Elements of Information Theory **2024**

Lecture 1: Capacity bounds

Lecturer: S. Yang

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

In this lecture, we introduce capacity bounding techniques for memoryless channels, and apply them to non-conventional multi-antenna channels, such as channels with quantized outputs.

1.1 Capacity of memoryless channels

We recall that for a memoryless channel $P_{Y|X}$, with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, the channel capacity is

$$C = \sup_{P_X \in \mathcal{P}} I(P_X, P_{Y|X}) \tag{1.1}$$

where $\mathcal{P}$ is the set of admissible input distributions, i.e., distributions satisfying some constraints (e.g., power constraints in wireless communications). The mutual information $I(P_X, P_{Y|X})$ can be written as $I(X;Y)$ when the joint distribution is clear from the context.

Note that the mutual information is

$$I(P_X, P_{Y|X}) = D(P_{Y|X}P_X \| P_X P_Y) = D(P_{Y|X} \| P_Y | P_X) = \mathbb{E}_{P_{XY}}\left[\log \frac{P(Y|X)}{P(Y)}\right] = \mathbb{E}_{P_{XY}}\left[\log \frac{P(X|Y)}{P(X)}\right] \tag{1.2}$$

where P_Y is the marginal distribution of $P_{XY} = P_{Y|X}P_X$. Here, we consider the discrete case or the continuous case where the above probabilities are with respect to the counting or Lebesgue measure. It can be extended to the general case with the Radon-Nikodym derivative. For simplicity, we use natual logarithm throughout the lecture.

Except for a few special cases, the mutual information is hard to evaluate in general: either $P(Y)$ or $P(X|Y)$ needs to be computed, before taking the expectation. In this lecture, we are interested in deriving computable bounds. To that end, we can rewrite the mutual information as follows:

$$I(P_X, P_{Y|X}) = \mathbb{E}_{P_{XY}}\left[\log \frac{P(Y|X)}{Q(Y)}\right] - D(P_Y \| Q_Y) \tag{1.3}$$

$$= \mathbb{E}_{P_{XY}}\left[\log \frac{Q(X|Y)}{P(X)}\right] + D(P_{X|Y} \| Q_{X|Y} | P_Y), \tag{1.4}$$

which holds for any distribution Q_Y and conditional distribution $Q_{X|Y}$.

From the positivity of the divergence, we obtain the following.

$$I(P_X, P_{Y|X}) = \min_{Q_Y} \mathbb{E}_{P_{XY}}\left[\log \frac{P(Y|X)}{Q(Y)}\right] = \max_{Q_{X|Y}} \mathbb{E}_{P_{XY}}\left[\log \frac{Q(X|Y)}{P(X)}\right]. \tag{1.5}$$

1.1.1 Capacity upper bound

For any Q_Y, we have the following upper bound on the mutual information:

$$I(P_X, P_{Y|X}) \le \mathbb{E}_{P_{XY}}\left[\log \frac{P(Y|X)}{Q(Y)}\right] \tag{1.6}$$

In most cases, we would like to find Q_Y that is easy to compute or analyze, or both. In particular, for a specific channel, we can find a smoothly parameterized family of distributions, $\{Q_\theta : \theta \in \Theta\}$ to optimize the upper bound.

$$I(P_X, P_{Y|X}) \le \inf_{\theta\in\Theta} \mathbb{E}_{P_{XY}}\left[\log \frac{P(Y|X)}{Q_\theta(Y)}\right] \tag{1.7}$$

1.1.2 Capacity lower bound

For any $Q_{X|Y}$, we have the following lower bounds on the mutual information:

$$I(P_X, P_{Y|X}) \ge \mathbb{E}_{P_{XY}}\left[\log \frac{Q(X|Y)}{P(X)}\right]. \tag{1.8}$$

In most cases, we would like to find $Q_{X|Y}$ that is easy to compute or analyze, or both. In particular, for a specific channel, we can find a smoothly parameterized family of distributions, $\{Q_\theta : \theta \in \Theta\}$ to optimize the lower bound.

$$I(P_X, P_{Y|X}) \ge \sup_{\theta\in\Theta} \mathbb{E}_{P_{XY}}\left[\log \frac{Q_\theta(X|Y)}{P(X)}\right] \tag{1.9}$$

1.1.2.1 Example 1: Mismatched decoding bound

One specific way to set $Q(x|y)$ is to let $Q(x|y) = \frac{P(x)Q(y|x)}{\mathbb{E}_P[Q(y|X)]}$ with

$$Q(y|x) = q(x,y)^t a(x) b(y). \tag{1.10}$$

for some positive functions q, a, b, and $t \ge 0$.

We have the following lower bounds.

For any $Q_{X|Y}$, we have the following lower bounds on the mutual information:

$$I(P_X, P_{Y|X}) \ge I_{\mathrm{LM}}(P_X, P_{Y|X}) := \sup_{t\ge 0, a(\cdot)\ge 0} \mathbb{E}_{P_{XY}}\left[\log \frac{q(X,Y)^t a(X)}{\mathbb{E}_{\bar{X}\sim P_X} q(\bar{X},Y)^t a(\bar{X})}\right] \tag{1.11}$$

$$\ge I_{\mathrm{GMI}}(P_X, P_{Y|X}) := \sup_{t\ge 0} \mathbb{E}_{P_{XY}}\left[\log \frac{q(X,Y)^t}{\mathbb{E}_{\bar{X}\sim P_X} q(\bar{X},Y)^t}\right] \tag{1.12}$$

The above bounds can be achieved with **mismatched decoders** using any product metric $q(x^n, y^n) = \prod_{i=1}^n q(x_i, y_i)$ for a given q, instead of the optimal metric (e.g., the likelihood function).

1.1.2.2 Example 2: Hard decoding bound for finite $\mathcal{X}$

When $\mathcal{X}$ is finite, one can set the auxiliary distribution as follows:

$$Q(x|y) = \left(\frac{\alpha}{M-1}\right)^{1(x\neq\hat{x}(y))}(1-\alpha)^{1(x=\hat{x}(y))}, \quad x \in \mathcal{X}. \tag{1.13}$$

Then, we obtain the following upper bound.

$$I(P_X, P_{Y|X}) \geq H(P_X) - \log(M-1)P_e - \left[P_e \log\frac{1}{\alpha} + (1-P_e)\log\frac{1}{1-\alpha}\right], \tag{1.14}$$

which can be maximized by setting $\alpha = P_e$ where $P_e := \mathbb{P}(X \neq \hat{x}(Y))$ that depends on both P_X and $\hat{x}$. By maximizing over P_X, we have the capacity lower bound

$$C \geq (1-P_e)\log M - H_2(P_e). \tag{1.15}$$

That is approximately $(1-P_e)\log M$ close to the capacity of an erasure channel. Intuitively, it is equivalent to first decode each symbols, then the whole sequence by considering the discrete symbol input symbol output channel after the hard decoding.

An alternative way to obtain the lower bound is from Fano's inequality $H(X|\hat{x}(Y)) \leq P_e \log(M-1) + H_2(P_e)$. Indeed, we can write

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (1.16)\\ &= H(X) - H(X|Y,\hat{x}(Y)) && (1.17)\\ &\geq H(X) - H(X|\hat{x}(Y)) && (1.18)\\ &\geq H(X) - P_e\log(M-1) - H_2(P_e) && (1.19) \end{aligned}$$

1.1.3 Capacity: Numerical computation

In some cases, e.g., the discrete case with finite alphabets, it is possible to compute the capacity numerically. We present the iterative algorithm, Blahut-Arimoto (BA) algorithm, for such cases.

With the variational characterization, the capacity can be written as

$$\max_{P_X}\max_{Q_{X|Y}} \mathbb{E}_{P_X P_{Y|X}}\left[\log\frac{Q(X|Y)}{P(X)}\right] \tag{1.20}$$

which is a concave problem. The BA algorithm works as follows:

- Intialize P_X, e.g., uniform over $\mathcal{X}$
- Compute $P_{X|Y}$ and let $Q_{X|Y} = P_{X|Y}$. Note that $P(x|y) \propto P(x)P(y|x)$, i.e., just compute the joint distribution and normalize for each $y \in \mathcal{X}$. This step just computes the mutual information $I(P_X, P_{Y|X})$, and solve the inner maximization.
- Update P_X by letting $P(x) = Q'(x) \propto \exp\left[\sum_y P(y|x)\log Q(x|y)\right]$. This step solves the outer maximization. To see this, let us rewrite

$$\begin{aligned} \max_{P_X}\mathbb{E}_{P_X P_{Y|X}}\left[\log\frac{Q(X|Y)}{P(X)}\right] &= \max_{P_X}\mathbb{E}_{P_X}\left[\log\frac{\exp\left(\mathbb{E}_{P_{Y|X}}\log Q(X|Y)\right)}{P(X)}\right] && (1.21)\\ &= -\log c + \max_{P_X}\mathbb{E}_{P_X}\left[\log\frac{c\exp\left(\mathbb{E}_{P_{Y|X}}\log Q(X|Y)\right)}{P(X)}\right] && (1.22)\\ &= -\log c + \max_{P_X}(-D(P_X\|Q'_X)), && (1.23) \end{aligned}$$

  where c is the normalization factor for the proposed distribution. The maximization is achieved when $P_X = Q'_X$.

1.2 Applications

1.2.1 Application 1: Non-coherent channel

In the following, we are interested in the single-input multiple-output (SIMO) channel.

$$\boldsymbol{y} = \boldsymbol{h}X + \boldsymbol{z} \tag{1.24}$$

where $\boldsymbol{h}, \boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{I})$ and $X \in \mathbb{C}$ is the input subject to the average power constraint P.

When the channel realization $\boldsymbol{h}$ is unknown to both the transmitter and the receiver, the channel is characterized by $p(\boldsymbol{y}|x) \sim \mathcal{CN}(0, (1+|x|^2)\boldsymbol{I}_{n_\mathrm{r}})$, i.e., the phase information of x is lost. Therefore, it is without loss of optimality to consider x real and positive. The conditional differential entropy is

$$h(\boldsymbol{y}|X) = n_\mathrm{r}\mathbb{E}\left[\log(\pi e(1+|X|^2))\right] = n_\mathrm{r}\log \pi e + n_\mathrm{r}\mathbb{E}\left[\log(1+|X|^2)\right] \tag{1.25}$$

First, let us recall the Gamma distribution that we will use repeatedly.

The Gamma distribution with parameter $\alpha > 0, \lambda > 0$ has the following pdf:

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \tag{1.26}$$

The Gamma-integral formula states that for parameters $a, b > 0$:

$$\int_0^\infty x^{a-1} e^{-bx}\, dx = \frac{\Gamma(a)}{b^a}. \tag{1.27}$$

1.2.1.1 Capacity upper bound

We apply (1.6) to otain an upper bound on the mutual information. To that end, we need to introduce a "rich" family of distributions Q_Y so that we refine the upper bound and tighten it. For the given model, we consider the following family.

$$Q_\theta(\boldsymbol{y}) := c_\theta \cdot (\|\boldsymbol{A}\boldsymbol{y}\|^2 + \delta)^{\alpha-1} \cdot \|\boldsymbol{A}\boldsymbol{y}\|^{2(1-n_\mathrm{r})} e^{-(\|\boldsymbol{A}\boldsymbol{y}\|^2+\delta)/\beta}, \quad \boldsymbol{y} \in \mathbb{C}^{n_\mathrm{r}},$$

where $\theta := (\alpha > 0, \beta > 0, \delta \geq 0, \boldsymbol{A} \in \mathbb{C}^{n_\mathrm{r}\times n_\mathrm{r}})$ is the parameter of the distribution;

$$c_\theta := \frac{\Gamma(n_\mathrm{r})|\det \mathbf{A}|^2}{\pi^{n_\mathrm{r}}\beta^\alpha \Gamma(\alpha, \delta/\beta)} \tag{1.28}$$

is the normalization factor;

$$\Gamma(\alpha, \xi) := \int_\xi^\infty t^{\alpha-1} e^{-t}\, dt, \tag{1.29}$$

is the incomplete Gamma function and $\Gamma(\alpha) := \Gamma(\alpha, 0)$ is the Gamma function.

Although the expression looks complicated, the idea behind it is quite intuitive. It is a linear transformation ($\boldsymbol{A}^{-1}$) of an isotropic vector whose squared norm has regularized (δ) Gamma distribution (α, β). The upper bound (1.6) becomes

$$I(P_X, P_{Y|X}) \le -h(\boldsymbol{y}|X) - \log c_\theta + (n_\mathrm{r} - 1)\mathbb{E}[\log \|\boldsymbol{A}\boldsymbol{y}\|^2] \tag{1.30}$$

$$+(1-\alpha)\mathbb{E}[\log(\|\boldsymbol{A}\boldsymbol{y}\|^2 + \delta)] + \frac{1}{\beta}\mathbb{E}[\|\boldsymbol{A}\boldsymbol{y}\|^2 + \delta]. \tag{1.31}$$

Since optimizing jointly over all feasible θ remains elusive, one can reduce the parameter set, e.g.,

$$\delta = 0, \quad \beta = \frac{1}{\alpha}\mathbb{E}[\|\boldsymbol{A}\boldsymbol{y}\|^2]. \tag{1.32}$$

In our application, we also know that the output must be isotropic. Therefore, we set $\boldsymbol{A} = \boldsymbol{I}$.

$$I(P_X, P_{Y|X}) \le \log \pi^{n_\mathrm{r}} - \log\Gamma(n_\mathrm{r}) + \log\Gamma(\alpha) - \alpha\log\alpha + \alpha \tag{1.33}$$

$$-h(\boldsymbol{y}|X) + \alpha\log\mathbb{E}[\|\boldsymbol{y}\|^2] + (n_\mathrm{r} - \alpha)\mathbb{E}[\log\|\boldsymbol{y}\|^2] \tag{1.34}$$

Note that $\|\boldsymbol{y}\|^2$ is equivalent in distribution to $\|\boldsymbol{z}\|^2(1+|X|^2)$, we have $\log\mathbb{E}[\|\boldsymbol{y}\|^2] = \log n_\mathrm{r} + \log\mathbb{E}[1+|X|^2]$, and $\mathbb{E}[\log\|\boldsymbol{y}\|^2] = \mathbb{E}\log\|\boldsymbol{z}\|^2 + \mathbb{E}[\log(1+|X|^2)] = \psi(n_\mathrm{r}) + \mathbb{E}[\log(1+|X|^2)]$, where $\psi(s) := \frac{\mathrm{d}}{\mathrm{d}s}\log\Gamma(s)$ is the digamma function. We have the following upper bound for the MISO channel.

$$I(P_X, P_{Y|X}) \le \alpha - n_\mathrm{r} + \log\frac{\Gamma(\alpha)}{\Gamma(n_\mathrm{r})} - \alpha\log\frac{\alpha}{n_\mathrm{r}} + (n_\mathrm{r} - \alpha)\psi(n_\mathrm{r}) \tag{1.35}$$

$$+\alpha(\log\mathbb{E}[1+|X|^2] - \mathbb{E}[\log(1+|X|^2)]) \tag{1.36}$$

$$\le \alpha - n_\mathrm{r} + \log\frac{\Gamma(\alpha)}{\Gamma(n_\mathrm{r})} - \alpha\log\frac{\alpha}{n_\mathrm{r}} + (n_\mathrm{r} - \alpha)\psi(n_\mathrm{r}) \tag{1.37}$$

$$+\alpha(\log\mathbb{E}[1+|X|^2]) \tag{1.38}$$

$$\le \alpha - n_\mathrm{r} + \log\frac{\Gamma(\alpha)}{\Gamma(n_\mathrm{r})} - \alpha\log\frac{\alpha}{n_\mathrm{r}} + (n_\mathrm{r} - \alpha)\psi(n_\mathrm{r}) + \alpha\log(1+P). \tag{1.39}$$

Note that the above upper bound does not depend on P_X, which implies that it is also an upper bound of the capacity. Since it is valide for all α, we can choose α such that the upper bound is minmized.

> For the non-coherent MISO channel, the capacity is upper bounded by
>
> $$C \le \inf_{\alpha>0}\,(\alpha - n_\mathrm{r}) + \log\frac{\Gamma(\alpha)}{\Gamma(n_\mathrm{r})} - \alpha\log\frac{\alpha}{n_\mathrm{r}} + (n_\mathrm{r} - \alpha)\psi(n_\mathrm{r}) + \alpha\log(1+P). \tag{1.40}$$

Remarkably, with the same setting, the upper bound remains the same even with multiple transmit antennas (check).

1.2.1.2 Capacity lower bound

Since the output only depends on $|X|^2$, we can redefine the input as $S := |X|^2$ with constraint $S \ge 0$ and $\mathbb{E}[S] \le P$. To obtain a lower bound, let us consider the GMI rate (1.12). For tractability, we let $S \sim \mathrm{Gamma}(\alpha, \lambda)$, and the metric q be a Gamma pdf, i.e.,

$$q(s, \boldsymbol{y}) = f_{\tilde\alpha,\tilde\lambda}(s), \quad \frac{\tilde\alpha}{\tilde\lambda} = \eta\|\boldsymbol{y}\|^2 \tag{1.41}$$

The reason for this choice is that one can now compute the expectation $\mathbb{E}[q(S,\boldsymbol{y})^t]$ in closed form. Thus, we obtain:

$$\mathbb{E}_S[q(S,\boldsymbol{y})^t] = \frac{(\tilde\lambda^{\tilde\alpha})^t\lambda^\alpha\Gamma(t(\tilde\alpha - 1) + \alpha)}{\Gamma(\tilde\alpha)^t\Gamma(\alpha)(t\tilde\lambda + \lambda)^{t(\tilde\alpha-1)+\alpha}}. \tag{1.42}$$

Next, we compute the normalized term:

$$\frac{f(S)^t}{E\left([f(S)]^t\right)} = \frac{\left(\frac{\tilde{\lambda}^{\tilde{\alpha}} S^{\tilde{\alpha}-1} e^{-\tilde{\lambda} S}}{\Gamma(\tilde{\alpha})}\right)^t}{E\left([f(S)]^t\right)}. \tag{1.43}$$

Finally, we get

$$\begin{aligned}
g(\alpha,\tilde{\alpha},\eta) &:= \log\Gamma(\alpha) - \alpha\log\lambda - \log\Gamma\Big(t(\tilde{\alpha}-1)+\alpha\Big) && (1.44)\\
&\quad + \mathbb{E}_{S,Y}\left\{[t(\tilde{\alpha}-1)+\alpha]\log\Big(t\frac{\tilde{\alpha}}{\eta\,\|\boldsymbol{y}\|^2}+\lambda\Big) + t(\tilde{\alpha}-1)\log S - t\,\frac{\tilde{\alpha}}{\eta\,\|\boldsymbol{y}\|^2}\,S\right\} && (1.45)\\
&= \log\Gamma(\alpha) - \alpha\log\lambda - \log\Gamma\Big(t(\tilde{\alpha}-1)+\alpha\Big) + t(\tilde{\alpha}-1)(\psi(\alpha)-\log\lambda) && (1.46)\\
&\quad - \frac{t\,\tilde{\alpha}}{(n_{\mathrm{r}}-1)\eta}\mathbb{E}\frac{S}{1+S} - [t(\tilde{\alpha}-1)+\alpha](\log\eta + \psi(n) + \mathbb{E}_S\left\{\log\Big(1+S\Big)\right\} && (1.47)\\
&\quad + [t(\tilde{\alpha}-1)+\alpha]\mathbb{E}_{S,Z}\left\{\log\Big(t\tilde{\alpha} + \lambda\eta\,(1+S)\|\boldsymbol{z}\|^2\Big)\right\} && (1.48)\\
&\geq \log\Gamma(\alpha) - \alpha\log\lambda - \log\Gamma\Big(t(\tilde{\alpha}-1)+\alpha\Big) + t(\tilde{\alpha}-1)(\psi(\alpha)-\log\lambda) && (1.49)\\
&\quad - \frac{t\,\tilde{\alpha}}{(n_{\mathrm{r}}-1)\eta}\frac{\mu}{1+\mu} - [t(\tilde{\alpha}-1)+\alpha]\Big(\log\eta + \psi(n) + \log\Big(1+\mu\Big)\Big) && (1.50)\\
&\quad + [t(\tilde{\alpha}-1)+\alpha]\mathbb{E}_{S,Z}\left\{\log\Big(t\tilde{\alpha} + \lambda\eta\,(1+S)\|\boldsymbol{z}\|^2\Big)\right\} && (1.51)
\end{aligned}$$

where $\lambda := \frac{\alpha}{\mu}$

1.2.2 Application 2: One-bit ADC

Let us now consider the case with one-bit ADC. For simplicity, we consider the case without fading and the channel is real, i.e.,

$$Y_i = \operatorname{sign}(X + Z_i), \quad i = 1,\dots,n_{\mathrm{r}}, \tag{1.52}$$

where $z_i \sim \mathcal{N}(0,1)$. Note that the conditional distribution of $\boldsymbol{y}$ given $X = x$ is i.i.d. $\mathrm{Bern}(\lambda(x))$ where $\lambda(x) := Q(x)$, i.e.,

$$P(\boldsymbol{y}|x) = \lambda(x)^{T(\boldsymbol{y})}(1-\lambda(x))^{n_{\mathrm{r}}-T(\boldsymbol{y})}, \tag{1.53}$$

where $T(\boldsymbol{y}) := \sum_{i=1}^{n_{\mathrm{r}}} \mathbf{1}\{Y_i = -1\}$ is the sufficient statistic for x and follows the binomial distribution $(n_{\mathrm{r}}, \lambda(x))$. Then the capacity of the channel is

$$C = \sup_{P_X} I(P_X, P_{\tilde{Y}|X}) = \sup_{P_\lambda} I(\lambda, \mathrm{Binom}(n,\lambda)). \tag{1.54}$$

1.2.2.1 Capacity upper bound

To derive a capacity upper bound, one may apply (1.7) with a family of auxiliary distributions. For instance, we can use the Beta-binomial distribution with parameters $\alpha = \beta$, i.e.,

$$Q_\alpha(k) := \binom{n_{\mathrm{r}}}{k}\frac{B(k+\alpha, n_{\mathrm{r}}-k+\alpha)}{B(\alpha,\alpha)} \tag{1.55}$$

while the true conditional distribution is

$$P(k|\lambda) := \binom{n_{\mathrm{r}}}{k}\lambda^k(1-\lambda)^{n_{\mathrm{r}}-k} \tag{1.56}$$

For any fixed distribution on λ, one can get the following upper bound:

$$n_{\mathrm{r}}\mathbb{E}[\lambda\log\lambda + (1-\lambda)\log(1-\lambda)] - \mathbb{E}[\log\Gamma(k+\alpha) + \log\Gamma(n_{\mathrm{r}}-k+\alpha)] + \log B(\alpha,\alpha) + \log\Gamma(n_{\mathrm{r}}+2\alpha) \tag{1.57}$$

1.2.2.2 Capacity lower bound

For the lower bound, one can compute the GMI lower bound with a well chosen metric or the hard decoding lower bound with a simple decoding strategy.

1.2.2.3 Exact capacity computation

When the input alphabet is finite, one can apply the BA algorithm to compute the exact capacity.

Elements of Information Theory **2024**

Lecture 2: Detection algorithms

Lecturer: S. Yang

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

In this lecture, we study the detection problem in wireless communication channels. It is related to integer programming problem.

2.1 Maximum likelihood detection

Let us assume that $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. In many cases, we are interested in estimating X from the observation Y. Let $\hat{X} := \hat{x}(Y)$ be the estimate with a given estimator $\hat{x}$. When $\mathcal{X}$ is discrete, a common distortion measure is the Hamming distortion $d(x, x') = \mathbf{1}(x \neq x')$. The average Hamming distortion is the probability of estimation error $P_e := \mathbb{P}(X \neq \hat{X})$.

We assume that given the source X, the observation is distributed according to $P_{Y|X}$. If the *prior* distribution of X, P_X, is known, then the probability of error is minimized with the maximum *a posteriori* detector, i.e.,

$$\begin{aligned} 1 - P_e^* &= \max_{\hat{x}(\cdot)} \mathbb{E}_Y \mathbb{E}_{X|Y} \left(\mathbf{1}(X = \hat{x}(Y)) \right) && (2.1) \\ &= \max_{\hat{x}(\cdot)} \mathbb{E}_Y \sum_x P(x|Y) \mathbf{1}(x = \hat{x}(Y)) && (2.2) \\ &= \max_{\hat{x}(\cdot)} \mathbb{E}_Y \max_x P(x|Y) && (2.3) \end{aligned}$$

where the optimal detector that achieves the maximum is $\hat{x}^*(Y) = \arg\max_x P(x|Y)$

When the prior is uniform, i.e., $P(x)$ is constant, we have

$$\hat{x}^*(Y) = \arg\max_x P(x|Y) = \arg\max_x P(Y|x) \tag{2.4}$$

from the Bayes' rule. This coincides with the maximum likelihood (ML) detection, which we will focus from now on.

Specifically, we will consider the following equivalent formulation:

$$\min_{x \in \mathcal{X}} -\log P(y|x) \tag{2.5}$$

2.2 Conventional MIMO channels

Consider the channel model

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z} \tag{2.6}$$

where $\boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{I}_{n_r})$ and $\boldsymbol{H}^{n_r \times n_t}$ is the channel matrix known at the decoder. The likelihood function is $P(\boldsymbol{y}|\boldsymbol{x}) = \pi^{-n_r} \exp(-\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2)$, and the ML detection problem (2.5) is equivalent to

$$\min_{x \in \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2. \tag{2.7}$$

Note that the objective function is convex but the feasible set $\mathcal{X}$ is not. Indeed, $\mathcal{X}$ is in general a discrete and finite constellation set. In this lecture, we assume that $\mathcal{X} := \mathcal{S}^{n_t}$ where $\mathcal{S}$ is M^2-QAM, i.e., the real and imaginary parts of $\boldsymbol{x}$ are from M-PAM. The problem (2.7) is NP hard in general, except in a few cases. For instance, when $\boldsymbol{H}$ has orthogonal columns, then the problem is equivalent to

$$\min_{x \in \mathcal{X}} \|\boldsymbol{H}^H \boldsymbol{y} - \boldsymbol{x}\|^2 \tag{2.8}$$

whose solution is $\hat{\boldsymbol{x}} = [\boldsymbol{H}^H \boldsymbol{y}]_{\mathcal{X}}$, known as the rounding of $\boldsymbol{H}^H \boldsymbol{y}$, i.e., find the closest constellation point to $\boldsymbol{H}^H \boldsymbol{y}$. In this case, the complexity is linear in n_t.

2.2.1 Zero-forcing

One popular suboptimal solution in the general case when $n_r \geq n_t$ is the rounding of the relaxed least-square (LS) solution. First, let

$$\hat{\boldsymbol{x}}_{\mathrm{LS}} := \arg\min_{\boldsymbol{x} \in \mathbb{C}^{n_t}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 = (\boldsymbol{H}^H \boldsymbol{H})^{-1} \boldsymbol{H}^H \boldsymbol{y}. \tag{2.9}$$

Then, we can round the relaxed LS solution to the closest constellation point. This is commonly called zero-forcing (ZF) in the literature:

$$\hat{\boldsymbol{x}}_{\mathrm{ZF}} := [(\boldsymbol{H}^H \boldsymbol{H})^{-1} \boldsymbol{H}^H \boldsymbol{y}]_{\mathcal{X}} \tag{2.10}$$

If, however, $\boldsymbol{H}$ is rank-deficient, i.e., $\boldsymbol{H}^H \boldsymbol{H}$ is not invertible, then one can apply the regularized ZF,

$$\hat{\boldsymbol{x}}_{\mathrm{RZF}} := [(\boldsymbol{H}^H \boldsymbol{H} + \lambda \boldsymbol{I})^{-1} \boldsymbol{H}^H \boldsymbol{y}]_{\mathcal{X}} \tag{2.11}$$

where $\lambda > 0$. When we set $\lambda = 1/P$ with P the average power of each input entry x_i, the scheme is commonly referred to as the minimum mean square error (MMSE) detection.

2.2.2 Sphere decoding

The naive way to solve the integer LS problem (2.5) is to enumerate exhaustively $\mathcal{X}$ and find the optimal value, which incurs a complexity with order $|\mathcal{X}| = M^{2n_t}$, prohibitive in when n_t is large. To reduce the complexity, one can consider only points in the ball $\mathcal{B}(\boldsymbol{y}, \rho)$ centered at $\boldsymbol{y}$ with a given radius ρ. Indeed, if $\mathcal{B}(\boldsymbol{y}, \rho) \bigcap \mathcal{X}$ is not empty, then it is without loss of optimality to only enumerate the points in $\mathcal{B}(\boldsymbol{y}, \rho)$. It will be simpler to consider the following equivalent problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{L}\boldsymbol{x} - \boldsymbol{s}\|^2 \tag{2.12}$$

where $\boldsymbol{L} \in \mathbb{C}^{n_t \times n_t}$ is a lower triangular matrix with $\boldsymbol{L} := \boldsymbol{P}\boldsymbol{R}\boldsymbol{P}$; $\boldsymbol{H}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ is the QR decomposition of $\boldsymbol{H}\boldsymbol{P}$, and $\boldsymbol{P} = [\mathbf{1}(j + k = n_t + 1)]_{j,k \in [n_t]}$ is the permutation matrix such that $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{I}_{n_t}$; $\boldsymbol{s} := \boldsymbol{P}\boldsymbol{Q}^H \boldsymbol{y}$. Each $\boldsymbol{x}$ is considered as a leaf node of a tree, and $\boldsymbol{x}_k$, $k = 1, \ldots, n_t - 1$ are the intermediate nodes from the root to the leaf node. We let $\boldsymbol{x}_{n_t} = \boldsymbol{x}$. Define $\boldsymbol{L}_k$ the $k \times k$ upper left submatrix of $\boldsymbol{L}$ and $\boldsymbol{s}_k$ the upper subvector of $\boldsymbol{s}$, we can define $d_k(\boldsymbol{x}) := \|\boldsymbol{L}_k \boldsymbol{x}_k - \boldsymbol{s}_k\|^2$, for $k \in [n_t]$. Note that $d_k(\boldsymbol{x})$ is increasing with k and only depends on $\boldsymbol{x}_k$. The sphere decoding algorithm exploits this property:

- Initialize the radius ρ, e.g., let $\rho = \sqrt{d(\hat{\boldsymbol{x}}_{\mathrm{ZF}})}$.
- Initialize $\hat{\boldsymbol{x}} \leftarrow \hat{\boldsymbol{x}}_{\mathrm{ZF}}$, $\boldsymbol{x}_0 \leftarrow []$, tree level $k \leftarrow 1$, "back track" status $b \leftarrow 0$
- While $k \geq 1$,
  - If $b = 0$ (first time visit),
    - find a list $\mathcal{S}_k := \{x_k \in \mathcal{S} : d_k([\boldsymbol{x}_{k-1}; x_k]) < \rho^2\}$ that is ordered with increasing distance.
    - If $\mathcal{S}_k = \varnothing$, $k \leftarrow k - 1$, $b \leftarrow 0$
    - Else, take the first element x from the list, and let $\boldsymbol{x}_k \leftarrow [\boldsymbol{x}_{k-1}; x]$.
      - If $k = n_t$, update $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x}_k$, update $\rho^2 \leftarrow d(\hat{\boldsymbol{x}})$, $k \leftarrow 1$, $b \leftarrow 1$
  - Else (back tracking)

    - * take the next element x from the list $\mathcal{S}_k$
    - * If failed (no more points), $k \leftarrow k-1$, $b=0$
    - * Else: $\boldsymbol{x}_k \leftarrow [\boldsymbol{x}_{k-1}; x]$
        - · If $d_k(\boldsymbol{x}_k) \geq \rho^2$, then $k \leftarrow k-1$, $b \leftarrow 1$
        - · Else $k \leftarrow k+1$, $b \leftarrow 0$

- return optimal point $\hat{\boldsymbol{x}}$, minimum squared distance ρ^2

2.3 Non-conventional MIMO channels

For non-conventional channels, the ML detection problem is still the same, while the log-likelihood function $f(\boldsymbol{x}) := -\log P(\boldsymbol{y}|\boldsymbol{x})$ is in general not quadratic as in (2.7), but in many cases is still convex in $\boldsymbol{x}$.

2.3.1 Zero-forcing

If $f(\boldsymbol{x})$ is convex, then the minimizer in $\mathbb{C}^{n_\mathrm{t}}$, if exists, should satisfy

$$\nabla f(\hat{\boldsymbol{x}}_\mathrm{S}) = \mathbf{0}. \tag{2.13}$$

This stationary point can be found either by gradient descent or in some cases analytically. And we can define a generalized zero-forcing solution as

$$\hat{\boldsymbol{x}}_\mathrm{GZF} = [\hat{\boldsymbol{x}}_\mathrm{S}]_{\mathcal{X}}. \tag{2.14}$$

2.3.2 Sphere decoding

If $f(\boldsymbol{x})$ is convex, then the minimizer in $\mathbb{C}^{n_\mathrm{t}}$, if exists, should also satisfy

$$\nabla^2 f(\hat{\boldsymbol{x}}_\mathrm{S}) \geq \mathbf{0}. \tag{2.15}$$

where ∇^2 is the Hessian. Then, around $\hat{xv}_\mathrm{S}$, one can have the quadratic approximation

$$f(\boldsymbol{x}) = f(\hat{\boldsymbol{x}}_\mathrm{S}) + \frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{x}}_\mathrm{S})^H \left[\nabla^2 f(\hat{\boldsymbol{x}}_\mathrm{S})\right](\boldsymbol{x}-\hat{\boldsymbol{x}}_\mathrm{S}) + o(\|\boldsymbol{x}-\hat{\boldsymbol{x}}_\mathrm{S}\|^2) \tag{2.16}$$

so that the minimization of $f(\boldsymbol{x})$ can be approximated by

$$\min_{\boldsymbol{x}\in\mathcal{X}}(\boldsymbol{x}-\hat{\boldsymbol{x}}_\mathrm{S})^H \left[\nabla^2 f(\hat{\boldsymbol{x}}_\mathrm{S})\right](\boldsymbol{x}-\hat{\boldsymbol{x}}_\mathrm{S}) = \min_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{A}\hat{\boldsymbol{x}}_\mathrm{S} - \boldsymbol{A}\boldsymbol{x}\|^2, \tag{2.17}$$

where $\boldsymbol{A} \in \mathbb{C}^{m\times n_\mathrm{t}}$ such that $\boldsymbol{A}^H\boldsymbol{A} = \nabla^2 f(\hat{\boldsymbol{x}}_\mathrm{S})$, e.g., the Cholesky decomposiiton. Then, one can simply solve the optimization problem with sphere decoding as for conventional MIMO systems.

2.4 Applications

2.4.1 Application 1: MIMO channel with one-bit ADC

Now, let us consider $\boldsymbol{y} = \mathrm{sign}(\boldsymbol{H}\boldsymbol{x}+\boldsymbol{z})$ where $\boldsymbol{H} \in \mathbb{R}^{n_\mathrm{r}\times n_\mathrm{t}}$ is a known real matrix, $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I}_{n_\mathrm{r}})$, and $\boldsymbol{x} \in M$-PAM. In this case, the negative log-likelihood function is

$$f(\boldsymbol{x}) = -\sum_{i=1}^{n_\mathrm{r}} \log(Q(-y_i\boldsymbol{h}_i^T\boldsymbol{x}/\sigma)), \tag{2.18}$$

where $\boldsymbol{h}_i^{\mathrm{T}}$, $i \in [n_{\mathrm{r}}]$, is the i-th row of the channel matrix $\boldsymbol{H}$; $Q(u) := \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-t^2/2}\, dt$ is the normal Q function. For convenience, we define

$$u_i := -y_i\, \boldsymbol{h}_i^T \boldsymbol{x}/\sigma, \quad i \in [n_{\mathrm{r}}]. \tag{2.19}$$

Then, the objective function becomes $f(\boldsymbol{x}) = -\sum_{i=1}^{n_r} \log\Big(Q(u_i)\Big)$. Differentiating, we obtain the gradient

$$\boxed{\nabla f(\boldsymbol{x}) = -\sum_{i=1}^{n_r} \frac{y_i}{\sigma} \frac{\phi(u_i)}{Q(u_i)}\, \boldsymbol{h}_i}$$

where $\phi(u)$ is the standard normal pdf. Next, we differentiate the gradient to obtain the Hessian. It follows that

$$\boxed{\nabla^2 f(\boldsymbol{x}) = \sum_{i=1}^{n_r} \frac{\phi(u_i)}{\sigma^2 Q(u_i)^2} \Big(\phi(u_i) - u_i\, Q(u_i)\Big)\, \boldsymbol{h}_i\, \boldsymbol{h}_i^T = \boldsymbol{H}^{\mathrm{T}} \boldsymbol{D} \boldsymbol{H}}$$

where $\boldsymbol{D} := \operatorname{diag}\Big(\frac{\phi(u_i)}{\sigma^2 Q(u_i)^2}\Big(\phi(u_i) - u_i\, Q(u_i)\Big),\ i \in [n_{\mathrm{r}}]\Big)$. Note that $uQ(u) < \phi(u)$. This is obvious when $u \le 0$. When $u > 0$, we have $uQ(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} u e^{-t^2/2}\, dt < \frac{1}{\sqrt{2\pi}} \int_u^{\infty} t e^{-t^2/2}\, dt = \phi(u)$. Therefore, the Hessian is positive definite for any $\boldsymbol{x}$.

In this case, we can apply the sphere decoding (2.17) with $\boldsymbol{A} := \sqrt{\boldsymbol{D}}\boldsymbol{H}$.

2.4.2 Application 2: Channel estimation error

Assume that $\boldsymbol{H} = \hat{\boldsymbol{H}} + \tilde{\boldsymbol{H}}$ where $\hat{\boldsymbol{H}}$ is known estimate and $\tilde{\boldsymbol{H}}$ is estimation noise with i.i.d. $\mathcal{CN}(0, \sigma^2)$ entries. Then, the MIMO channel becomes

$$\boldsymbol{y} = \hat{\boldsymbol{H}}\boldsymbol{x} + \tilde{\boldsymbol{H}}\boldsymbol{x} + \boldsymbol{z}, \tag{2.20}$$

and the negative log-likelihood function becomes

$$f(\boldsymbol{x}) = n_{\mathrm{r}} \log \pi + n_{\mathrm{r}} \log(1 + \sigma^2 \|\boldsymbol{x}\|^2) + \frac{\|\boldsymbol{y} - \hat{\boldsymbol{H}}\boldsymbol{x}\|^2}{1 + \sigma^2 \|\boldsymbol{x}\|^2} \tag{2.21}$$

We consider the negative log-likelihood function expressed in terms of the real representation

$$\tilde{\boldsymbol{x}} := \begin{pmatrix} \mathrm{Re}\{\boldsymbol{x}\} \\ \mathrm{Im}\{\boldsymbol{x}\} \end{pmatrix} \in \mathbb{R}^{2n_{\mathrm{t}}},$$

and let $\tilde{\boldsymbol{y}} \in \mathbb{R}^{2n_{\mathrm{r}}}$ and $\hat{\boldsymbol{H}}_r \in \mathbb{R}^{2n_{\mathrm{r}} \times 2n_{\mathrm{t}}}$ be the real-lifts of $\boldsymbol{y} \in \mathbb{C}^{n_{\mathrm{r}}}$ and $\hat{\boldsymbol{H}} \in \mathbb{C}^{n_{\mathrm{r}} \times n_{\mathrm{t}}}$, respectively. We define

$$A(\tilde{\boldsymbol{x}}) = 1 + \sigma^2 \|\boldsymbol{x}\|^2, \qquad \boldsymbol{s}(\tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{y}} - \hat{\boldsymbol{H}}_r \tilde{\boldsymbol{x}}\,.$$

Then the objective function is given by

$$\boxed{F(\tilde{\boldsymbol{x}}) = n_{\mathrm{r}} \log\Big(1 + \sigma^2 \|\tilde{\boldsymbol{x}}\|^2\Big) + \frac{\|\tilde{\boldsymbol{y}} - \hat{\boldsymbol{H}}_r \tilde{\boldsymbol{x}}\|^2}{1 + \sigma^2 \|\tilde{\boldsymbol{x}}\|^2}\,.}$$

Its gradient with respect to $\tilde{\boldsymbol{x}}$ can be written as

$$\boxed{\nabla F(\tilde{\boldsymbol{x}}) = \frac{2}{A(\tilde{\boldsymbol{x}})^2} \Big[A(\tilde{\boldsymbol{x}})\Big(n_{\mathrm{r}}\sigma^2\, \tilde{\boldsymbol{x}} - \hat{\boldsymbol{H}}_r^T \boldsymbol{s}(\tilde{\boldsymbol{x}})\Big) - \sigma^2 \|\boldsymbol{s}(\tilde{\boldsymbol{x}})\|^2\, \tilde{\boldsymbol{x}}\Big].}$$

And the Hessian is

$$\boxed{\begin{aligned} \nabla^2 F(\tilde{\boldsymbol{x}}) = \frac{2}{A}\Big(n_{\mathrm{r}}\sigma^2 \boldsymbol{I} + \hat{\boldsymbol{H}}_r^T \hat{\boldsymbol{H}}_r\Big) - \frac{2\sigma^2 \|\boldsymbol{s}\|^2}{A^2} \boldsymbol{I} \\ + \frac{4\sigma^2}{A^2}\Big[\tilde{\boldsymbol{x}}\Big(\boldsymbol{s}^T \hat{\boldsymbol{H}}_r\Big) + \hat{\boldsymbol{H}}_r^T \boldsymbol{s}\, \tilde{\boldsymbol{x}}^T\Big] \\ + \Big(-\frac{4n_{\mathrm{r}}\sigma^4}{A^2} + \frac{8\sigma^4 \|\boldsymbol{s}\|^2}{A^3}\Big) \tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T. \end{aligned}}$$