Seminar 5

Info Theory - Notes Sept. 2007
July 2009

Gelfand - Pinsker
Costa.

Note: If $(x^n, y^n, z^n) \in A_\epsilon^{(n)}$ then $(x^n, y^n) \in A_\epsilon^{(n)}$, $(y^n, z^n) \in A_\epsilon^{(n)}$. BUT the converse is not true

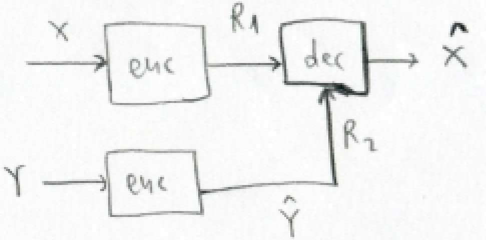
$(x^n, y^n) \in A_\epsilon^{(n)}$ and $(y^n, z^n) \in A_\epsilon^n$ does not in general imply that $(x^n, y^n, z^n) \in A_\epsilon^{(n)}$.

Lemma: Let $X \to Y \to Z$, i.e., $p(x,y,z) = p(x,y) \cdot p(z|y)$. If for a given $(y^n, z^n) \in A_\epsilon^{(n)}$, X^n is drawn $\sim \prod_{i=1}^n p(x_i|y_i)$, then $\Pr\{(X^n, y^n, z^n) \in A_\epsilon^{(n)}\} > 1-\epsilon$ for sufficiently large n.

Remark: Theorem is true if $X^n \sim \prod_{i=1}^n p(x_i|y_i, z_i)$. $X \to Y \to Z$ is used to show that $X^n \sim \prod_{i=1}^n p(x_i|y_i)$ is sufficient for the same conclusion

Gelfand - Pinsker, Costa

Source coding with side information



Only X is to be recovered

If $R_2 > H(Y)$, Y can be described perfectly, and from S-W: $R_1 > H(X|Y)$.

If $R_2 = 0$, then $R_1 > H(X)$ necessary to describe X.

In general, we use $R_2 = I(Y, \hat{Y})$ bits to describe an approximate version of Y. Then, we need $R_1 > H(X|\hat{Y})$, to describe X given Y.

Theorem: $(X, Y) \sim p(x,y)$. If Y is encoded at rate R_2 and X at rate R_1, we can recover X with arbitrarily small probability of error iff
$$R_1 \geq H(X|U)$$
$$R_2 \geq I(Y;U)$$
for some $p(x,y)\,p(u|y)$, with $|\mathcal{U}| \leq |\mathcal{Y}| + 2$.

Achievability Proof:

- Fix $p(u|y)$. Calculate $p(u) = \sum_y p(y) \cdot p(u|y)$. $\boxed{\text{Also, fixed is } p(u,y) = p(y) \cdot p(u|y)}$
- Codebook: Generate 2^{nR_2} codewords $\underline{U}(w_2)$, iid $p(u)$.

  Randomly bin the X^n seqs into 2^{nR_1} bins. Let $B(i)$ the set of X^n-seqs alloted to bin i.
- Encode: - X sender sends bin i of X^n.

  - Y sender looks for s such that $(Y^n, U^n(s)) \in A_\epsilon^{*(n)}(Y, U)$. If there are more than one s, send least. Otherwise, $s = 1$.

Decode: Receiver looks for unique $X^n \in B(i)$ such that $(X^n, U^n(s)) \in A_\epsilon^{*(n)}(X,U)$. If there is none or more than one, declares error.

Probability of error:

Error sources:

1. (X^n, Y^n) not typical. Prob $< \epsilon$.

2. Y^n is typical but there does not exist a $U^n(s)$ in the codebook which is jointly typical with it. If $R_2 > I(Y;U)$, then probability of this event is very small. $\boxed{\text{Why?}}$

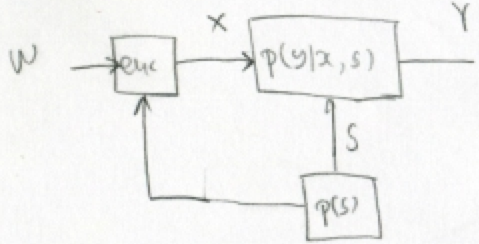
   Given Y^n and $p(u|y)$, we generate $U^n(s) \sim$ iid $p(u) = \sum_y p(y)p(u|y)$. Thus, $Y^n, U^n(i)$ independent with marginals those corresponding to $p(y,u)$. Thus, $P((Y^n, U^n(i)) \in A_\epsilon^{(n)}) \simeq 2^{-nI(Y,U)}$. If we generate 2^{nR_2} $U^n(i)$ with $R_2 > I(Y,U)$, then probability that Y^n j.t. with some of the $U^n(i)$ is close to 1.

3. $U^n(s)$ is j.t. with y^n but not with x^n. Since $X \to Y \to U$ forms a Markov chain, the probability of this event is small.

4. $\exists$ another $X^n \in B(i)$ j.t. with $U^n(s)$. Probability that any other X^n j.t. with $U^n(s)$ is $\simeq 2^{-nI(U;X)}$ and thus prob. of this kind of error is upper bounded by

$$|B(i) \cap A_\epsilon^{*(n)}(x)|\, 2^{-nI(X;U)} \leq \frac{2^{nH(X)}}{2^{nR_1}} 2^{-nI(X;U)}$$ which goes to 0 if $R_1 > H(X|U)$.

Gelfand-Pinsker coding:



$x \in \mathcal{X}$, $y \in \mathcal{Y}$, $s \in \mathcal{S}$.

$S_1, \dots, S_n$ known at the transmitter non-causaly.

$$p(s^n) = \prod_{i=1}^{n} p(s_i)$$

$$p(y^n|x^n, s^n) = \prod_{i=1}^{n} p(y_i|x_i, s_i).$$

$w \in I_M = \{1, \dots, M\}$, $M = 2^{nR}$.

We introduce an auxiliary r.v. U

We define the triple (U, S, X) with joint pdf $p(u,s,x)$ such that $\sum_{s,x} p(u,s,x) = p(s)$.

Quadruple: (U,S,X,Y), with joint pdf

$$p(u,s,x,y) = p(u,s,x) \cdot p(y|x,s)$$

For every triple $A = (U,S,X)$, we define

$$\boxed{R(A) = I(U;Y) - I(U;S)}$$

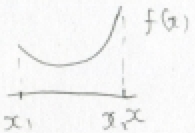
Let $\boxed{C = \max_{p(u,s,x)} R(A)}$ In fact, since $p(s)$ is given $\boxed{C = \max_{p(ux|s)} R(A).}$

Θεώρημα: C is the capacity of the Gelfand-Pinsker channel.

Useful proposition (i) For fixed $p(x|u,s)$, $R(A)$ is $\cap$-convex function of $p(u|s)$

(ii) For fixed $p(u|s)$, $R(A)$ is $\cup$-convex function of $P_{x|us}$.

Note $p(u,x|s) = p(u|s) \cdot p(x|u,s)$. Thus, for fixed? $p(u|s)$, $\max_{p(u,x|s)} = \max_{p(x|u,s)}$

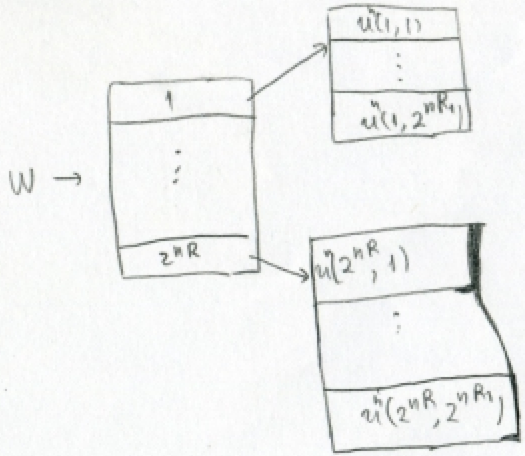


maximization of $\cup$-convex $f(x)$ over the closed set $[x_1, x_2]$ is achieved at an extreme point of the set.

In our case $\max_{p(\cdot|u,s)}$ is achieved at an extreme point of the form $[0 \cdots 0 1 0 \cdots 0]$.

Thus at max, $\exists$ $f(u,s)$ such that $x = f(u,s)$.

sketch of achievability: (1) Fix $p(u|s)$. Compute $p(u) = \sum_s p(s)\, p(u|s)$. Generate codewords iid from $p(u)$ and make them available at Tx, Rx.



(2) For any $s^n \in A_\epsilon^{(n)}(S)$ and $w \in \{1,\dots,2^{nR}\}$, look into the w-th bin for a codeword $u^n(w,*)$ jointly typical with s^n. If $\boxed{R_1 > I(U;S)}$, then we can find at least one such codeword with high prob.

(3) Given s^n and $u^n(w,j)$ jointly typical compute x^n such that $x_n = f(s_n, u(w,j,n))$, with f defined by $\max\limits_{p(x|u,s)}$

Note: x^n jointly typical with s^n, $u^n(w,j)$. what is the purpose of f?

(4) The output of the channel is y^n constructed by x^n and s^n from $p(y|x,s)$

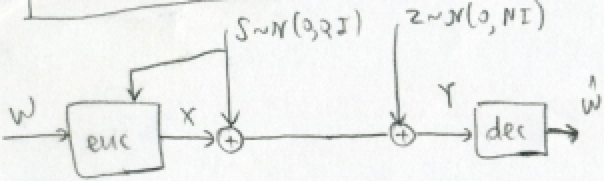
(5) Decoding: Look for $u^n(\cdot,\cdot)$ j.t. with y^n. If all such u^n's belong to the same bin $\hat{w}$, return $\hat{w}$. Otherwise, return error.

Probability of error: $P(\epsilon) = P\left[\, (u^n(w,j), y^n) \notin A_\epsilon^{(n)}(U,Y) \text{ or } \exists\, \hat{w} \neq w \text{ and } j \text{ such that } (u^n(\hat{w},j), y^n) \in A_\epsilon^{(n)}(U,Y) \right]$

$$\leq \epsilon + \sum_{\hat{w} \neq w, j} P\left[(u^n(\hat{w},j), y^n) \in A_\epsilon^{(n)}(U,Y) \right] \leq \epsilon + (2^{nR}-1)\, 2^{nR_1}\, 2^{-nI(U;Y)} \leq 2\epsilon$$

if n sufficiently large and $\boxed{R + R_1 < I(U;Y) \Rightarrow R < I(U;Y) - R_1 < I(U;Y) - I(U;S)}$

Costa coding.



If $P > Q$, we may use part of P to cancel S and the rest $P-Q$ to send information. Capacity is

$$\frac{1}{2}\log\left(1+\frac{P-Q}{N}\right).$$

In general, we may "partially cancel" Q. But this is not optimal.

We remind

$$C = \max_{p(u,x|s)} \{ I(U;Y) - I(U;S) \}.$$

problems: find U and $X = f(U,S)$.

Costa considered the case

$$U = X + \alpha S.$$

$X \sim \mathcal{N}(0,P)$, $S \sim \mathcal{N}(0,Q)$, X, S indep

There could be loss of generality, but it doesn't.

Recall, $Y = X + S + Z$. In order to compute $I(U;Y), I(U;S)$, we must compute $f(u,y)$, $f(u,s)$.

$$\begin{bmatrix} U \\ S \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P+\alpha^2 Q & \alpha Q \\ \alpha Q & Q \end{bmatrix} \right)$$

$$\begin{bmatrix} U \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P+\alpha^2 Q & P+\alpha Q \\ P+\alpha Q & P+Q+N \end{bmatrix} \right)$$

Then
$$I(U;Y) = \frac{1}{2}\ln \frac{(P+Q+N)(P+\alpha^2 Q)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}$$

$$I(U;S) = \frac{1}{2}\ln \frac{P+\alpha^2 Q}{P}.$$

Define
$$R(\alpha) = I(U;Y) - I(U;S) = \frac{1}{2}\ln \frac{P(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}$$

Maximizing $R(\alpha)$ over α, we obtain

$$\max_{\alpha} R(\alpha) = R(\alpha^*) = \frac{1}{2}\ln\left(1+\frac{P}{N}\right) = C^*, \quad \alpha^* = \frac{P}{P+N}.$$

If S were known to both Tx and Rx, the achievable capacity would be C^*.

Thus, the chosen U and input X achieve capacity.

Actual coding scheme.

① Generate $e^{nI(U;Y)}$ iid codewords $\sim \mathcal{N}(0, P+\alpha^{*2}Q)$, and distribute them into e^{nR} bins such that each bin contains the same number of seqs.

② Given $\underline{S}_0$ and message k, search in bin k to find $\underline{U}$ j.t. with $\underline{S}_0$.

Actually, this is equivalent to looking for a seq. U such that

$$|(\underline{U} - \alpha^* \underline{S}_0)^T \underline{S}_0| \leq \delta \quad \text{(for some small } \delta).$$

With high prob, we can find such a seq. Call it $\underline{U}_0$.

Encoder computes $\underline{X}_0 = \underline{U}_0 - \alpha^* \underline{S}_0$. With high prob. $\underline{X}_0$ will be typical, which says that $\frac{1}{n}\|\underline{X}_0\|^2 \leq P$. Encoder sends $\underline{X}_0$.

Σχέση (10) σε Costa dirty-paper

Ελέγχουμε αν δύο ακολουθίες x^n και y^n είναι από κοινού τυπικές υπολογίζοντας την απόλυτη τιμή

$$\left| -\frac{1}{n} \sum_{i=1}^{n} \ln p(x_i, y_i) - H(X,Y) \right|$$

όπου η εντροπία $H(X,Y)$ είναι υπολογισμένη βάσει της $p(x,y)$.

Για τη συγκεκριμένη σχέση, έχουμε ότι $U^n = X^n + aS^n$, που δίνει

$$\begin{bmatrix} U_i \\ S_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P + a^2 Q & aQ \\ aQ & Q \end{bmatrix} \right)$$

Έχουμε

$$H(U,S) = \frac{1}{2} \ln(2\pi e)^2 PQ = \frac{1}{2} \ln(2\pi)^2 PQ + 1.$$

και

$$\begin{aligned} \ln p(u_i, s_i) &= -\frac{1}{2} \ln(2\pi)^2 PQ - \frac{1}{2} [u_i \; s_i] \, C_{U,S}^{-1} \begin{bmatrix} u_i \\ s_i \end{bmatrix} \\ &= -\frac{1}{2} \ln(2\pi)^2 PQ - \frac{1}{2PQ} \left(Q(u_i - as_i)^2 + Ps_i^2 \right) \\ &= -\frac{1}{2} \ln(2\pi)^2 PQ - \frac{1}{2P} (u_i - as_i)^2 - \frac{1}{2Q} s_i^2 \end{aligned} \tag{7}$$

Συνεπώς

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^{n} \ln p(u_i, s_i) &= \frac{1}{2} \ln(2\pi)^2 PQ + \frac{1}{2Pn} (U^n - aS^n)^T (U^n - aS^n) + \frac{1}{2Qn} S^{nT} S^n \\ &\to \frac{1}{2} \ln(2\pi)^2 PQ + \frac{1}{2Pn} (U^n - aS^n)^T (U^n - aS^n) + \frac{1}{2}. \end{aligned} \tag{8}$$

Για να είναι τα U^n και S^n από κοινού τυπικά θα πρέπει

$$(U^n - aS^n)^T (U^n - aS^n) \approx Pn$$

Κάνοντας πράξεις, λαμβάνουμε

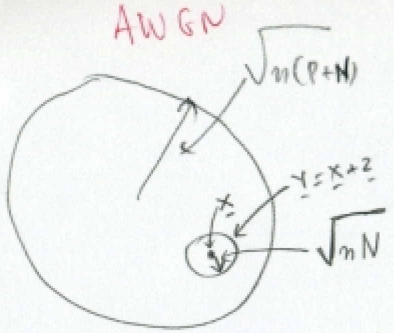
$$\begin{aligned} &U^{nT} U - a\,U^{nT} S^n - a\,S^{nT} U^n + a^2 S^{nT} S^n \approx Pn \implies \\ &(P + a^2 Q)n - 2a\,U^{nT} S^n + a^2 Qn \approx Pn \implies \\ &2a^2 Qn - 2aU^{nT} S^n \approx 0 \implies aQn - U^{nT} S^n \approx 0 \implies U^{nT} S^n \approx aQn. \end{aligned} \tag{9}$$

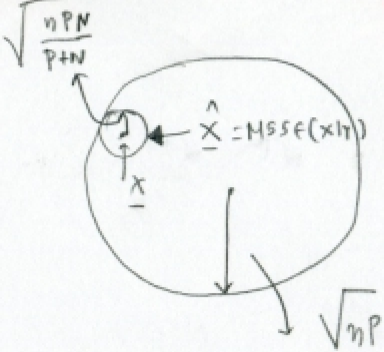
Οπότε

$$|(U^n - aS^n)^T S^n| \approx |U^{nT} S^n - aS^{nT} S^n| \approx |aQn - aQn| \approx 0. \tag{10}$$

Συνεπώς, η είσοδος στο σύστημα $X^n = U^n - aS^n$ είναι κάθετη στο S^n!

Costa Precoding from Tse





Detection in AWGN

$$\boxed{\hat{\underline{X}} = MMSE(X|Y) = \alpha Y = \frac{P}{P+N} Y}$$

$$(\hat{X}, X) \in A_\epsilon^{(n)}$$

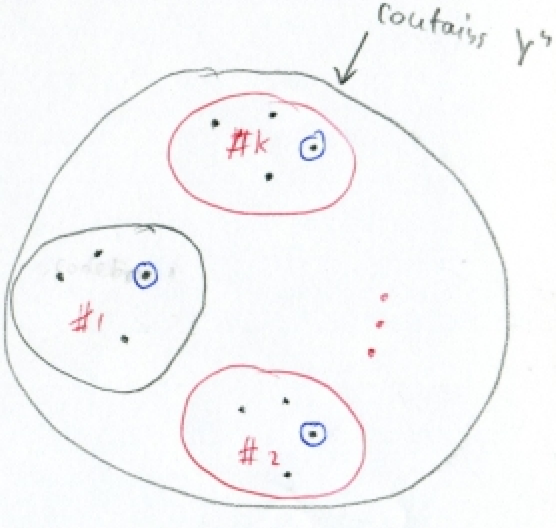
Costa and MMSE estimation.

$$Y^n = X^n + S^n + Z^n$$

Consider a domain $V \in R^N$ large enough for the received Y^n to lie inside

In this domain we replicate the basic codebook of M codewords K times

Each initial codeword corresponds to an equivalence class of points in R^N



o points in the same equivalence class.

- With known interference S^n, and given message to transmit, i, the Tx finds the extended codeword in $\mathcal{E}_i$ (equivalence class of i), $\underline{P}$, and transmits

$$\boxed{\underline{X}_1 = \underline{P} - \underline{S}}$$

- Based on $\underline{y}$, the decoder finds the point in the extended constellation that is closest to $\underline{y}$ and decodes to the info bits corresponding to the equivalence class.

Performance

To estimate the max-rate for given P we observe

- **Sphere packing:** To avoid confusing $\underline{x}_1$ with any other of the $K(M-1)$ points in the extended constellation, that belong to other equivalence classes the noise spheres of radius $\sqrt{N\sigma^2}$ around each point should be disjoint

$$KM < \frac{Vol(V)}{Vol(B_N(\sqrt{N\sigma^2}))} \quad (1)$$

- **Sphere covering:** To maintain transmit power $<P$, the quantization error should be no more than $\sqrt{NP}$ for $\forall \underline{s}$. Thus, the spheres of radius $\sqrt{NP}$ around the K replicas of a codeword should cover the whole domain. Thus

$$K > \frac{Vol(V)}{Vol(B_N(\sqrt{NP})} \quad (2)$$

$$(1),(2) \Rightarrow M < \frac{vol(B_N(\sqrt{NP}))}{vol(B_N(\sqrt{N\sigma^2}))} \Rightarrow R = \frac{\log_2 M}{N} = \frac{1}{2}\log\frac{P}{\sigma^2}.$$

suboptimal for finite P.

Performance enhancement via MMSE estimation

To meet the average power constraint, the density of the replication cannot be reduced beyond (2).

On the other hand, (1) is a direct consequence of the nearest neighbor decoding rule, and this is suboptimal for the problem at hand.

Let us use an estimate $\alpha\underline{y}$ of $\underline{x}_1$.

$$\alpha\underline{y} = \alpha(\underline{x}_1 + \underline{s} + \underline{z}) = \alpha(\underline{x}_1 + \underline{w}) + \alpha\underline{s} \stackrel{\Delta}{=} \underline{x}_{MMSE} + \alpha\underline{s}$$

where $\underline{x}_{MMSE}$ is the estimate of $\underline{x}_1$ from $\underline{y}$ assuming $\underline{s}=0$.

Since $\alpha\underline{s}$ is not known, it must be pre-subtracted.

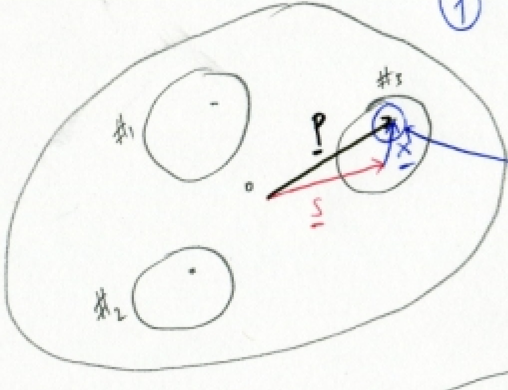
Let $\underline{x}_1 = \underline{p} - \alpha\underline{s}, \quad \underline{y} = \underline{x}_1 + \underline{s} + \underline{z} = \underline{p} - \alpha\underline{s} + \underline{s} + \underline{z}$

Then $\alpha\underline{y} = \hat{x}_{MMSE} + \alpha\underline{s}$

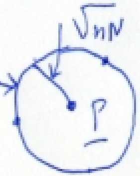
and $\underline{p} - \alpha\underline{y} = \underline{x}_1 - \underline{x}_{MMSE}$.

Receiver finds constellation point nearest to $\alpha\underline{y}$ and decodes. Error occurs only if there is another constellation point closer to $\alpha\underline{y}$ than $\underline{p}$. The MMSE radius is $\sqrt{nPN}/\sqrt{P+N}$. This gives capacity.

① Πρώτη προσέγγιση: Δεδομένου του $\underline{s}$, βρίσκουμε το κοντινότερο σημείο της κλάσης ισοδυναμίας $\underline{p}$ και κατόπιν το $\underline{x} = \underline{p} - \underline{s}$.

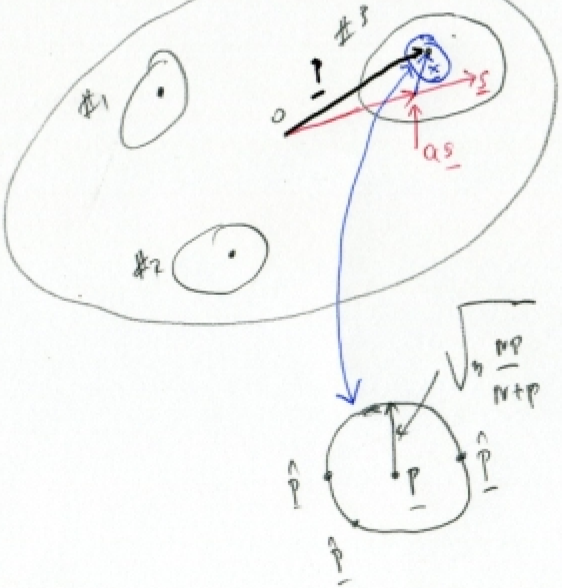


Λαμβάνουμε $\underline{y} = \underline{x} + \underline{s} + \underline{w} = \underline{p} + \underline{w}$. Αν $w_i \sim \text{iid } \mathcal{N}(0, N)$ τότε το $\underline{y}$ είναι κάποιο σημείο στην επιφάνεια σφαίρας με κέντρο το $\underline{p}$ και ακτίνα $\sqrt{nN}$.



② Δεύτερη προσέγγιση

Με δεδομένο $\underline{s}$, βρίσκουμε το κοντινότερο σημείο της κλάσης ισοδυναμίας στο $\alpha \underline{s}$, $\underline{p}$, και κατόπιν κατασκευάζουμε το $\underline{x} = \underline{p} - \alpha \underline{s}$.



Λαμβάνουμε $\underline{y} = \underline{x} + \underline{s} + \underline{w}$. Για αποκωδικοποίηση χρησιμοποιούμε το

$$\hat{\underline{p}} = \alpha \underline{y} = \alpha \underline{s} + \alpha(\underline{x} + \underline{w}) = \alpha \underline{s} + \underline{x}_{MMSE}.$$

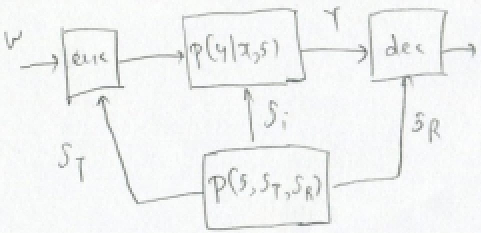
Άρα $\hat{\underline{p}} - \underline{p} = \alpha \underline{s} + \underline{x}_{MMSE} - \alpha \underline{s} - \underline{x} = \underline{x}_{MMSE} - \underline{x} = \underline{\varepsilon}_{MMSE}$.

Άρα, το $\hat{\underline{p}}$ βρίσκεται πάνω σε σφαίρα με κέντρο το $\underline{p}$ και ακτίνα $\sqrt{n\frac{NP}{N+P}}$. Αυτό οδηγεί σε

$$M \le \frac{A(\sqrt{nP})^n}{A\left(\sqrt{n\frac{PN}{P+N}}\right)^n} \Rightarrow R = \frac{1}{n}\log_2 M \le \frac{1}{2}\log\left(1 + \frac{P}{N}\right).$$

Capacity with (non)-causal CSI.

Jafar, IT, Dec. 2006.



$p(s^n, s_T^n, s_R^n) = \prod_i p(s_i, s_{T,i}, s_{R,i})$ iid states (memoryless)

$p(y^n|x^n,s^n) = \prod_i p(y_i|x_i,s_i)$

$$C^{noncausal} = \max_{P_{noncausal}} I(U;Y) - I(U;S)$$

$$P_{noncausal} = \{P(U,X|S_T) = P(U|S_T)\cdot P(X|U,S_T)\}$$

If $S_T = \phi$ (no Tx-side information)

$$C = \max_{p(U,X)} I(U;Y) = \max_{p(x)} I(X;Y)$$

Availability of TX-side info permits the Tx to match its input to channel state by picking U,X conditioned on S_T.

We remind that $C^{causal} = \max_{P_T(t)} I(T;Y)$

with T extended alphabet of mappings from channel state to input alphabet

Rx-side info can be incorporated by replacing Y with (Y,S_R).

Recent results have shown that

$$C^{noncausal} = \max_{P_{noncausal}} I(U;Y,S_R) - I(U;S_T)$$

$$C^{causal} = \max_{P_{causal}} I(U;Y,S_R) - \cancelto{0}{I(U;S_T)}$$

$$P_{noncausal} = \{P(U,X|S_T) = P(U|S_T)\cdot P(X|U,S_T)\}$$

$$P_{causal} = \{P(U,X|S_T) = P(U)\cdot P(X|U,S_T)\}$$

<u>Result:</u> If $S_T = f(S_R)$, $f(\cdot)$ deterministic, then capacity with causal side information <u>equals</u> capacity with non-causal side information

<u>Proof:</u>

$$C^{noncausal} = \max_{p(u|s_T)\cdot p(x|u,s_T)} I(U;Y,S_R) - I(U;S_T) = \max_{p(u|s_T)\cdot p(x|u,s_T)} I(U;S_R) + I(U;Y|S_R) - I(U;S_T)$$

$$= \max_{p(u|s_T)\cdot p(x|u,s_T)} I(U;S_R,S_T) + I(U;Y|S_R) - I(U;S_T) = \max_{p(u|s_T)\, p(x|u,s_T)} I(U;S_R|S_T) + I(U;Y|S_R)$$

$$= \max_{p(u|s_T)\, p(x|u,s_T)} I(U;Y|S_R) = \max_{p(x|s_T)} I(X;Y|S_R) = \max_{p(u)\cdot p(x|u,s_T)} I(U;Y|S_R)$$

- Capacity with causal and non-causal side info

Theorem: $C^{noncausal}(S_T, S_R) - C^{causal}(S_T, S_R) \leq H(S_T | S_R)$