Information Theory with Applications, Math6397
Lecture Notes from November 13, 2014

taken by Sabrine Assi

Recall from last time we discussed

- Capacity of AWGN
- Channel with correlated noise
- Treat eigenvalues of covariance matrix like variances in independent case

Warm up : Matrix Theory
We need to prove this statement:

5.7.32 Proposition

For any n×k matrix T over $\mathbb{R}$ det $(I_n + TT^t) =$ det $(I_k + T^tT)$.

Proof. Consider $\varepsilon \geq 0$. We want to show det $(I_n + \varepsilon TT^t) = det\,(I_k + \varepsilon T^tT)$ so if $\varepsilon = 0$ the statement is proven because the LHS=RHS=1.

Now consider $\varepsilon > 0$. First we want to define the functions $f_n(\varepsilon) = det\,(I_n + \varepsilon TT^t)$ and $f_k(\varepsilon) = det\,(I_k + \varepsilon T^tT)$ then we will compute the derivative of f_n at 0.

$f_n'(0) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,|_{\varepsilon=0} \det\,(I_n + \varepsilon TT^t)$
$= Tr[TT^t]$
$= Tr[T^tT]$ by elementary matrix production also
$= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\,|_{\varepsilon=0} \det\,(I_k + \varepsilon T^tT)$ which satisfies the same differential equation as f_k at $t = 0$.

More generally, if we choose A to be the inverse $A = (I_n + \tilde{\varepsilon}TT^t)^{-1}$ then

$$f_n(\tilde{\varepsilon} + \varepsilon) = \det(I_n + \tilde{\varepsilon}TT^t + \varepsilon TT^t) = \det(A^{-1})\det(I_n + A\varepsilon TT^t)$$

and taking the derivative at $\epsilon = 0$ gives

$$f_n'(\tilde{\varepsilon}) = \frac{d}{d\varepsilon}|_{\varepsilon=0}\det(I_n + \tilde{\varepsilon}TT^t + \varepsilon TT^t) = \det(I_n + \tilde{\varepsilon}TT^t)\mathrm{tr}[(I_n + \tilde{\varepsilon}TT^t)^{-1}\varepsilon TT^t]$$

We note that by equating terms in the power series expansion,

$$\mathrm{tr}[(I_n + \tilde{\varepsilon}TT^t)^{-1}\varepsilon TT^t] = \mathrm{tr}[(I_k + \tilde{\varepsilon}T^tT)^{-1}\varepsilon T^tT]$$

thus reversing the steps gives $f'_n(\tilde{\varepsilon}) = f_n(\tilde{\varepsilon})g(\tilde{\varepsilon})$ and $f'_k(\tilde{\varepsilon}) = f_k(\tilde{\varepsilon})g(\tilde{\varepsilon})$ with the same g.

By the uniqueness of the solution to ordinary differential equations with Lipschitz-continuous expressions for the derivative, we see that f_k and f_n coincide. □

5.7.33 Theorem

For a fixed n× k matrix H (linear model coding) and $\mathbb{R}^n$ -valued zero mean Gaussian Random Variable N with covariance

$$\mathbb{E}\left[N_i N_j\right] = \delta_{i,j}$$

The capacity of γ, $\gamma(X)$= HX+N subject to the input constraint $\sum_{j=1}^{k} \mathbb{E}\left[X_j^2\right] \leq S$ is

$$C(S) = \sum_{j=1}^{k} \left(\ln\left(\mu\lambda_j\right)\right)^+$$

where $\{\lambda_j\}_{j=1}^k$ are eigenvalues of $H^t H$ and μ is chosen such that $\sum_{j=1}^{k} \left(\mu - \lambda_j^{-1}\right)^+ = S.$

Proof. Based on Mutual Information, as usual $I(X;Y)$= $h(Y)$ -$h(N)$. So we need to maximize $h(Y)$. Using short hand notation $X = (x_1, x_2, ...x_k)^t$ $C_X = \mathbb{E}\left[XX^t\right]$ because $(C_X)_{i,j} = \mathbb{E}\left[X_i X_j\right]$. By the linearity of $X \to HX$ we have

$$\begin{aligned}(C_X)_{ij} &= \mathbb{E}\left[(HX+N)_i\,(HX+N)_j\right]\\ &= \mathbb{E}\left[(HX)_i\,(HX)_j\right] + \delta_{i,j}\\ C_Y &= HC_XH^t + I\end{aligned}$$

Thus, we have an upper bound for $h(Y)$ by the differential entropy of a corresponding Gaussian with the same covariance matrix, $h(Y) \leq \frac{1}{2}ln\left((2\pi e)^n det(I + HC_XH^t)\right)$ and this bound can be achieved by choosing X Gaussian, which implies that Y is Gaussian.
For a given H take SVD(Singular Value Decomposition) such that:

$$H = \mathcal{O}DW$$

where W is isometry, $\mathcal{O}$ is orthogonal, D is diagonal and D contains entries $\left\{\sqrt{\lambda_i}\right\}_{i=1}^n$ where λ_i are non negative eigenvalues of $HH^t = \mathcal{O}D^2\mathcal{O}^t$ where D^2 has diagonal matrix λ_i and 0 elsewhere.
Using the orthogonal invariance of determinants, we then have $h(Y)$= $\frac{1}{2}ln\left((2\pi e)^n det(I + DWC_XW^tD)\right)$.
Next we factor $C_X = VV^t$ by positive definiteness of C_X. So reordering in the determinant as shown above results in $det(I_n + DWC_XW^tD) = det(I_n + DWVV^tW^tD)$

$$= det(I_k + V^tW^tD^2WD)\,.$$

To continue, we choose an SVD of V such that V= $\mathcal{O}'\Delta\mathcal{O}''$ where Δ is the diagonal matrix and again after using the invariance of determinants we apply Hadamard's inequality, $det(I_k + V^tW^tD^2WV) = det(I_k + \Delta W^tD^2W\Delta)$

$$\leq \prod_{j=1}^{k} \left(1 + \underbrace{(\Delta W^tD^2W\Delta)_{j,j}}_{\Delta_{j,j}^2 (W^tD^2W)_{j,j}}\right)$$

So Equality holds in Hadamard's Inequality if and only if WD^2W^t is diagonal. To achieve equality we let VV^t have same eigenbasis such as $W^tD^2W = H^tH$.
Next we wish to maximize

$$\prod_{j=1}^{k}\left(1+\underbrace{(\Delta^2)_{j,j}}_{eigenvalue\ of C_X}\underbrace{(W^tD^2W)_{j,j}}_{eigenvalue\ of H^tH}\right)$$

subject to

$$\sum_{j=1}^{k}\left(\Delta^2\right)_{j,j}=Tr[C_X]\leq S$$

The optimal choice of C_X gives the Euler-Lagrange Equation :

$$\lambda_j(1+(VV^t)_{j,j}\lambda_j)^{-1}=\begin{cases}\mu,(VV^t)_{j,j}>0\\ \lambda_j, \qquad \text{else}\end{cases}$$

So

$$(VV^t)_{j,j}=\begin{cases}\frac{1}{\mu}-\frac{1}{\lambda_j}(VV^t)_{j,j}>0\\ 0, \qquad \text{else}\end{cases}$$

and μ is chosen such that $\sum_{j=1}^{k}(VV^t)_{j,j}=S.$
To end this the Hadamard's inequality achieved if and only if the matrix is diagonal matrix. Thus, the upper bound is achieved if the matrix is diagonal with corresponding entries such that Euler Lagrange equations are satisfied. □