X) - \lambda\phi'_{G_0}(G_X) \le 0. \qquad (38)$$

Using the results obtained for $I'_{G_0}(G_X)$ in (31) and $\phi'_{G_0}(G_X)$ in (32) we get the following inequality

$$\int_0^\infty \left\{ \int_0^\infty p_{Y|X}(y|x)\log p_Y(y;G_0)dy \right\} dG_X(x)$$
$$+ \frac{1}{2}\int_0^\infty \log(1+x^2)dG_X(x) + C - V(n_r)$$
$$+ \lambda\int_0^\infty (x^2 - P_{av})dG_X(x) \ge 0 \quad \forall G \in \mathcal{G} \qquad (39)$$

in order to have a optimal point. The following theorem is given in [6].

Theorem 3: Let E_0 be the points of increase of a distribution function F_0. Then

$$\int [I(x;F_0) - \lambda x^2]dF_X(x) \le C - \lambda P_{av} \qquad (40)$$

for all $F \in \mathcal{F}$ if and only if

$$I(x;F_0) \le C + \lambda(x^2 - P_{av}), \quad \forall x, \qquad (41)$$

and

$$I(x;F_0) = C + \lambda(x^2 - P_{av}), \quad \forall x \in E_0. \qquad (42)$$

Using Theorem 3 in (39), we obtain the necessary and sufficient condition (6) for the function (30) to have a maximum. This is known as the Kuhn-Tucker condition.

VII. ACKNOWLEDGEMENT

REFERENCES

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. on Info. Theory*, vol. 45, no. 01, pp. 139–157, Jan. 1999.

[2] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Info. Theory*, vol. 48, no. 02, pp. 359–383, Feb. 2002.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 06, pp. 585–595, Nov. 1999.

[4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 06, pp. 311–335, Mar. 1998.

[5] S. Ray, M. Medard, L. Zheng, and J. Abounadi, "On the sublinear behavior of MIMO channel capacity at low SNR," *IEEE Int. Symp. on Info. Theory*, pp. 1031–1034, Parma, Italy, Oct. 2004.

[6] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. on Info. Theory*, vol. 47, no. 04, pp. 1290–1301, May 2001.

[7] R. R. Perera, T. S. Pollock, and T. D. Abhayapala, "Upper bound on non-coherent MIMO channel capacity in Rayleigh fading," *in Proc. 11th Asia-Pacific Conference on Communications, APCC 2005*, pp. 72–76, Perth, Australia, Oct. 3-5, 2005.

[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Inc, New York, 1965.

[9] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic press San Diego, USA, sixth edition, 2000.

[10] S. Boyd and L. Vandenberghe, *Convex Optimisation*, Cambridge University Press, United Kingdom, 2004.

[11] J. G. Smith, "The information capacity of peak and average power constrained Gaussian channels," *Inform. Contr.*, vol. 18, no. 107, pp. 203–219, 1971.