9.2.2 MIMO–The multi-dimensional Gaussian channel

The Multiple In, Multiple Out (MIMO) channel referes to a radio communication setup where several transmit and receive antennas are used together. The transmissions over the antennas use the the same communication bandwidth, which mean that all receive antenas get contributions from all transmit antennas, see Figure 9.7. For each path, from one transmit antenna to one receive antenna, there is an attenuation factor, and if all transmission paths can be considered independent there are significant gains compared to the single antenna alternative. In many standards today, for example in WiFi 802.11n and LTE, there are suport for MIMO in the physical link. For the next mobile system, 5G, it will be an essential part of increasing the available bit rates in the system. In this section the scheme will be considered from an information theoretic view and the capacity for the system will be derived.

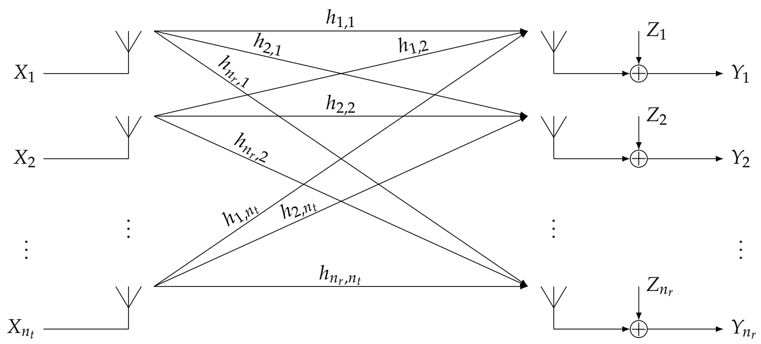

Figure 9.7: The antenna grid in a MIMO setup.

The transmission system shown in Figure 9.7 uses n_t transmit antennas and n_r receive antennas. The attenuation factors between the transmit and received nodes can be given in the $n_r \times n_t$ matrix

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n_t} \\ h_{21} & h_{22} & \cdots & h_{2n_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r1} & h_{n_r2} & \cdots & h_{n_rn_t} \end{pmatrix} \tag{9.51}$$

For simplicity it is assumed that the channel attenuation matrix has full rank, $\text{rank}(H) = \min\{n_t, n_r\}$. Assign the input, noise and the output as the column vectors

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n_t} \end{pmatrix} \quad \boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{n_r} \end{pmatrix} \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n_r} \end{pmatrix} \tag{9.52}$$

where the noise is an n_r-dimensional Gaussian vector, $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \Lambda_Z)$ where $\Lambda_Z = E\big[\boldsymbol{Z}\boldsymbol{Z}^T\big]$ is the covariance matrix. The output is derived as $\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{Z}$ and hence, in this format the MIMO channel constitute a multi-dimensional Gaussian channel with attenuation matrix H. As in the one-dimensional case, the total power used at the transmitter side is the limiting factor of the communication link. Constraining it to P, the sum of the power consummations for the transmitted vector must not exceed this. Since the diagonal elements in Λ_X is $P_i = E\big[X_i^2\big]$, it can be written as $\mathrm{tr}\Lambda_X \leq P$. The notation $\mathrm{tr}A$ denotes the trace of the matrix A, which is the sum of the diagonal elements. Furthermore, it can be shown that the trace also equals the sum of the eigenvalues of the matrix A.

The first step in deriving the channel capacity for the MIMO channel, the mutual information between the input and output vectors can be bounded as

$$\begin{aligned}
I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) \\
&= H(\boldsymbol{Y}) - H(\boldsymbol{Z}) \\
&= H(\boldsymbol{Y}) - \frac{1}{2}\log\big|2\pi e \Lambda_Z\big| \\
&\leq \frac{1}{2}\log\big|2\pi e \Lambda_X\big| - \frac{1}{2}\log\big|2\pi e \Lambda_Z\big| \\
&= \frac{1}{2}\log\big|\Lambda_X \Lambda_Z^{-1}\big|
\end{aligned} \tag{9.53}$$

where it is used in the second equality that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent, in the third equality the differential entropy derived in (8.79) and (8.84), and in the inequality Theorem 8.10. The bound is fulfilled with equality if and only if $\boldsymbol{Y}$ is Gaussian, i.e. $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{0}, \Lambda_Y)$. To achieve this distribution, notice that the sum of two Gaussian variables is again Gaussian. Thus, by letting the input vector be Gaussian, $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{0}, \Lambda_X)$, the output $\boldsymbol{Y}$ will also be Gaussian with covariance matrix

$$\begin{aligned}
\Lambda_Y &= E\big[\boldsymbol{Y}\boldsymbol{Y}^T\big] = E\big[(H\boldsymbol{X}+\boldsymbol{Z})(H\boldsymbol{X}+\boldsymbol{Z})^T\big] \\
&= E\big[(H\boldsymbol{X}+\boldsymbol{Z})(\boldsymbol{X}^T H^T + \boldsymbol{Z}^T)\big] \\
&= HE\big[\boldsymbol{X}\boldsymbol{X}^T\big]H^T + E\big[\boldsymbol{Z}\boldsymbol{Z}^T\big] \\
&= H\Lambda_X H^T + \Lambda_Z
\end{aligned} \tag{9.54}$$

where it is used that since $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent, $E\big[H\boldsymbol{X}\boldsymbol{Z}^T\big] = \boldsymbol{0}$. By maximising the mutual information using a Gaussian distributed input vector gives that the general form of the channel capacity can be written as

$$\begin{aligned}
C &= \max_{f(\boldsymbol{x}), \mathrm{tr}\Lambda_X = P} I(X;Y) \\
&= \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2}\log\big|(H\Lambda_X H^T + \Lambda_Z)\Lambda_Z^{-1}\big| \\
&= \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2}\log\big|I + H\Lambda_X H^T \Lambda_Z^{-1}\big|
\end{aligned} \tag{9.55}$$

Since all transmissions in Figure 9.7 work over the same bandwidth it is reasonable to assume that the noise for the received symbols are identically distributed and independent. Then the covariance matrix becomes a diagonal matrix with the noise variance N, i.e. $\Lambda_Z = NI_{n_r}$, where I_{n_r} is the $n_r \times n_r$ unit matrix. The capacity in (9.55) becomes

$$C = \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2} \log\Big| I + \frac{1}{N} H \Lambda_X H^T \Big| \tag{9.56}$$

To get a better understanding of how the distribution of $\boldsymbol{X}$ can be assign, the channel attenuation matrix can be composed by singular value decomposition, SVD[5],

$$H = USV^T \tag{9.57}$$

where U and V are orthogonal matrices and S a diagonal (in general non-square) matrix with the singular values s_i, $i = 1, \dots, n$ along the diagonal. The fact that U and V are orthogonal means they have unit determinant and that the transpose is the inverse, i.e.

$$|U| = |U^T| = |V| = |V^T| = 1 \tag{9.58}$$

and

$$UU^T = U^T U = I_{n_r} \text{ and } VV^T = V^T V = I_{n_t} \tag{9.59}$$

Inserting (9.57) in (9.56) gives

Thus, the capacity can be rewritten as

$$\begin{aligned} C &= \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2} \log\Big| I + \frac{1}{N} (USV^T) \Lambda_X (USV^T)^T \Big| \\ &= \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2} \log\Big| UU^T + \frac{1}{N} USV^T \Lambda_X V S^T U^T \Big| \\ &= \max_{\mathrm{tr}\Lambda_X = P} \frac{1}{2} \log\Big| I + \frac{1}{N} SV^T \Lambda_X V S^T \Big| \end{aligned} \tag{9.60}$$

Introducing a basis change of the input vector, such that $\tilde{\boldsymbol{X}} = V^T \boldsymbol{X}$, gives the covariance matrix $\Lambda_{\tilde{X}} = E\big[V^T \boldsymbol{X} (V^T \boldsymbol{X})^T\big] = E\big[V^T \boldsymbol{X}\boldsymbol{X}^T V\big] = V^T \Lambda_X V$.

Two matrices A and B, they are said to be *similar* if there exists a nonsingular matrix T such that $A = TBT^T$. From matrix theory it is known that similar matrices have the same set of eigenvalues, and hence the same trace and determinant. This means that Λ_X and $\Lambda_{\tilde{X}}$ are similar and that they have the same trace, $\mathrm{tr}\Lambda_X = \mathrm{tr}\Lambda_{\tilde{X}}$. Thus, the power constraint in the capacity formula can be considered over $\tilde{\boldsymbol{X}}$ instead of $\boldsymbol{X}$, and the capacity becomes

$$C = \max_{\mathrm{tr}\Lambda_{\tilde{X}} = P} \frac{1}{2} \log\Big| I + \frac{1}{N} S \Lambda_{\tilde{X}} S^T \Big| \tag{9.61}$$

[5] In MATLAB the command `[U,S,V]=svd(A)` gives the singular value decomposition.

Since $\Lambda_{\tilde{X}}$ is a covariance matrix it is positive semi-definite, i.e. $\boldsymbol{a}^T \Lambda_{\tilde{X}} \boldsymbol{a} \geq 0$ for all vectors $\boldsymbol{a}$. From $\boldsymbol{a}^T S \Lambda_{\tilde{X}} S^T \boldsymbol{a} = \tilde{\boldsymbol{a}}^T \Lambda_{\tilde{X}} \tilde{\boldsymbol{a}} \geq 0$, where $\tilde{\boldsymbol{a}} = S^T \boldsymbol{a}$, it is seen that also $S \Lambda_{\tilde{X}} S^T$ is positive semi-definite.

The Hadamard inequality states that if a matrix A is positive semi-definite, then the determinant is bounded by the product of the diagonal entries, $|A| \leq \prod_i a_{ii}$. Clearly there is equality if A is a diagonal matrix. Since both I and S in the argument $I + \frac{1}{N} S \Lambda_{\tilde{X}} S^T$ are diagonal, the capacity can obtained as

$$C = \max_{\sum_i \tilde{P}_i = P} \frac{1}{2} \log \prod_i \Big(1 + \frac{s_i^2}{N} \tilde{P}_i\Big) = \max_{\sum_i \tilde{P}_i = P} \sum_i \frac{1}{2} \log \Big(1 + \frac{s_i^2}{N} \tilde{P}_i\Big) \tag{9.62}$$

where $\Lambda_{\tilde{X}} = \operatorname{diag}(\tilde{P}_1, \ldots, \tilde{P}_{n_t})$. From the assumption that H has full rank the number of non-zero singular values is $n = \min\{n_t, n_r\}$.

Hence, the MIMO channel is equivalent to a channel model containing parallel Gaussian channels with attenuation given by the singular values of H. As before, to optimise the usage of the channel the power levels $\tilde{P}_i$ can be found by water-filling. The result is summarised in the following theorem.

THEOREM 9.5 Given a MIMO channel with n_t transmit antennas, n_r receive antennas and the attenuation matrix H, the capacity is given by

$$C = \sum_{i=1}^{n} \frac{1}{2} \log \Big(1 + \frac{s_i^2}{N} \tilde{P}_i\Big) \tag{9.63}$$

where the power levels $\tilde{P}_i$ is found by

$$\begin{cases} \tilde{P}_i = \Big(B - \frac{N}{s_i^2}\Big)^+ \\ \sum_i \tilde{P}_i = P \end{cases} \tag{9.64}$$

and s_i are the $n = \min\{n_t, n_r\}$ singular values in the singular value decomposition $H = USV^T$. □

The optimising distribution is $\tilde{\boldsymbol{X}} \sim \mathrm{N}(\boldsymbol{0}, \Lambda_{\tilde{X}})$ where $\Lambda_{\tilde{X}} = \operatorname{diag}(\tilde{P}_1, \ldots, \tilde{P}_{n_t})$. This corresponds to the input distribution $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{0}, \Lambda_X)$ where $\Lambda_X = V \Lambda_{\tilde{X}} V^T$.

The above derivation for the MIMO channel assumes that both the transmitter and receiver have full knowledge of the channel matrix H. However, MIMO is a typical radio channel and often the attenuation changes rapidly over time, which makes estimation a hard task. Another reasonable assumption would then be that the receiver has full knowledge of the channel, while the transmitter does not know H. In this case the optimal distribution on the power levels is obtained from a uniform distribution, $\tilde{P}_i = P/n_t$. In this case the corresponding capacity is given by the next theorem.

THEOREM 9.6 Given a MIMO channel with n_t transmit antennas and n_r receive antennas. If the channel attenuation matrix H is not known by the transmitter but perfectly known by the receiver, the capacity is given by

$$C = \frac{n}{2} \log\Big(1 + \frac{s_i^2 P}{N n_t}\Big) \tag{9.65}$$

where s_i are $n = \min\{n_t, n_r\}$ the singular values in the singular value decomposition $H = USV^T$. $\square$