6.962 Graduate Seminar in Communications - Week 8

Space-Time Modulation for Unknown Fading

- Bertrand M. Hochwald and Thomas L. Marzetta

presenter: Huan Yao

November 1, 2000

Last week, Aaron talked about coherent multiple antenna channel. This week, I will talk about the non-coherent case, also for single user to single user.

1 Motivation

In a coherent communication setup, the receiver is assumed to know the channel coefficients. So typically, they apply to quasi-stationary environment, so that receiver can track the channel. When the channel varies fast, i.e., we have fast fading, it would be hard to train the channel coefficients quickly and reliably, especially where there is noise. Furthermore, when there are multiple antennas, there are more coefficients to train, the training job will take even more effort.

The motivation for studying non-coherent channel is to not separate the channel estimation and data decoding into two different steps, but rather consider the channel as unknown and do the estimations all together (or just the data part), taking a more general approach, making one big box rather than two boxes.

2 Signal model

Figure 1 displays a communication link with M transmitter antennas and N receiver antennas that operates in a Rayleigh flat-fading environment. The flat-fading assumption assumes that the signal bandwidth is much smaller than the *coherence bandwidth* (the width of the band that the channel frequency response stay approximately flat) of the channel, so the effect of the channel is multiplicative.

With this assumption, we can further model the connection between every receiver antenna and every transmitter antenna to be an independent identically Rayleigh distributed, zero mean, unit variance, circularly-symmetric, complex gain, h_{mn}, $p(h_{mn}) = (1/\pi)exp\{-|h_{mn}|^2\}$. All coefficients are unknown to both transmitter and receiver. The fading channel, characterized by the fading coefficient matrix, $H = \{h_{mn}\}$, is assumed to be piecewise constant. The entries, h_{mn}, remain constant for T symbol periods, after which they change to new independent values. The value of T is related to the *coherence time* (how long the channel stays approximately constant) of the fading channel. This piecewise constant model is not entirely accurate, but convenient for analysis.

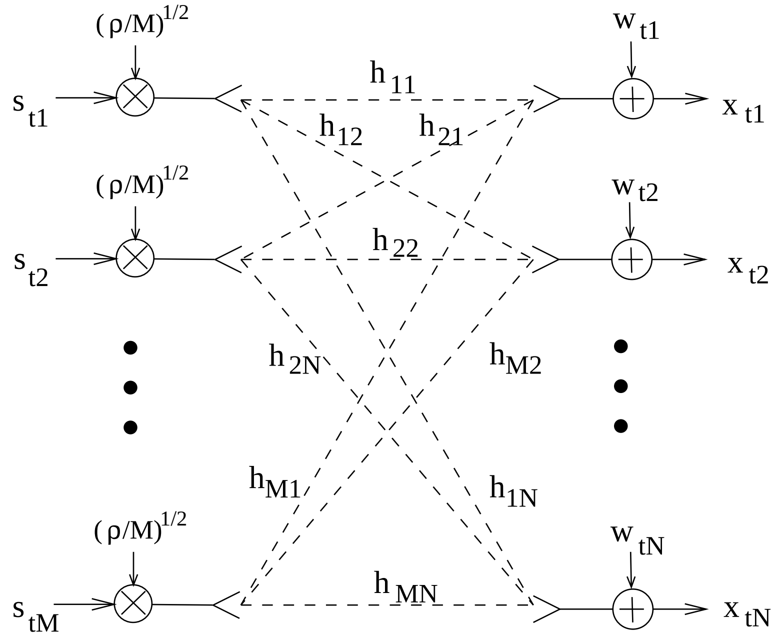

Figure 1: Wireless link comprising M transmitter and N receiver antennas. Every receiver antenna is connected to every transmitter antenna through an independent identically Rayleigh distributed complex gain, h_{ij}, which is unknown to both transmitter and receiver.

At the receiver end, the signals are corrupted by additive noise, w_{ij}, which are also independent identically Rayleigh distributed, zero mean, unit variance, circularly-symmetric, complex numbers. They are also independent from the channel coefficients.

At each time t, the signal transmitted by antenna m is $\sqrt{\frac{\rho}{M}} s_{tm}$. The signal is normalized so that

$$\frac{1}{M} \sum_{m=1}^{M} E(|s_{tm}|^2) = 1, \text{ for } \quad t = 1, \ldots, T.$$

The scalar ρ denotes the expected signal-to-noise ratio at each receiver. Normalization ensures that the total expected transmitted power is independent of M, the number of antennas used, for a fixed ρ.

To introduce the matrix notation of this signal model, let

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1M} \\ \vdots & \ddots & \vdots \\ s_{T1} & \cdots & s_{TM} \end{bmatrix}, \qquad X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{TN} \end{bmatrix},$$

$$H = \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{M1} & \cdots & h_{MN} \end{bmatrix}, \qquad W = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{T1} & \cdots & w_{TN} \end{bmatrix},$$

where

- S is the $T \times M$ matrix of transmitted signal by M antennas in time T. Each column corresponds to what one antenna transmits. Each row corresponds to what all antenna transmits at one time.
- H is the $M \times N$ matrix of propagation coefficients between the M transmitter antennas and N receiver antennas. The entire matrix is constant in time for T samples.
- W is the $T \times N$ matrix of noise.
- X is the $T \times N$ matrix of received signal by N antennas in time T.

The matrix representation of this signal model is

$$X = \sqrt{\frac{\rho}{M}} SH + W.$$

One important thing I would like to point out is that, in the absence of noise, $X = \sqrt{\frac{\rho}{M}} SH$. The column space of X is the same as the column space of S. No matter what H, this is always true. As David Tse pointed out in his talk on [4], it is this column space that carries all the information. In this talk, I will sometimes think of the signal matrix S as a subspace.

3 How to pick capacity achieving distribution

To have a sense of what the capacity achieving distribution should be, let's first look at the channel input output probability,

$$p(X|S) = \frac{\exp\left(-\mathrm{tr}\{\left[I_T + (\rho/M)SS^\dagger\right]^{-1} XX^\dagger\}\right)}{\pi^{TN} \det^N \left[I_T + (\rho/M)SS^\dagger\right]}.$$

The effect of the transmitted signals on the conditional probability density of the received signal is through the $T \times T$ matrix $SS^\dagger$.

The above statement is a very important insight into code design. One immediate corollary is that we don't need to make M greater than T. Since $SS^\dagger$ is $T \times T$, any $SS^\dagger$ can be realized by using no more than $M = T$ antennas. (Zheng and Tse [4] actually showed that not more than $T/2$ antennas are needed. It is not obvious just yet, but I will show you some intuition later.)

The main conclusion that follows from this insight and some rotational symmetry arguments is that the capacity achieving distribution for S is of the form $S = \Phi V$, where Φ is an $T \times M$ *isotropically distributed* unitary matrix, and V is an independent $M \times M$ real, nonnegative, diagonal matrix (with some distribution).

Isotropically distributed (i.d.) unit vectors and unitary matrices : An i.d. unit vector is a unit vector that is equally likely to point in any direction in the (complex) space. Think of picking a point uniformly on a sphere. An i.d. unitary matrix is a matrix formed by first picking an i.d. unit vector to be the first column, then pick another i.d. unit vector in the rest of the space, i.e., orthogonal to the first vector, to be the second column, and so on. Equivalently, we can pick all the vectors and then do Gram-Schmidt. One property of i.d. unitary matrix is rotational symmetry, i.e., $p(\Phi) = p(\Theta^\dagger \Phi),\ \forall \Theta : \Theta^\dagger \Theta = I$.

As for the capacity achieving distribution of V, they only gave a limiting result: for fixed M, as $T \to \infty$, that $V \to \sqrt{T} I_M$.

I will provide a quick proof and some intuition of these results in my presentation. See [1] for more detailed proofs.

4 How to pick deterministic codebook

Now, we want to design deterministic codebook using some of the capacity achieving distribution results as a guide line.

Say we want to design $L = 2^{RT}$ signals, so we transmit at rate R bits per time. We need to design signal matrices $S_1 = \Phi_1 V, \ldots, S_L = \Phi_L V_L$. Since in the limiting case, $V \to \sqrt{T} I_M$, we could choose $V_1 = \cdots = V_L = \sqrt{T} I_M$ for our deterministic codebook, so that $S_1 = \sqrt{T}\Phi_1, \ldots, S_L = \sqrt{T}\Phi_L$, where $\Phi_l^\dagger \Phi_l = I_M$. This scheme is called *Unitary Space-Time Modulation*.

It is shown in [2] that the maximum likelihood decoder for this scheme is

$$\Phi_{ml} = \text{argmax}_{\Phi_l = \Phi_1, \ldots, \Phi_L} tr\{X^\dagger \Phi_l \Phi_l^\dagger X\}.$$

This ML receiver maximizes the energy contained in the inner product $\Phi_l X$. (This decoder reminds me of a match filter.) I can also consider it as finding the "nearest" valid codeword subspace to the received signal. In addition, we can also obtain the ML estimate of the channel (if we need to), $H_{ml} = (\frac{\rho T}{M})^{-frac12} \Phi_{ml}^\dagger X$

Now that we have a decoder, we can compute the probability of error. The two-signal probability of error (the probability of confusing between S_1 and S_2 as if they were the only two signals) has Chernoff upper bound

$$P_e \leq \frac{1}{2} \prod_{m=1}^{M} \left[\frac{1}{1 + \frac{(\rho T/M)^2 (1 - d_m^2)}{4(1 + \rho T/M)}} \right]^N .$$

where $1 \geq d_1 \geq \ldots \geq d_M \geq 0$ are the singular values of the $M \times M$ correlation matrix $\Phi_2^\dagger \Phi_1$. I will show that these $d_i\ 's$ corresponds to the cosines of a set of *principle angles* defined between the two subspaces Φ_1 and Φ_2.

From the equation above, we can see that the probability of error decreases with decreasing d_m, which corresponds to increased separation between the subspaces in terms of the angles between them. One iterative way of finding a good deterministic codebook is described in [2]. The basic idea is to first compute all $L(L-1)/2$ correlation matrices $\Phi_i^\dagger \Phi_j$ and identify the worst pair, then try to "move them apart". Repeat until improvements stop. One example with $M = 1$, $T = 5$, and $L = 32$ is shown in the paper in Figure 3.

I can consider the code design as a task of trying to fit L M-dimensional subspaces in T dimensions. When $M > \frac{T}{2}$, any two subspaces would have to share dimensions, i.e, the worst singular values of the correlation matrix are 1's, which contributes nothing towards decreasing the probability of error. This hints that having $M > \frac{T}{2}$ might be counter productive. We might want to use less antennas. In [4], Zheng and Tse showed that $M = \lfloor \frac{T}{2} \rfloor$ is optimal (assume $N = M$). In my presentation, I will give you several examples and a quick proof of why $M = \lfloor \frac{T}{2} \rfloor$ is optimal.

5 Systematic design of unitary space-time signals

The iterative design method described in last section is too cumbersome, especially when $L = 2^{RT}$ gets really large (exponentially!).

A systematic method of designing good Unitary space-time codebook was developed in [3]. These designs might not be optimal, but they seem to be reasonably good and are easy to design.

They propose to use

$$\Phi_l = \Theta^{l-1}\Phi_1, \quad l = 1, \cdots, L$$

where Θ is a $T \times T$ diagonal unitary matrix such that $\Theta^L = I_T$, and Φ_1 is a unitary matrix. This way, for any two Φ_i and Φ_j, the correlation is $\Phi_1^\dagger \Theta^{(j-i) \bmod L}\Phi_1$. Every signal matrix forms the same set of correlation matrices with all the other signal matrices. (This reminds me of binary linear block code with Hamming distance metric.) Now to see whether a codebook is good, there are only a total of $L-1$ correlation matrices to check, as opposed to $L(L-1)/2$.

To pick the diagonal matrix Θ, since $\Theta^L = I_T$, the diagonal entries are

$$[\Theta]_{tt} = e^{i\frac{2\pi}{L}u_t}, \quad t = 1, \cdots, T,$$

and the u_t are integers such that $0 \leq u_1, \cdots, u_T \leq L-1$. This makes the search much easier, since we only need to search over a finite set of integers.

The choice of the initial signal Φ_1 is less critical. In [3], the authors proposed to use a Fourier-based construction. I will not go over the detail of this method in this talk. I will show you simple examples.

References

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", IEEE Trans. on Information Theory, vol. 45(1), pp. 139-157, Sep. 1999.

[2] B. M. Hochwald and T. L. Marzetta, "Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading", IEEE Trans. on Information Theory, vol. 46(2), pp. 543-564, March 2000.

[3] B. M. Hochwald, T. L. Marzetta, and T. J. Richardson, "Systematic Design of Unitary Space-Time Constellations", IEEE Trans. on Information Theory, vol. 46(6), pp. 1962-1973, Sep. 2000.

[4] L. Zheng and D. N.C. Tse, "Packing Spheres in the Grassmann Manifold: A Geometric Approach to the Non-coherent Multi-Antenna Channel", submitted to the IEEE Transactions on Information Theory, April 2000.