Information Theory with Applications, Math6397
Lecture Notes from November 11, 2014

taken by Kelley Walker

Last Time

Gaussian Channels as the"worst" additive noise.

Partially noisy channels

Capacity for parallel AWGN channel, water filling principle

Capacity for parallel AWGN channe (continued)

Last time we saw

$$I(X;Y) \leq \sum_{j=1}^{k}(I_j;Y_j) \leq \sum_{j=1}^{k}\frac{1}{2}\ln\left(1+\frac{s_j}{\sigma_j^2}\right). \tag{1}$$

Recall that given k channels, additive Gaussian white noise of variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$, the power constraint S given by $S = \sum_{j=1}^{k} s_j$ is fixed.

We wish to maximize the RHS of (1) subject to the power constraint S as above. To accomplish this we use Lagrange Multipliers. Specifically, we wish to find

$$\max_{\substack{s_j \geq 0 \\ \lambda \in \mathbb{R}}}\left\{\sum_{j=1}^{k}\frac{1}{2}\ln\left(1+\frac{s_j}{\sigma_j^2}\right)+\lambda\left(\sum_{j=1}^{k}s_j - S\right)\right\}.$$

Whenever $s_j > 0$ we may take derivatives:

$$\frac{\partial}{\partial s_i}\left\{\sum_{j=1}^{k}\frac{1}{2}\ln\left(1+\frac{s_j}{\sigma_j^2}\right)+\lambda\left(\sum_{j=1}^{k}s_j - S\right)\right\} = 0+\cdots+\frac{1}{2\sigma_i^2}\left(\frac{1}{1+\frac{s_i}{\sigma_i^2}}\right)+\lambda+\cdots+0.$$

So if $\frac{1}{2\sigma_i^2}\left(\frac{1}{1+\frac{s_i}{\sigma_i^2}}\right)+\lambda = 0$ then $\frac{1}{2} = -\lambda(\sigma_i^2+s_i)$. Since $\sigma_i^2+s_i > 0$ for each i it follows that $\lambda < 0$. Moreover, since $\sigma_i^2+s_i$ is constant, denote it θ so that $s_i = \theta - \sigma_i^2$ whenever $s_i > 0$. Then choose θ such that $S = \sum_{j=1}^{k} s_j$. To accommodate all indices let $s_j = \max_{j=1,\dots,k}\{\theta - \sigma_j^2, 0\}$.

Alternatively, Since $\sum(1/2)\ln(1+(s_j/\sigma_j^2))$ is concave the sum is maximized when $s_j+\sigma_j^2$ has a constant value for all j for which $s_j>0$. This is precisely the construction of θ as above.

To conclude the argument, observe the maximum can be achieved for $I(X;Y)$ in general by choosing X to have $N(\mu,\sigma^2)$ distribution. Thus the capacity becomes

$$C(S)=\sum_{j=1}^{k}\frac{1}{2}\ln\left(1+\frac{s_j}{\sigma_j^2}\right).$$

5.7 Matrix Theory and Linear Algebra Review

We wish to answer the question "What happens when noise is correlated?". To do so, we need some results from matrix theory and linear algebra. Specifically, we will extend notions of convexity to operators.

5.7.27 Theorem. *Let $\mathcal{H}$ be a Hilbert Space. Suppose $x\in\mathcal{H}$ with $\|x\|=1$. Let f be convex functions and suppose that A is a bounded Hermitian operator. Then*

$$f(\langle Ax,x\rangle)\le\langle f(A)x,x\rangle.$$

Proof. Consider the spectral family $\{\mathcal{F}_\lambda\}_{\lambda\in\mathbb{R}}$ where each $\mathcal{F}_\lambda$ is strongly right-continuous and $\mathcal{F}_\lambda\to I$ as $\lambda\to\infty$ so that

$$A=\int_{\mathbb{R}}\lambda\,d\mathcal{F}_\lambda.$$

Then by continuity of f,

$$f(A)=\int_{\mathbb{R}}f(\lambda)\,d\mathcal{F}_\lambda$$

and

$$\langle Ax,x\rangle=\int_{\mathbb{R}}\lambda\,d\langle\mathcal{F}_\lambda x,x\rangle.$$

Let $\langle\mathcal{F}_\lambda x,x\rangle=\mu_x(\lambda)$. Then

$$\langle(f(A)x),x\rangle=\int_{\mathbb{R}}f(\lambda)\,d\mu_x(\lambda).$$

with spectral probability measure μ_x. So by Jensen's Inequality

$$\langle Ax,x\rangle=\int\lambda\,d\mu_x(\lambda)\le\int f(\lambda)\,d\mu_x(\lambda)=\langle f(A)x,x\rangle,$$

as desired. □

5.7.28 Example. If A is Hermitian and compact then $A=\sum_{n=1}^{\infty}\lambda_nP_n$ where each P_n is a projection and $\sum_{n=1}^{\infty}P_n=I$. (This follows from the spectral theorem for compact operators.) For convex f and x satisfying $\|x\|=1$ we have

$$f(\langle Ax,x\rangle)=f\left(\sum_{n=1}^{\infty}\lambda_n\langle P_nx,x\rangle\right).$$

Yet, $\langle P_n(x), x\rangle = \|P_n(x)\|^2 = P_n$, all $P_n \geq 0$, and $\sum P_n = 1$ so P_n's are all probabilities.
By Jensen's Inequality,

$$f\left(\langle Ax, x\rangle\right) = f\left(\sum_{j=1}^{\infty} \lambda_n \langle P_n x, x\rangle\right) \overset{J}{\leq} \sum_{n=1}^{\infty} f(\lambda_n) P_n = \left\langle \sum f(\lambda_n) P_n x, x \right\rangle.$$

5.7.29 Corollary. *Given a compact operator A with strictly positive eigenvalues and an orthonormal basis $\{e_j\}_{j=1}^{n}$ such that $\{\langle Ae_n, e_n\rangle\}$ is log summable, we have*

$$\mathit{Tr}[\ln A] \leq \sum_{j=1}^{\infty} \ln \langle Ae_j, e_j\rangle < \infty.$$

5.7.30 Corollary (Hadamard's Inequality). *If $\mathcal{H}$ is d-dimensional then for any positive definite operator A we have $\det A \leq \prod_{j=1}^{d} A_{jj}$.*

Proof. If A is singular, then the inequality is trivial, because the diagonal values $A_{i,i} \geq 0$, are by $A_{i,i} = \langle Ae_i, e_i\rangle = \sum_{j=1}^{d} \lambda_j \|P_j e_i\|^2$ convex combinations of all (non-negative) eigenvalues, thus the right-hand side is non-negative and the left hand side is zero. Thus, we can assume A is non-singular. Denote the eigenvalues of A by $\lambda_1, \lambda_2, \ldots, \lambda_d$. Then

$$\ln(\det A) = \sum_{j=1}^{d} \ln(\lambda_j + \epsilon) = \mathsf{Tr}\left(\ln A\right) \leq \sum_{j=1}^{d} \ln A_{jj} = \ln \prod_{j=1}^{d} A_{jj}.$$

Hence $\det A \leq \prod_{j=1}^{d} A_{jj}$. □

Correlated Noise

Suppose noise variable $\{N_1, N_2, \ldots, N_k\}$ have $\mu = 0$ for all N_i and covariance matrix C_N given by $(C_N)_{ij} = \mathbb{E}[N_i N_j]$. Since C_N is a real valued and symmetric matrix it is Hermitian. The input variables $\{X_1, \ldots, X_k\}$ have $\mu = 0$ and covariance C_X. The eigenvalues are variance in the direction of the associated eigenbasis vector. This motivates the power constraint in the following theorem.

5.7.31 Theorem. *Given k parallel channels with noise $\{N_1, N_2, \ldots, N_k\}$ as described above and power constraint $\mathit{Tr}[C_X] = S$, then the capacity is given by*

$$C(S) = \sum_{j=1}^{k} \frac{1}{2} \ln\left(1 + \frac{s_i}{\sigma_i^2}\right)$$

where σ_i^2 are the eigenvalues of C_N and s_i is as before: $s_i = \max\{\theta - \sigma_i^2, 0\}$ with $\theta = s_i + \sigma_i^2$ such that $\sum_{j=1}^{k} s_j = S$.

Proof. We wish to maximize $I(X;Y)$. Since $I(X;Y) = h(Y) - h(Y|X)$ where $Y = N + X$ we have $I(X;Y) = h(Y) - h(N)$ for fixed N. Then the problem reduces to maximizing Y.
The covariance matrix of Y$Y = X + N$ is, by independence, $C_Y = C_X + C_N$ because

$$\mathbb{E}[(X_i + N_i)(X_j + N_j)] = \mathbb{E}[X_i X_j] + \mathbb{E}[N_i N_j].$$

Then differential entropy is bounded by the Gaussian:

$$h(Y) \leq \frac{1}{2} \ln \left((2\pi e)^k |C_X + C_N|\right)$$

with equality when Y is Gaussian.
Since C_N is Hermitian, there exists an orthonormal basis for which C_N is diagonal. Let $\mathcal{O}$ be an orthonormal basis such that $D = \text{diag}\,(\sigma_1^2, \ldots, \sigma_k^2)$ and $C_N = \mathcal{O}^t D \mathcal{O}$. Then

$$\det(C_X + C_N) = \det(C_X + \mathcal{O}^t D \mathcal{O}) = \det(C_X + \mathcal{O}^t D \mathcal{O}) \det(\mathcal{O}^t \mathcal{O})$$

$$= \det((C_X + \mathcal{O}^t D \mathcal{O})(\mathcal{O}^t \mathcal{O})) = \det(\mathcal{O}(C_X + \mathcal{O}^t D \mathcal{O})\mathcal{O}^t)$$
$$= \det(\mathcal{O} C_X \mathcal{O}^t + D).$$

Note that $\text{Tr}\,[\mathcal{O} C_X \mathcal{O}^t] = \text{Tr}\,[C_X]$. Let $\mathcal{O} C_X \mathcal{O}^t = A$. We wish to maximize $\det[A + D]$ subject to the constraint $\text{Tr}\,[A] \leq S$. Hadamard's inequality gives

$$\det[A + D] \leq \pi_{j=1}^k (A_{i,i} + \sigma_i^2).$$

To achieve equality let $A_{i,i} = \max\{\theta - \sigma_i^2, 0\}$ with $\sum A_{ii} = S$. The upper bound is achieved when A is diagonal with entries A_{ii} so C_X and C_N are simultaneously diagonalizable and eigenvalues s_i, σ_i^2 satisfy the relationship as in the case of independent noise components.

□