

Capacity of non-coherent Rayleigh fading MIMO channels

R.R. Perera, K. Nguyen, T.S. Pollock and T.D. Abhayapala

Abstract: The capacity of discrete time uncorrelated Rayleigh fading multiple input multiple output (MIMO) channels was investigated without channel state information (CSI) at either the transmitter or the receiver. 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. It is shown 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, it is shown that at low SNR, the capacity with two mass points is optimal. As the number of receiver antennas increases, the maximum SNR at which two mass points are optimal decreases. Using this result it is argued that on–off keying is optimal in non-coherent Rayleigh fading MIMO channels at low SNR.

1 Introduction

In wireless communications, the knowledge of channel state information (CSI) at either the receiver or at the transmitter, or at both, is considered to be a vital part in information transfer across the channel. It appears from the results of [1] that it is possible to obtain accurate CSI using pilots in slowly changing channels. It is also beneficial if the statistics of the channels are known. However, the statistics of wireless channels are highly invariant and finding a general model which holds in all scenarios seems to be a very difficult, if not impossible, task. Therefore, there are scenarios or applications where coherent detection is not plausible.

Consider the case of a mobile receiver, with channel variation due to the surroundings and movement of the receiver. Here, the time between the independent fades may be too short to permit reliable estimation of the fading coefficient. In such a situation, the channel becomes non-coherent with no knowledge of channel.

The capacity achieving input distribution of non-coherent Rayleigh fading MIMO channels has been an open problem for some time. Early work of [2] 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. It is shown in [2] 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.

The work in [2] was extended in [3] 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 [4, 5]. 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 [6], 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 signalling is optimal. However, no proof was given. The SISO non-coherent Rayleigh fading channel is extensively studied in [7] showing the optimal input is discrete with a finite number of mass points; (defined as distinct points with nonzero values). Capacity is computed numerically choosing the optimal number of mass points, their probabilities and locations.

In this paper, we prove for the first time, that the capacity achieving input distribution (i.e. the amplitude of the multiple inputs) of a non-coherent Rayleigh fading MIMO channel is discrete with a finite number of mass points, one necessarily located at the origin. The numerical simulation work in optimising the channel capacity is described in detail, extending the work presented in [7] for a single antenna system to multiple antennas. 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.

2 Channel model

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

$$Y = HX + N \quad (1)$$

where the output $\mathbf{Y}$ is $n_r \times 1$ and the channel gain matrix $\mathbf{H}$ is $n_r \times n_t$. The input $\mathbf{X}$ is $n_t \times 1$ and the noise $\mathbf{N}$, which is assumed to be zero mean complex Gaussian, is $n_r \times 1$. Each element of $\mathbf{H}$, h_{ij} , $i = 1, \dots, n_r$, $j = 1, \dots, n_t$ is assumed to be zero mean circular complex Gaussian random variables with a unit variance in each dimension, where n_t and n_r denote the number of transmit and receive antennas, respectively. We use $X = |\mathbf{X}|$ and $Y = |\mathbf{Y}|$ to 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 $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ to indicate the Gamma and Psi functions. The differential entropy of $\mathbf{x} \in \mathbf{X}$ is denoted by $h(\mathbf{X})$ and the mutual information between $\mathbf{x}$ and $\mathbf{y} \in \mathbf{Y}$ is designated by $I(\mathbf{X}; \mathbf{Y})$. All the differential entropies and the mutual information are defined to the base 'e', and 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.

3 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 co-ordinate transformation is applied on $2n_r$ dimensions [8].

3.1 Mutual information

Using the pdf of the channel output given input (2), the mutual information between the input and output of the channel model (1) can be written as

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

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 [8]. The channel capacity is the supremum of (3) over the set of all input distributions satisfying the input power constraint $\int x^2 p_X(x) dx \leq P_{av}$

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

where $I(G_X) \triangleq I(\mathbf{X}, \mathbf{Y})$ and $p_Y(y; G_X) = \int_0^\infty p_{Y|X}(y|x) \times (y|x) dG_X(x)$ 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) 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 [7].

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 constraint P_{av} , G_0 is the capacity achieving input amplitude distribution if and only if there exist, λ such that the following is satisfied $\forall x \geq 0$

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

with equality if $x \in E_0$, 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 behaviour.

Proof: See the Appendix.

3.2 Input distribution

We adopt the same principle in proving the discrete character of the optimal input X^* given in [7] for a single antenna system. Therefore, X^* should possess one of the following properties:

- (i) the support set contains an interval
- (ii) it is discrete, with an infinite number of mass points on some bounded interval
- (iii) it is discrete and infinite, but with only a finite number of mass points on any bounded interval, or
- (iv) it is discrete with a finite number of mass points.

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

Let us assume (1) and (2) hold, and define $u = (1+x^2)$, $z = y^2/2$. The support set U_X has an infinite number of distinct points and the Kuhn–Tucker condition holds with equality for all real $u \in (0, 1]$ [7]. In this case, using the equality in (6) we write

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

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

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

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 transformation with the solutions [9, pp. 1020–1030]

$$L^{-1} \left[\frac{n!}{s^{n+1}} \right] = t^n$$

and

$$L^{-1} \left[\frac{\log s}{s^n} \right] = \frac{t^{n-1}}{\Gamma(n)} (\Psi(n) - \log t) \quad n > 0 \quad (9)$$

we obtain the output pdf

$$p_Y(y) = \frac{\sqrt{2}}{y} M e^{-\frac{\lambda y^2}{2n_r}} \quad (10)$$

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 (10) cannot be a valid pdf, negating our original assumptions (1) and (2).

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

$$\begin{aligned} p_Y(y) &= \sum_{i=0}^{\infty} p_i \left\{ \frac{y^{2n_r-1} u_i^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r-1} \Gamma(n_r)} \right\} \\ &> p_i(y) = p_i \left\{ \frac{y^{2n_r-1} u_i^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r-1} \Gamma(n_r)} \right\}, \\ &\forall y \geq 0, i = 1, 2, \dots \end{aligned} \quad (11)$$

Using the property of logarithmic function $\log t_1 > \log t_2$, for all $t_1 > t_2 > 0$, we can pose the following inequality:

$$\begin{aligned} \int_0^{\infty} p_{Y|U}(y|u) \log[p_Y(y)] &> \int_0^{\infty} p_{Y|U}(y|u) \\ &\times \log \left[\frac{y^{2n_r-1} u_i^{n_r}}{2^{n_r-1} \Gamma(n_r)} \exp\left(-\frac{u_i y^2}{2}\right) \right] dy \end{aligned} \quad (12)$$

With the integral solutions [10, pp. 260–265]

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

and

$$\begin{aligned} \int_0^{\infty} x^t e^{-u_1 x^2} \log[v_1 x^t] dx &= \frac{1}{2} u_1^{-\frac{1+t}{2}} \log v_1 \Gamma\left(\frac{1+t}{2}\right) \\ &+ \frac{u_1^{-\frac{1}{2}}}{4\sqrt{u_1}} \Gamma\left(\frac{1+t}{2}\right) \left\{ \Psi\left(\frac{1+t}{2}\right) - \log u_1 \right\} \end{aligned} \quad (13)$$

(12) can be simplified to

$$\begin{aligned} \int_0^{\infty} p_{Y|U}(y|u) \log[p_Y(y)] &> \left(n_r - \frac{1}{2} \right) \Psi(n_r) - n_r \\ &+ \log \left[\frac{\sqrt{2} p_i u_i^{n_r}}{2^{n_r-1} \Gamma(n_r)} \right] - n_r \left(\frac{u_i}{u} \right) \end{aligned} \quad (14)$$

The result in (14) can be used to derive the following bound on LHS of (6):

$$\begin{aligned} \text{LHS} &\geq \frac{(\lambda - n_r u_i)}{u} - \lambda(1 + P_{av}) + C - V(n_r) \\ &+ \left(n_r - \frac{1}{2} \right) \Psi(n_r) - n_r + \log \left[\frac{\sqrt{2} p_i}{\Gamma(n_r)} \left(\frac{u_i}{u} \right)^{n_r} \right] \end{aligned} \quad (15)$$

$$= \frac{(\lambda - n_r u_i)}{u} + O\left(\frac{1}{u}\right) \quad (16)$$

This lower bound diverges to ∞ for $\lambda > n_r u_i$ when $u \rightarrow 0$, 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 $u \rightarrow 0$. Therefore, the only possibility is that $\lambda = 0$ if the input X^* is discrete with infinite mass points. However, the Kuhn–Tucker theorem [11] for convex functions (the mutual information

and hence the channel capacity is concave [7]) 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. A question arises, where the mass points are located and their probabilities. Locating the mass points analytically would be very difficult and even hard using numerical methods. The existence of a mass point at zero is proven for a single-antenna system, as well as for a diversity receiver [7].

3.3 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 a finite number of mass points, we use the distribution function

$$G_X^*(x) = \sum_{i=1}^N p_i \delta(x - x_i) \quad (17)$$

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

$$I(\mathbf{X}; \mathbf{Y}) = \sum_{i=0}^N p_i \int_0^{\infty} p_{Y|X}(y|x_i) \log \left[\frac{p_{Y|X}(x|x_i)}{\sum_{j=1}^N p_j p_{Y|X}(x|x_j)} \right] dy \quad (18)$$

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

$$\frac{\partial I(\mathbf{X}; \mathbf{Z})}{\partial x_0} = p_0 \int_0^{\infty} \frac{\partial}{\partial x_0} p_{Z|X}(z|x_0) \log \left[\frac{p_{Z|X}(z|x_0)}{\sum_{j=1}^N p_j p_{Z|X}(z|x_j)} \right] dz \quad (19)$$

where

$$\frac{\partial}{\partial x_0} p_{Z|X}(z|x_0) = \frac{2x_0}{(1+x_0^2)^2} [z - n_r(1+x_0^2)] p_{Z|X}(z|x_0) \quad (20)$$

Let us define

$$J(z) \triangleq \log \left[\frac{p_{Z|X}(z|x_0)}{\sum_j p_j p_{Z|X}(z|x_j)} \right] \quad (21)$$

then (19) becomes

$$\frac{\partial I(\mathbf{X}; \mathbf{Z})}{\partial x_0} = \frac{2x_0 p_0}{(1+x_0^2)} \int_0^{\infty} [z - n_r(1+x_0^2)] p_{Z|X}(z|x_0) J(z) dz \quad (22)$$

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

Corollary 1: The function $J(z)$ in (21) is a decreasing function for $0 \leq x_0 < x_1 < \dots < x_N$.

Proof: The ratio

$$\frac{p_Z(z)}{p_{Z|X}(z|x_0)} = p_0 + \sum_{i=1}^N \frac{(1+x_0^2)^{n_r}}{(1+x_i^2)^{n_r}} \exp \left[z \left(\frac{1}{1+x_0^2} - \frac{1}{1+x_i^2} \right) \right] \quad (23)$$

is an increasing function since

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

for $0 \leq x_0 < x_1 < \dots < x_N$. Therefore, $J(z)$ in (21) is a decreasing function due to logarithm of the reciprocal of ratio given in (23).

Using Corollary 1 with Lemma 1 in [7], we conclude that the derivative is negative with respect to x_0 for $0 \leq x_0 < x_1 < \dots < x_N$. Therefore, the X^* with $x_0 \geq 0$ cannot produce a local maximum, and hence the input distribution $G_X^*(x)$ necessarily has a mass point located at the origin.

4 Numerical results and simulation

4.1 Numerical results

The optimal amplitude input, discrete with a finite number mass points can be used to compute the MIMO channel capacity numerically. 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.

Figure 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. Figure 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. The same is shown in [7] for SISO Rayleigh fading channels.

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, a zero mass point dominates with a high probability. Figures 4 and 5 depict the Kuhn–Tucker condition (6) for $P_{av} = \{1.4, 2.2\}$ with a single receiver. As claimed in Lemma 1, it is above zero except for the optimal mass point locations where the Kuhn–Tucker condition equals zero.

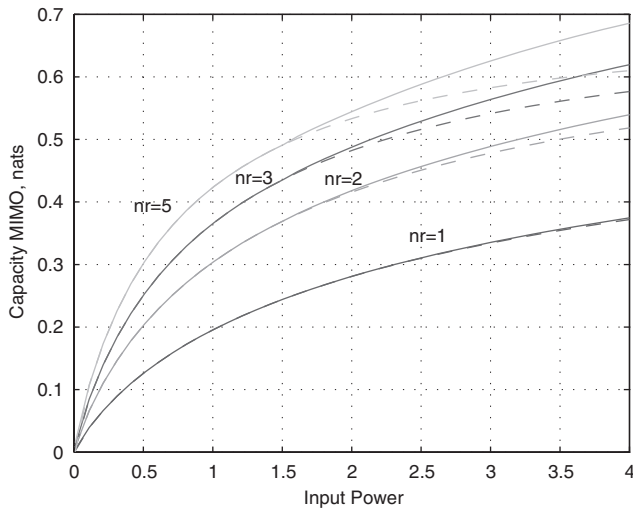


Fig. 1 Capacity of noncoherent Rayleigh fading MIMO channel against input power using two and three mass points for a different number of receiver antennas $n_r = \{1, 2, 3, 5\}$

The dashed lines show the capacity having two mass points for each receiver configuration

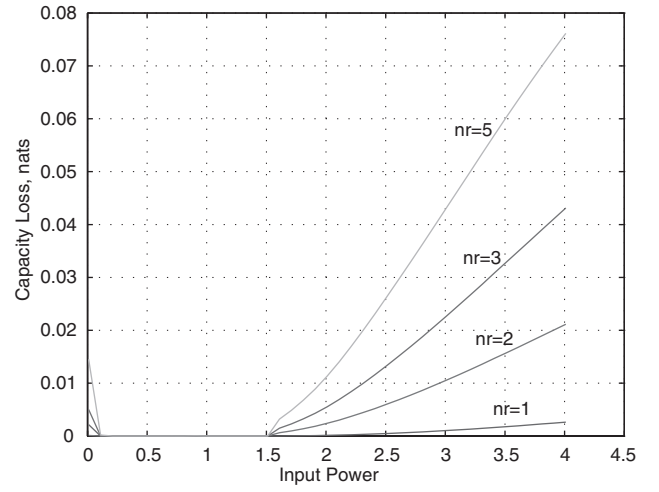


Fig. 2 Loss in channel capacity with two mass points against three against input power for $n_r = \{1, 2, 3, 5\}$

High values shown at very low SNR are due to oscillation of optimal mass points and should be ignored

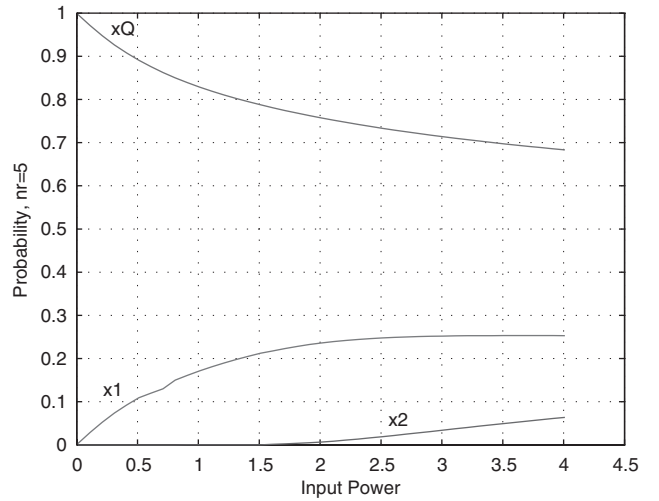


Fig. 3 Probability distribution of optimal mass points against input power for $n_r = 5$

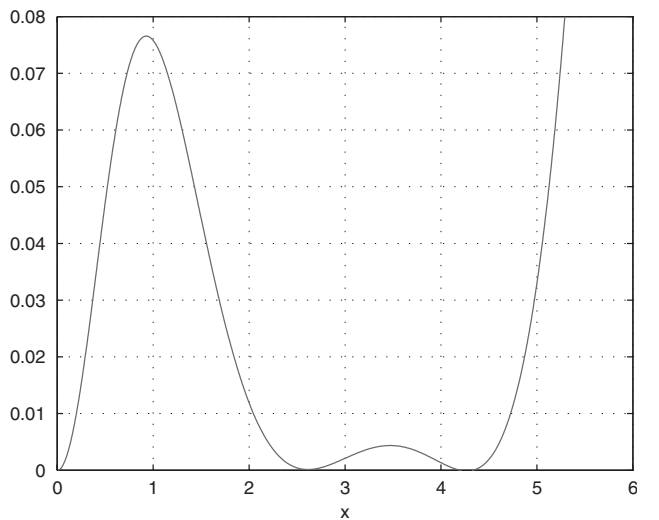


Fig. 4 Kuhn–Tucker condition (6) for $P_{av} = 1.4$, $\lambda = 0.08819532394409$ and $p_X(x) = 0.79549244782344\delta(x) + 0.20444300113415\delta(x - 2.6160) + 0.0000645510564\delta(x - 4.1451)$ Channel capacity $C = 0.23549289482526$

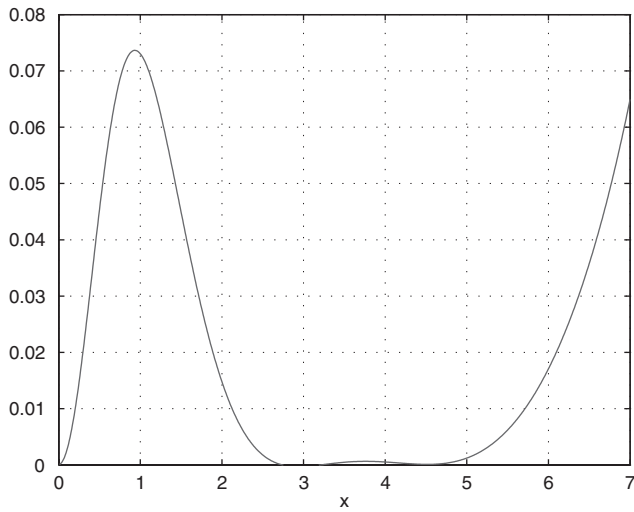


Fig. 5 The Kuhn-Tucker condition (6) for $P_{av}=2.6$, $\lambda=0.0158$ and $p_X(x)=0.7294457069010\delta(x)+0.24599313181750\delta(x-2.9278)+0.0245611664924\delta(x-4.4728)$
Channel capacity $C=0.29355664132426$

4.2 Simulation

4.2.1 Capacity curve: Having analytically derived the mutual information for the MIMO case, we can numerically compute the capacity of a Rayleigh fading MIMO channel as a function of input power. In [7], the results are given only for the single-antenna case. In this paper, we extend the computation for MIMO systems using the theory derived.

We introduce the Lagrange multiplier and find the optimum solution for the function $f=I(G_X)-\lambda\phi(G_X)$ over the whole range of input X^* where $\phi(\cdot)$ represents the input power constraint. We apply the Gauss-Laguerre quadrature method to estimate all necessary integrals in f . To maximise it over the whole range of X , we used steepest descent method [11] to find the minimum solution for function $-f$. To guarantee the convergence, we limit the range between zero and $C'(0)$, which equals one if the input noise variance is normalised to unity, ensuring that the algorithm will not try to increase indefinitely. Since the optimal input is discrete with a finite number of mass points, to find a solution we start the program with a random vector $[x_1, x_2, \dots, x_N, p_1, p_2, \dots, p_N]^T$ subject to the constraints $x_1 < x_2 < \dots < x_N$, $0 < p_i$ for all i , $\sum_{i=1}^{N-1} p_i = 1$, and keep λ fixed. For each value of power constraint P_{av} , $\lambda_1 \in \lambda$ is calculated using the quadrant division method. A projected gradient method was implemented to ensure the input probability is positive and all the boundary constraints are satisfied.

The order of the Gauss-Laguerre quadrature method is critical to achieve the specified accuracy for any given integral approximation. In our simulation, the program was able to converge when the order was set to 85, which provided sufficient accuracy for the exiting condition in the projected gradient method to be satisfied. Note that the program failed to compute if the order was set higher than 150.

4.2.2 Kuhn-Tucker condition: The Kuhn-Tucker condition given in Lemma 1 is a necessary and sufficient condition for optimality (refer to the Appendix for more details).

This condition is used to determine if the local maximum obtained from the steepest decent method is actually the global maximum. To illustrate the optimality condition, the LHS of (6) should be plotted as a function of x . If the condition holds, the graph obtained must be non-negative and must touch zero at the atoms of X^* .

The order of Gauss-Laguerre has a significant effect on the accuracy of Kuhn-Tucker analysis. Figure 6 shows the Kuhn-Tucker test for $P_{av}=1.4$ with sensitivity to the order of Gauss-Laguerre quadrature. The results show that depending on the Gauss-Laguerre order, the function crosses over the horizontal axis with varying degrees of magnitude. As the order increases, the third 'ditch' goes below the zero boundary and vice versa. The third 'ditch' of the function converges to the point on the horizontal axis that corresponds to the third mass point when the order equals 38.

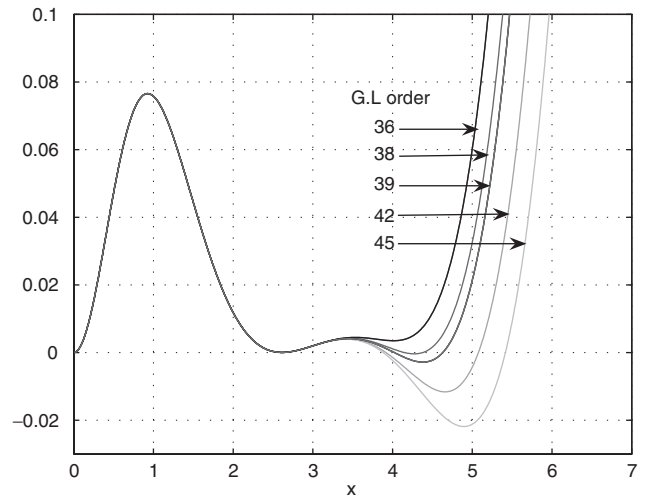


Fig. 6 Kuhn-Tucker condition (6) with sensitivity to Gauss-Laguerre order {36, 38, 39, 42, 45}

It is worth mentioning that the integrity of the Kuhn-Tucker test also depends very much on the magnitude and the corresponding probability of the third mass point. As the probability drops below 10^{-4} , the Kuhn-Tucker condition tends to be unstable. It is claimed in [7] that the dashed lines (projected capacity, probability and locations) are due to the instability of the Kuhn-Tucker condition. We have increased the accuracy of simulation by setting the order of Gauss-Laguerre quadrature to be 120. The results obtained are then tested using the Kuhn-Tucker condition. We claim that the capacity in the dashed line range [7] is actually achievable.

For the MIMO case, in some occasions, the Kuhn-Tucker condition seems to be unstable. The inaccuracy incurred in using the built in $\Gamma(\cdot)$ and $\Psi(\cdot)$ functions in Matlab may have accounted for this problem. In these cases, the Kuhn-Tucker condition for the MIMO case is always non-negative, but the ditches of the Kuhn-Tucker functions do not converge to the horizontal axis at which the mass points are positioned. The converging of the Kuhn-Tucker condition for the MIMO case requires further analysis.

For example, Figs. 7 and 8 show that the optimum input random variable with discrete distribution function can achieve the capacity of an average power-limited channel in the dashed line regions. This is because the resulting Kuhn-Tucker test graphs are non-negative and touch zero at the atoms of X^* . Verifying that the optimum input satisfies the Kuhn-Tucker condition for various values of input power in the dashed line ranges, we conclude that the capacity is actually achievable.

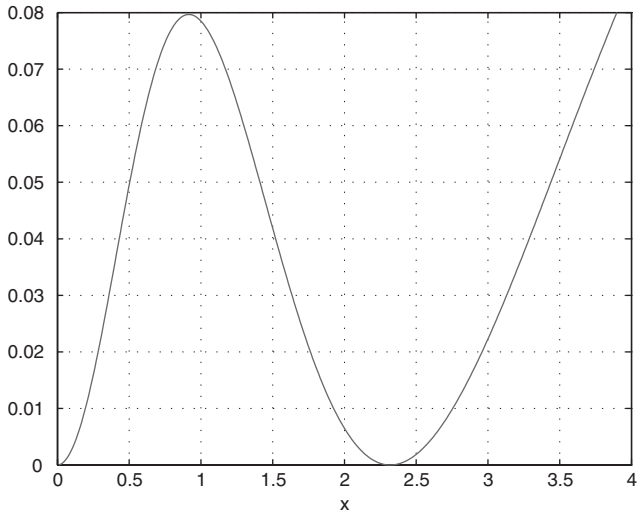


Fig. 7 Kuhn–Tucker condition (6) for $P_{av}=0.8$, $\lambda=0.131734$ and $p_X(x)=0.85118882\delta(x)+0.1488111\delta(x-2.3186)$ Channel capacity $C=0.171$

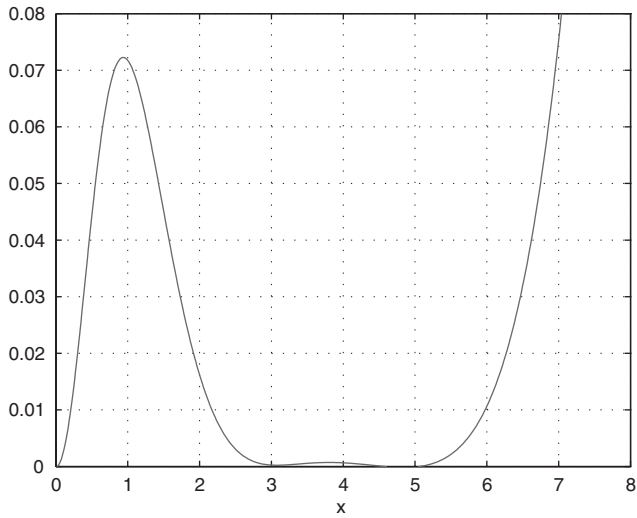


Fig. 8 Kuhn–Tucker condition (6) for $P_{av}=4$, $\lambda=0.03433$ and $p_X(x)=0.68347\delta(x)+0.253523\delta(x-3.07943)+0.063\delta(x-5.03258)$ Channel capacity $C=0.3746$

5 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 necessarily located at the origin. The channel capacity is computed numerically finding the optimal number of mass points, their probabilities and locations. The optimality is guaranteed when the optimal input distribution satisfies the Kuhn–Tucker condition.

The simulation work is briefly highlighted with the improvements made in the single-antenna case compared to the work of Abou-Faycal *et al.* [7]. The conjectured capacities in [7] 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 tabulated form in which the optimal mass point properties and the capacity are readily available for a given input power constraint.

6 Acknowledgments

National ICT Australia (NICTA) is funded through the Australian Government's Backing Australia's Ability Initiative, in part through the Australian Research Council.

7 References

- 1 Tarokh, V., Naguib, A., Seshadri, N., and Calderbank, A.R.: 'Space-time codes for high data rates wireless communications: Performance criteria in the presence of channel estimation errors, mobility and multiple paths', *IEEE Trans. Commun.*, 1999, **47**, pp. 199–207
- 2 Marzetta, T.L., and Hochwald, B.M.: 'Capacity of a mobile multiple-antenna communication link in Rayleigh at fading', *IEEE Trans. Inf. Theory*, 1999, **45**, (1), pp. 139–157
- 3 Zheng, L., and Tse, D.N.C.: 'Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel', *IEEE Trans. Inf. Theory*, 2002, **48**, (2), pp. 359–383
- 4 Telatar, I.E.: 'Capacity of multi-antenna Gaussian channels', *European Trans. Telecommun.*, 1999, **10**, (6), pp. 585–595
- 5 Foschini, G., and Gans, M.: 'On limits of wireless communications in a fading environment when using multiple antennas', *Wireless Pers. Commun.*, 1998, **6**, pp. 311–335
- 6 Ray, S., Medard, M., Zheng, L., and Aounadi, J.: 'On the sublinear behavior of MIMO channel capacity at low SNR'. *IEEE Int. Symp. Inf. Theory*, October 2004, pp. 1031–1034
- 7 Abou-Faycal, I.C., Trott, M.D., and Shamai, S.: 'The capacity of discrete time memoryless Rayleigh fading channels', *IEEE Trans. Inf. Theory*, 2001, **47**, (4), pp. 1290–1301
- 8 Perera, R.R., Pollock, T.S., and Abhayapala, T.D.: 'Upper bound on noncoherent MIMO channel capacity in Rayleigh fading'. *Proc. 11th Asia-Pacific Conf. on Communications, APCC 2005*, Perth, Australia, 3–5 Oct. 2005, pp. 72–76
- 9 Abramowitz, M., and Stegun, I.A.: 'Handbook of mathematical functions' (Dover Publications, Inc., USA, 1965)
- 10 Gradshteyn, I.S., and Ryzhik, I.M.: 'Table of integrals, series, and products' (Academic Press, San Diego, USA, 2000, 6th edn.)
- 11 Boyd, S., and Vandenberghe, L.: 'Convex optimisation' (Cambridge University Press, UK, 2004)
- 12 Smith, J.G.: 'The information capacity of peak and average power constrained Gaussian channels', *Inf. Control.*, 1971, **18**, (3), pp. 203–219

8 Appendix: Proof of Lemma 1

The necessary and sufficient condition for an input distribution to be optimal is derived here proving Lemma 1 stated with the Kuhn–Tucker condition. The following definition is given in [12] 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 $L'_{x_0} : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$L'_{x_0}(x) \triangleq \lim_{\theta \rightarrow 0} \frac{L[(1-\theta)x_0 + \theta x] - L(x_0)}{\theta} \quad (24)$$

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 [11, p. 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

$$Q(y) \geq Q(x_0) + Q'(x_0)(y - x_0) \quad (25)$$

Let X denote the feasible set, i.e.

$$X = \{x | Q_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\} \quad (26)$$

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

$$Q'(x_0)(y - x_0) \geq 0 \quad \text{for all } y \in X \quad (27)$$

Geometrically, if $Q'_0(x_0) \neq 0$ it means that $-Q'(x_0)$ defines a supporting hyperplane to the feasible set at x_0 .

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

Corollary 2: Assume L is a 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 [11, pp. 215–218] commonly being used to find optimal solutions in both convex and nonconvex functions. The Lagrangian theorem is valuable since it 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

$$\mu_0 = \sup_{\substack{x \in \Omega \\ G(x) < 0}} f(x) \quad (28)$$

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

$$\mu_0 = \sup_{x \in \Omega} \{f(x) \langle G(x), Z_0^* \rangle\} \quad (29)$$

Furthermore, if the supremum is achieved in (28) by an $x \in \Omega$, $G(x_0) \leq 0$, it is achieved by x_0 in (29) and $\langle G(x), Z_0^* \rangle = 0$.

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

$$C = \sup_{\substack{G_X \in \mathcal{G} \\ E\{|x|\}^2 \leq P_{av}}} I(G_X) \quad (30)$$

$$= \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \phi(G_X) \quad (31)$$

$$= \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \left(\int_0^\infty x^2 dG_X(x) - P_{av} \right) \quad (32)$$

Note in here that all the conditions of the Lagrangian theorem are satisfied. The set of input distributions of non-negative random variables forms a convex set, the mutual information is a concave function of the input distribution [7, 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 (32) are weakly differentiable functionals on $\mathcal{G}$ with weak derivatives

$$\begin{aligned} I'_{G_0}(G_X) &= - \int_0^\infty p_Y(y; G_X) \log p_Y(y; G_0) dy \\ &\quad - \frac{1}{2} \int_0^\infty \log(1+x^2) dG_X(x) - I(G_0) + V(n_r) \end{aligned} \quad (33)$$

and

$$\phi'_{G_0}(G_X) = \phi(G_X) - \phi(G_0) \quad (34)$$

Proof: We define

$$G_\theta = (1 - \theta)G_0 + \theta G_X, \quad \theta \in (0, 1) \quad (35)$$

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

$$\begin{aligned} I(G_\theta) - I(G_0) &= \int_0^\infty p_Y(y; G_0) \log p_Y(y; G_0) dy \\ &\quad - \int_0^\infty p_Y(y; G_\theta) \log p_Y(y; G_\theta) dy \\ &\quad + \frac{1}{2} \left\{ \int_0^\infty \log(1+x^2) dG_0(x) - \int_0^\infty \log(1+x^2) dG_\theta(x) \right\} \end{aligned} \quad (36)$$

Using

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

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

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \left[\frac{I(G_\theta) - I(G_0)}{\theta} \right] &= \int_0^\infty p_Y(y; G_X) \log p_Y(y; G_0) dy \\ &\quad - \int_0^\infty p_Y(y; G_X) \log p_Y(y; G_X) dy \\ &\quad + \frac{1}{2} \left\{ \int_0^\infty \log(1+x^2) dG_0(x) - \int_0^\infty \log(1+x^2) dG_X(x) \right\} \end{aligned} \quad (37)$$

Also note that

$$\begin{aligned} I(G_0) &= - \int_0^\infty p_Y(y; G_0) \log p_Y(y; G_0) dy \\ &\quad - \frac{1}{2} \int_0^\infty \log(1+x^2) dG_0(x) + V(n_r) \end{aligned} \quad (38)$$

Substituting $I(G_0)$ in (38) into (37), we get (33). Similarly we can write the first derivative of $\phi(\cdot)$

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \left[\frac{\phi(G_\theta) - \phi(G_0)}{\theta} \right] &= \int_0^\infty x^2 dG_X(x) - \int_0^\infty x^2 dG_0(x) \\ &= \phi(G_X) - \phi(G_0) \end{aligned} \quad (39)$$

proving (34). 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 $\phi(G_X)$, (31) achieves its maximum if and only if

$$I'_{G_0}(G_X) - \lambda \phi'_{G_0}(G_X) \leq 0 \quad (40)$$

Using the results obtained for $I'_{G_0}(G_X)$ in (33) and $\phi'_{G_0}(G_X)$ (34) we get the following inequality:

$$\begin{aligned} &\int_0^\infty \left\{ \int_0^\infty p_{Y|X}(y|x) \log p_Y(y; G_0) dy \right\} dG_X(x) \\ &\quad + \frac{1}{2} \int_0^\infty \log(1+x^2) dG_X(x) + C - V(n_r) \\ &\quad + \lambda \int_0^\infty (x^2 - P_{av}) dG_X(x) \geq 0 \quad \forall G_X \in \mathcal{G} \end{aligned} \quad (41)$$

in order to have an optimal point. The following theorem is given in [7].

Theorem 3: Let E_0 be the points of increase of a distribution function G_0 . Then

$$\int [I(x; G_0) - \lambda x^2] dG_X(x) \leq C - \lambda P_{av} \quad (42)$$

for all $G_X \in \mathcal{G}$ if and only if

$$I(x; G_0) \leq C + \lambda(x^2 - P_{av}), \quad \forall x \quad (43)$$

and

$$I(x; G_0) = C + \lambda(x^2 - P_{av}), \quad \forall x \in E_0 \quad (44)$$

Using Theorem 3 in (41), we obtain the necessary and sufficient condition (6) for the function (32) to have a maximum. This is known as the Kuhn–Tucker condition.