Derivation of MIMO Log-Det Formula for Channel Capacity with Real Input / Output Signals

H.L. Rappaport

November 5, 2014

Consider a MIMO (multiple input multiple output) system of N_t transmitters and N_r receivers and assume

$$\mathbf{R} = \mathbf{H}_0\mathbf{S} + \mathbf{Z}, \tag{1}$$

where $\mathbf{R}$, $\mathbf{S}$ and $\mathbf{Z}$ are column vectors of real random variables representing received signals, transmitted signals and noise. Thus $\mathbf{R}$ and $\mathbf{Z}$ are column vectors with N_r elements and $\mathbf{S}$ is a column vector with N_t elements. The quantity $\mathbf{H}_0$ is the $N_r \times N_t$ channel matrix, i.e., $\mathbf{H}_0$ has N_r rows and N_t columns.

The mutual information of this channel is defined

$$I\left(\mathbf{S}, \mathbf{R}\right) = h\left(\mathbf{R}\right) - h\left(\mathbf{R}|\mathbf{S}\right), \tag{2}$$

where $h\left(\mathbf{R}\right)$ is the differential entropy of the received signal vector and $h\left(\mathbf{R}|\mathbf{S}\right)$ is a conditional differential entropy of $\mathbf{R}$ and $\mathbf{S}$ defined as follows

$$h\left(\mathbf{R}\right) = -\int f_{\mathbf{R}}\left(\mathbf{r}\right)\log_2 f_{\mathbf{R}}\left(\mathbf{r}\right)\, d^{N_r}r, \tag{3}$$

$$h\left(\mathbf{R}|\mathbf{S}\right) = -\int f_{\mathbf{SR}}\left(\mathbf{s}, \mathbf{r}\right)\log_2 f_{\mathbf{R}}\left(\mathbf{r}|\mathbf{s}\right)\, d^{N_t}s\, d^{N_r}r. \tag{4}$$

Here $f_{\mathbf{R}}\left(\mathbf{r}\right)$, $f_{\mathbf{SR}}\left(\mathbf{s}, \mathbf{r}\right)$, $f_{\mathbf{R}}\left(\mathbf{r}|\mathbf{s}\right)$ are marginal, joint and conditional probability density functions (p.d.f.s) of received and/or transmitted signals respectively.

Likewise, let $h\left(\mathbf{Z}\right)$ be the differential entropy of the noise defined

$$h\left(\mathbf{Z}\right) = -\int f_{\mathbf{Z}}\left(\mathbf{z}\right)\log_2 f_{\mathbf{Z}}\left(\mathbf{z}\right)\, d^{N_r}z, \tag{5}$$

where $f_{\mathbf{Z}}\left(\mathbf{z}\right)$ is the p.d.f. of the noise.

The capacity of the MIMO channel is given by [1]

$$C_{\mathrm{MIMO}} = \max_{f_{\mathbf{S}}(\mathbf{s})} I\left(\mathbf{S}, \mathbf{R}\right) = \max_{f_{\mathbf{S}}(\mathbf{s})}\left[h\left(\mathbf{R}\right) - h\left(\mathbf{R}|\mathbf{S}\right)\right], \tag{6}$$

where $f_{\mathbf{S}}(\mathbf{s})$ is the p.d.f. of the transmitted signal vector $\mathbf{S}$.

Show $h(\mathbf{R}|\mathbf{S}) = h(\mathbf{Z})$

The first step in evaluating the channel capacity is to assume that the noise is independent of the transmitted signal and show $h(\mathbf{R}|\mathbf{S}) = h(\mathbf{Z})$. To do this, use

$$f_{\mathbf{R}}(\mathbf{r}|\mathbf{s}) = \frac{f_{\mathbf{SR}}(\mathbf{s}, \mathbf{r})}{f_{\mathbf{S}}(\mathbf{s})}, \tag{7}$$

in Eq. (4) to produce

$$h(\mathbf{R}|\mathbf{S}) = -\int\int f_{\mathbf{SR}}(\mathbf{s}, \mathbf{r}) \left[\log_2 f_{\mathbf{SR}}(\mathbf{s}, \mathbf{r}) - \log_2 f_{\mathbf{S}}(\mathbf{s})\right] d^{N_t}s\, d^{N_r}r. \tag{8}$$

Since

$$\mathbf{r} = \mathbf{H}_0\mathbf{s} + \mathbf{z}, \qquad d^{N_r}r = d^{N_r}z, \tag{9}$$

$$h(\mathbf{R}|\mathbf{S}) = -\int\int f_{\mathbf{SR}}(\mathbf{s}, \mathbf{H}_0\mathbf{s} + \mathbf{z}) \left[\log_2 f_{\mathbf{SR}}(\mathbf{s}, \mathbf{H}_0\mathbf{s} + \mathbf{z}) - \log_2 f_{\mathbf{S}}(\mathbf{s})\right] d^{N_t}s\, d^{N_r}z. \tag{10}$$

It turns out that

$$f_{\mathbf{SR}}(\mathbf{s}, \mathbf{H}_0\mathbf{s} + \mathbf{z}) = f_{\mathbf{SZ}}(\mathbf{s}, \mathbf{z}). \tag{11}$$

To show Eq. (11) we turn to the random variable theorem given by Gillespie [2]. Introduce $\mathbf{s}'$ so that

$$\mathbf{s} = \mathbf{s}', \qquad \mathbf{z} = \mathbf{r} - \mathbf{H}_0\mathbf{s}'. \tag{12}$$

Then by the random variable theorem [2]

$$f_{\mathbf{SZ}}(\mathbf{s}, \mathbf{z}) = \int\int f_{\mathbf{SR}}(\mathbf{s}', \mathbf{r})\, \delta(\mathbf{s} - \mathbf{s}')\, \delta\left[\mathbf{z} - (\mathbf{r} - \mathbf{H}_0\mathbf{s}')\right] d^{N_t}s'\, d^{N_r}r, \tag{13}$$

$$= f_{\mathbf{SR}}(\mathbf{s}, \mathbf{H}_0\mathbf{s} + \mathbf{z}), \tag{14}$$

as stated in Eq. (11). Substituting Eq. (11) into Eq. (10) yields

$$h(\mathbf{R}|\mathbf{S}) = -\int\int f_{\mathbf{SZ}}(\mathbf{s}, \mathbf{z}) \left[\log_2 f_{\mathbf{SZ}}(\mathbf{s}, \mathbf{z}) - \log_2 f_{\mathbf{S}}(\mathbf{s})\right] d^{N_t}s\, d^{N_r}z, \tag{15}$$

Since $\mathbf{S}$ and $\mathbf{Z}$ are statistically independent

$$f_{\mathbf{SZ}} = f_{\mathbf{S}}(\mathbf{s})\, f_{\mathbf{Z}}(\mathbf{z}), \tag{16}$$

so that Eq. (15) becomes

$$h(\mathbf{R}|\mathbf{S}) = -\int\int f_{\mathbf{S}}(\mathbf{s})\, f_{\mathbf{Z}}(\mathbf{z}) \log_2 f_{\mathbf{Z}}(\mathbf{z})\; d^{N_t}s\, d^{N_r}z, \tag{17}$$

$$= h(\mathbf{Z}), \tag{18}$$

as indicated above.

Evaluate $h(\mathbf{Z})$

Since the noise vector $\mathbf{Z}$ is assumed to be a random vector with components that are jointly-normal random variables with zero mean the p.d.f. for $\mathbf{Z}$ satisfies [3]

$$f_{\mathbf{Z}}(\mathbf{z})_{=} \frac{1}{\sqrt{(2\pi)^{N_r}\,|\mathbf{V_Z}|}} \exp\left[-\frac{1}{2}\mathbf{z}^T\mathbf{V}_{\mathbf{Z}}^{-1}\mathbf{z}\right], \tag{19}$$

and the differential entropy $h(\mathbf{Z})$ satisfies

$$h(\mathbf{Z}) = \frac{1}{2}\log_2\left[(2\pi e)^{N_r}\,|\mathbf{V_Z}|\right], \tag{20}$$

where $\mathbf{V_Z}$ is the covariance matrix of the noise, e.g.,

$$\mathbf{V_Z} = E\left(\mathbf{Z}\mathbf{Z}^T\right). \tag{21}$$

Further assuming that the components of the noise vector $\mathbf{Z}$ are i.i.d. (independent identically distributed) random variables

$$\mathbf{V_Z} = \sigma_Z^2 \mathbf{I}_{N_r}, \tag{22}$$

where I_{N_r} is the $N_r \times N_r$ identity matrix and $h(\mathbf{Z})$ becomes

$$h(\mathbf{Z}) = \frac{1}{2}\log_2\left[\left(2\pi e\sigma_Z^2\right)^{N_r}\right]. \tag{23}$$

Maximizing Entropy

With Eq. (18) Eq. (6) becomes

$$C_{\mathrm{MIMO}} = \max_{f_{\mathbf{S}}(\mathbf{s})} \left[h\left(\mathbf{R}\right) - h\left(\mathbf{Z}\right)\right] = \max_{f_{\mathbf{S}}(\mathbf{s})} \left[h\left(\mathbf{R}\right)\right] - h\left(\mathbf{Z}\right), \tag{24}$$

since the entropy of $\mathbf{Z}$ does not depend on the probability density function of the input variable $\mathbf{S}$.

In the SISO (single input single output) problem the mutual information was maximized while holding the variance of the output p.d.f. constant [4]. In the MIMO problem, the mutual information is maximized while holding the covariance matrix $\mathbf{V_R}$ of the output signal vector $\mathbf{R}$ constant and setting the mean of $\mathbf{R}$ to zero. The method for finding the maximum is similar to the method outlined in the references [4, 5].

To find the maximum of $h\left(\mathbf{R}\right)$ suppose

$$E\left(R_i\right) = 0, \qquad E\left(R_i R_j\right) = \sigma_{ij}, \qquad 1 \le i, j \le N_r, \tag{25}$$

and let $f_{\mathbf{R}\mathrm{arb}}$ be any zero mean density satisfying

$$\int f_{\mathbf{R}\mathrm{arb}}\left(\mathbf{r}\right) r_i r_j \, d^{N_r} r = \sigma_{ij}, \tag{26}$$

where the σ_{ij} are the elements of the $\mathbf{V_R}$ matrix. Also let $f_{\mathbf{R}\mathrm{norm}}$ be the p.d.f. of a random vector with components that are jointly-normal random variables with zero mean

$$f_{\mathbf{R}\mathrm{norm}} = \frac{1}{\sqrt{\left(2\pi\right)^{N_r} \left|\mathbf{V_R}\right|}} \exp\left[-\frac{1}{2}\mathbf{r}^T \mathbf{V_R}^{-1} \mathbf{r}\right]. \tag{27}$$

The goal is to show that $h\left(\mathbf{R}\right)$ is maximized when the components of $\mathbf{R}$ are jointly normal. To this end the difference

$$h_{\mathrm{norm}}\left(\mathbf{R}\right) - h_{\mathrm{arb}}\left(\mathbf{R}\right) = -\int f_{\mathbf{R}\mathrm{norm}} \log_2 f_{\mathbf{R}\mathrm{norm}} \, d^{N_r} r + \int f_{\mathbf{R}\mathrm{arb}} \log_2 f_{\mathbf{R}\mathrm{arb}} \, d^{N_r} r, \tag{28}$$

is considered. The first step is to simplify this expression with the observation that

$$\int f_{\mathbf{R}\mathrm{arb}} \log_2 f_{\mathbf{R}\mathrm{norm}} \, d^{N_r} r = \int f_{\mathbf{R}\mathrm{norm}} \log_2 f_{\mathbf{R}\mathrm{norm}} \, d^{N_r} r. \tag{29}$$

To see the validity of Eq. (29) consider that

$$-\int f_{\mathbf{R}\mathrm{arb}} \log_2 f_{\mathbf{R}\mathrm{norm}} \, d^{N_r} r = -\frac{1}{\ln 2} \int f_{\mathbf{R}\mathrm{arb}} \left\{-\frac{1}{2} \ln\left[\left(2\pi\right)^{N_r} \left|\mathbf{V_R}\right|\right] - \frac{1}{2}\mathbf{r}^T \mathbf{V_R}^{-1} \mathbf{r}\right\} d^{N_r} r. \tag{30}$$

$$= -\frac{1}{\ln 2}\int f_{\mathbf{R}\text{arb}}\left\{-\frac{1}{2}\ln\left[(2\pi)^{N_r}\left|\mathbf{V_R}\right|\right]-\frac{1}{2}\sum_{i,j=1}^{N_r} r_i\left[\mathbf{V}_\mathbf{R}^{-1}\right]_{ij} r_j\right\} d^{N_r}r, \tag{31}$$

$$= \frac{1}{2}\log_2\left[(2\pi)^{N_r}\left|\mathbf{V_R}\right|\right]+\frac{1}{2\ln 2}\sum_{i,j=1}^{N_r}\left[\mathbf{V_R}\right]_{ij}\left[\mathbf{V}_\mathbf{R}^{-1}\right]_{ij}, \tag{32}$$

$$= \frac{1}{2}\log_2\left[(2\pi)^{N_r}\left|\mathbf{V_R}\right|\right]+\frac{1}{2\ln 2}\sum_{j=1}^{N_r}\delta_{jj}. \tag{33}$$

where the symmetric nature of the $\mathbf{V_R}$ matrix has been used. Continuing

$$-\int f_{\mathbf{R}\text{arb}}\log_2 f_{\mathbf{R}\text{norm}}\, d^{N_r}r = \frac{1}{2}\log_2\left[(2\pi)^{N_r}\left|\mathbf{V_R}\right|\right]+\frac{1}{2}\log_2 e^{N_r}, \tag{34}$$

the same as the result given previously for the entropy of a jointly-normal random vector [3] thus verifying Eq. (29).

Using Eq. (29) in Eq. (28) yields

$$h_{\text{norm}}(\mathbf{R}) - h_{\text{arb}}(\mathbf{R}) = -\int f_{\mathbf{R}\text{arb}}\log_2 f_{\mathbf{R}\text{norm}}\, d^{N_r}r + \int f_{\mathbf{R}\text{arb}}\log_2 f_{\mathbf{R}\text{arb}}\, d^{N_r}r, \tag{35}$$

$$= \int f_{\mathbf{R}\text{arb}}\log_2\frac{f_{\mathbf{R}\text{arb}}}{f_{\mathbf{R}\text{norm}}}\, d^{N_r}r, \tag{36}$$

$$\geq \log_2(e)\int f_{\mathbf{R}\text{arb}}\left(1-\frac{f_{\mathbf{R}\text{norm}}}{f_{\mathbf{R}\text{arb}}}\right) d^{N_t}r = 0, \tag{37}$$

where the inequality and condition for equality in (37) is discussed in the appendix of Rappaport [4].

Using the jointly-normal p.d.f for the maximum indicated in Eq. (24) produces

$$C_{\text{MIMO}} = h_{\text{norm}}(\mathbf{R}) - h(\mathbf{Z}), \tag{38}$$

$$= \frac{1}{2}\log_2\left|\mathbf{V_R}\right| - \frac{1}{2}\log_2\left|\mathbf{V_Z}\right| = \frac{1}{2}\log_2\left|\mathbf{V_R}\right| - \frac{1}{2}\log_2\left|\sigma_Z^2\mathbf{I}_{N_r}\right|. \tag{39}$$

Next, the covariance matrix $\mathbf{V_R}$ of the received signal $\mathbf{R}$ is evaluated in terms of the covariance matrix $\mathbf{V_S}$ of the input signal $\mathbf{S}$ and the covariance matrix $\mathbf{V_Z}$ of the noise $\mathbf{Z}$ using the relationship of Eq. (1). Thus

$$\mathbf{V_R} = E\left(\mathbf{R}\mathbf{R}^T\right) = E\left[\left(\mathbf{H}_0\mathbf{S}+\mathbf{Z}\right)\left(\mathbf{H}_0\mathbf{S}+\mathbf{Z}\right)^T\right], \tag{40}$$

$$= E\left[\mathbf{H}_0\mathbf{S}\left(\mathbf{H}_0\mathbf{S}\right)^T + \mathbf{H}_0\mathbf{S}\mathbf{Z}^T + \mathbf{Z}\left(\mathbf{H}_0\mathbf{S}\right)^T + \mathbf{Z}\mathbf{Z}^T\right], \tag{41}$$

$$= E\left[\mathbf{H}_0\mathbf{S}\mathbf{S}^T\mathbf{H}_0^T + \mathbf{H}_0\mathbf{S}\mathbf{Z}^T + \mathbf{Z}\mathbf{S}^T\mathbf{H}_0^T + \mathbf{Z}\mathbf{Z}^T\right], \tag{42}$$

$$= \mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T + \mathbf{H}_0 E\left(\mathbf{S}\mathbf{Z}^T\right) + E\left(\mathbf{Z}\mathbf{S}^T\right)\mathbf{H}_0^T + \mathbf{V}_\mathbf{Z}, \tag{43}$$

$$\mathbf{V}_\mathbf{R} = \mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T + \mathbf{V}_\mathbf{Z} = \mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T + \sigma_Z^2\mathbf{I}_{N_r}, \tag{44}$$

where two terms have been dropped from Eq. (43) because the noise and the transmitted signal are uncorrelated and the noise is assumed to have zero mean. Using Eq. (44) in Eq. (39) shows

$$C_{\text{MIMO}} = \frac{1}{2}\log_2\left|\mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T + \sigma_Z^2\mathbf{I}_{N_r}\right| - \frac{1}{2}\log_2\left|\sigma_Z^2\mathbf{I}_{N_r}\right|, \tag{45}$$

$$= \frac{1}{2}\log_2\left[\left|\mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T + \sigma_Z^2\mathbf{I}_{N_r}\right|\left|\sigma_Z^2\mathbf{I}_{N_r}\right|^{-1}\right], \tag{46}$$

$$= \frac{1}{2}\log_2\left|\mathbf{I}_{N_r} + \frac{1}{\sigma_Z^2}\mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T\right|, \tag{47}$$

the capacity per channel use or capacity in bits per sample. The capacity per unit time (or bits/second) is therefore [4]

$$C_{t\text{MIMO}} = B\log_2\left|\mathbf{I}_{N_r} + \frac{1}{\sigma_Z^2}\mathbf{H}_0\mathbf{V}_\mathbf{S}\mathbf{H}_0^T\right|, \tag{48}$$

where B is the signal bandwidth and $2B$ is the sampling rate. In the SISO case, this reduces to

$$C_t = B\log_2\left(1 + \frac{\sigma_S^2}{\sigma_Z^2}\right) = B\log_2\left(1 + \frac{P_S}{P_Z}\right), \tag{49}$$

where σ_S^2 is the variance of the input signal and P_S/P_Z is the ratio of the input signal power to noise power. Thus the result for C_t in the MIMO case reduces to C_t for the SISO case [4] as expected.

References

[1] J. R. Hampton, *Introduction to MIMO Communications*, Cambridge University Press, N.Y. (2014).

[2] D. T. Gillespie, "A Theorem for Physicists in the Theory of Random Variables," Am. J. Phys. 51 (6), (1983), p. 520.

[3] H. L. Rappaport, "Multivariate Distributions and Associated Differential Entropy of Jointly-Normal Random Variables," 7G Communications, 7GCTN04, October 6, 2014.

[4] H. L. Rappaport, "Notes on Information Theory II and the Geometric Interpretation of the Shannon Channel Capacity," 7G Communications, 7GCTN02, October 2014.

[5] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," Euro. Trans. Telecommun., 10 (6), (1999), pp. 585 - 595.