Logistics

Class Road Map - IT-I

- L19 (1/6): Overview, Communications, Gaussian Channel
- L20 (1/8): Gaussian Channel, band limitation, parallel channels, optimization and duality
- L21 (1/13): parallel channels, colored noise, feedback, matrix inequalities
- L22 (1/15):
  - (1/20): Monday holiday
- L23 (1/22):
- L24 (1/27):
- L25 (1/29):
- L26 (2/3):
- L27 (2/5):

- L28 (2/10):
- L29 (2/12):
  - (2/17): Monday, Holiday
- L30 (2/19):
- L31 (2/24):
- L32 (2/26):
- L33 (3/3):
- L34 (3/5):
- L35 (3/10):
- L36 (3/12):

Cumulative Outstanding Reading

- Read Ch. 10 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 17 in our book (Cover & Thomas, “Information Theory”) on matrix inequalities.
- Read Ch. 9 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 5 in Boyd and Vandenberghe’s Convex Optimization book.
- Read all readings assigned in EE514a, Fall 2013. (see later lectures on our previous web page (http://j.ee.washington.edu/~bilmes/classes/ee514a_fall_2013/)).

Homework

- No homework yet.
Office hours on Mondays, 3:30-4:30 but not today, 1/13.

As always, email me if you want to skype/google hangout rather than come to office hours, also at different times.

Parallel Channels

We are given $k$ independent Gaussian channels

- So the noises are uncorrelated since $Z_i \perp Z_j$ for $i \neq j$.
- Without constraints, capacity is $C = \log(\sum_i 2^{C_i})$ if we use one channel at a time, and $\sum_i C_i$ if we can use the channels simultaneously.
- With common power constraint (is common to all), i.e., with

$$E[\sum_{j=1}^k X_j^2] = \sum_{j=1}^k EX_j = \sum_i P_i \leq P$$

(21.22)

what is the capacity?
Parallel Channels

Goal: find maximum capacity in this case. I.e., find

\[ C = \max_{f(x_{1:k}) : \sum_i E X_i^2 \leq P} I(X_{1:k}; Y_{1:k}) \]  \hspace{1cm} (21.22)

We have the following:

\[ I(X_{1:k}; Y_{1:k}) = h(Y_{1:k}) - h(Y_{1:k}|X_{1:k}) = h(Y_{1:k}) - h(Z_{1:k}|X_{1:k}) \]  \hspace{1cm} (21.23)

\[ \leq \sum_j (h(Y_j) - h(Z_j)) \leq \sum_j \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \]  \hspace{1cm} (21.25)

with \( P_i = E X_i^2 \) and \( \sum_i P_i = P \).

The way to solve this is to solve

\[
\text{maximize} \quad \sum_j \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \\
\text{subject to} \quad \sum_i P_i = P 
\]  \hspace{1cm} (21.22)

\[
J(P_{1:n}) = \sum_j \frac{1}{2} \log(1 + \frac{P_i}{N_i}) + \lambda(\sum_j P_i - P) 
\]  \hspace{1cm} (21.24)
Lagrangian optimization

\[
\begin{align*}
\min_x & \quad f_0(x) \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ quadratic \ function.
\end{align*}
\]

subject to

\[
\begin{align*}
& f_i(x) \leq 0 \text{ for } i = 1, \ldots, m \\
& h_i(x) = 0 \text{ for } i = 1, \ldots, p
\end{align*}
\]

Lagrangian form

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\]

and define

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)
\]

Weak duality and duality gap

- Then the weak duality condition is if the max dual is no more than the min primal, i.e.,

\[
d_{opt} \leq p_{opt} \iff \text{weak duality}
\]

- Note, weak duality always holds since, for any \( x', \lambda', \nu' \) we have:

\[
g(\lambda', \nu') = \inf_x L(x, \lambda', \nu') \leq L(x', \lambda', \nu') \leq \sup_{\lambda>0, \nu} L(x', \lambda, \nu) \leq L^\square(x')
\]

- Hence,

\[
d_{opt} = \sup_{\lambda>0, \nu} \inf_x L(x, \lambda, \nu) \leq \inf_x \sup_{\lambda>0, \nu} L(x, \lambda, \nu) = \inf_x L^\square(x) = p_{opt}
\]

- From this, we can define the optimal duality gap

\[
p_{opt} - d_{opt} \geq 0 \text{ is the duality gap}
\]
Strong duality and zero duality gap

- Strong duality, tightness, zero gap. I.e., $p_{\text{opt}} = d_{\text{opt}}$.
- We like strong duality since it means that we can keep maximizing the dual and minimizing the primal, and if they meet we know we are optimal (or if they are close, we can bound quality of solution).
- Unfortunately, strong duality does not always hold, while weak duality does always hold.

Strong duality - Summary

- Notation: $\lambda_{\text{opt}} = \lambda^*$, $\nu_{\text{opt}} = \nu^*$
- Strong duality means $f(x_{\text{opt}}) = g(\lambda_{\text{opt}}, \nu_{\text{opt}})$
- If this holds and all functions differentiable, then we have the KKT necessary conditions for optimality, namely that:
  
  \[
  f_i(x_{\text{opt}}) \leq 0 \quad \text{for} \quad i = 1, \ldots, m \quad (21.53)
  \]
  \[
  h_i(x_{\text{opt}}) = 0 \quad \text{for} \quad i = 1, \ldots, p \quad (21.54)
  \]
  \[
  \lambda_i^* \geq 0 \quad \text{for} \quad i = i, \ldots, m \quad (21.55)
  \]
  \[
  \lambda_i^* f_i(x_{\text{opt}}) = 0 \quad \text{for} \quad i = 1, \ldots, m \quad (21.56)
  \]

  and

  \[
  \nabla_x L|_{x=x_{\text{opt}}, \lambda=\lambda_{\text{opt}}, \nu=\nu_{\text{opt}}} = 0 \quad (21.57)
  \]

- These conditions are necessary, they must be true at any optimum.
Back to our problem

- Consider again our problem from before, repeated again on the next slide.
- \( x = (P_1, P_2, \ldots, P_m) \) vector of power values.
- \( N_i \) noises given (before we called these \( \sigma_i^2 \)).
- We want to do

\[
\text{minimize } - \sum_{i=1}^{k} \log(1 + P_i/N_i) \quad (21.53)
\]

subject to inequality constraints \( P_i \geq 0 \) \( \forall i \) (so \( f_i(P_i) = -P_i \)), and equality constraints \( \sum_{i=1}^{k} P_i = P \) (i.e., \( h = (\sum_{j=1}^{n} P_i - P) = 0 \)).

- This problem is convex.
- Also, there exists a strictly feasible point, so strong duality holds, as do the KKT conditions for optimality.

Parallel Channels

- The way to solve this is to solve

\[
\text{maximize } \sum_{j} \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \quad (21.22)
\]

subject to \( \sum_{i} P_i = P \) \( (21.23) \)

- Or in Lagrangian form

\[
J(P_1, n) = \sum_{j} \frac{1}{2} \log(1 + \frac{P_i}{N_i}) + \lambda(\sum_{j} P_i - P) \quad (21.24)
\]
Back to our problem

- Consider again our problem from before, repeated again on the next slide.
- \( x = (P_1, P_2, \ldots, P_m) \) vector of power values.
- \( N_i \) noises given (before we called these \( \sigma_i^2 \)).
- We want to do

\[
\text{minimize} \ - \sum_{i=1}^{k} \log(1 + P_i/N_i) \quad (21.53)
\]

subject to inequality constraints \( P_i \geq 0 \ \forall i \) (so \( f_i(P_i) = -P_i \)), and equality constraints \( \sum_{i=1}^{k} P_i = P \) (i.e., \( h = (\sum_{j=1}^{n} P_i - P) = 0 \)).

- This problem is convex.
- Also, there exists a strictly feasible point, so strong duality holds, as do the KKT conditions for optimality.

We get Lagrangian

\[
L(x, \lambda, \nu) = -\sum_{i=1}^{k} \log(1 + P_i/N_i) - \sum_{i=1}^{k} \lambda_i P_i + \nu(\sum_{i=1}^{k} P_i - P) \quad (21.1)
\]

KKT conditions are:

\[
\forall i : P_i^* \geq 0, \quad \sum_{i} P_i^* = P, \quad \lambda_i^* \geq 0 \ \forall i, \quad \text{and} \quad \lambda_i^* P_i^* = 0 \quad (21.2)
\]

and also \( \forall i \)

\[
-\frac{1}{1 + P_i/N_i} - \frac{1}{N_i} = \lambda_i^* + \nu^* = 0 \quad (21.3)
\]
KKT Conditions

- From the Lagrangian gradient conditions we can further get
  
  \[- \frac{1}{N_i + P_i} - \lambda_i^* + \nu^* = 0 \quad (21.4)\]
  
  \[\Rightarrow \quad - \frac{1}{N_i + P_i} + \nu^* = \lambda_i^* \geq 0 \quad (21.5)\]

- We then eliminate \( \lambda_i^* \) (the slack variable) to get KKT conditions in form
  
  \[\forall i : P_i^* \geq 0, \quad \sum_i P_i^* = P \quad (21.7)\]
  
  \[
  \left( \nu^* - \frac{1}{N_i + P_i} \right) P_i^* = 0 \quad \nu^* \geq \frac{1}{N_i + P_i^*} \quad (21.8)
  \]

Cases

We have two cases

- Case 1: From condition \( \nu^* \geq \frac{1}{N_i + P_i^*} \), if \( \nu^* < 1/N_i \), then we must have \( P_i^* > 0 \) to achieve the condition.

- In such case, since \( \left( \nu^* - \frac{1}{N_i + P_i} \right) P_i^* = 0 \), we must have
  
  \[\nu^* = \frac{1}{N_i + P_i^*}.\]

- Hence, \( P_i^* = \frac{1}{\nu^*} - N_i \)

- Case 2: If \( \nu^* \geq 1/N_i \), then \( P_i^* = 0 \) since otherwise \( (P_i^* > 0) \) would mean
  
  \[
  \left( \nu^* - \frac{1}{N_i + P_i} \right) \times \frac{P_i^*}{>0} \geq 0 \quad (21.9)
  \]

  \[
  >0 \text{ since } \nu^* \geq 1/N_i \text{ and } P_i^* > 0
  \]
**Condition on $P_i$**

- So, $P_i^*$ must have form

\[
P_i^* = \begin{cases} 
\frac{1}{\nu^*} - N_i & \text{if } \nu^* < \frac{1}{N_i} \\
0 & \text{if } \nu^* \geq \frac{1}{N_i} 
\end{cases} \quad (21.10)
\]

\[
= \max\{0, \frac{1}{\nu^*} - N_i\} \quad (21.11)
\]

- With the last constraint, we have that

\[
\sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \quad (21.12)
\]

where $a^+ = \max\{0, a\}$.

- This leads to the famous water filling idea for parallel channels.

**Water Filling**

\[
\sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P \quad (21.13)
\]

- Think of this as gradually increasing $1/\nu^*$ until Equation 21.13 is satisfied.
Water Filling

\[
\sum_i \left( \frac{1}{\nu^*} - N_i \right)^+ = P
\]  

(21.14)

- We can think of this as filling up a bin with uniformly dispersed rain water and stopping the rain until we achieve equality in Eq. 21.21

- the red area indicated above is equal to \( P \).

Capacity of Parallel Channels

- So the final capacity is then

\[
C_n = \frac{1}{2} \sum_{j=1}^{k} \log(1 + P_i/N_i)
\]

(21.15)

\[
= \frac{1}{2} \sum_{j=1}^{k} \log \left( 1 + \frac{(1/\nu^* - N_i)^+}{N_i} \right) \text{ bits per parallel channel use}
\]

(21.16)

- In units of bits per transmission (bits per single channel transmission, take the average):

\[
C_n = \frac{1}{2n} \sum_{j=1}^{k} \log \left( 1 + \frac{(1/\nu^* - N_i)^+}{N_i} \right) \text{ bits per transmission}
\]

(21.17)
Before and After

- Before we had \( k \) independent Gaussian channels with a common power constraint and uncorrelated noise.
- Now, have colored noise, i.e.,
  \[
  Z_{1:k} \sim N(0, K_Z)
  \]
  where \( K_Z \) is not necessarily diagonal (i.e., noise is correlated from one time step to the next).
- We still assume \( X_{1:n} \perp Z_{1:n} \) (so noise and signal are independent).

Colored Noise

- Define \( K_X = EXX^\top \)
- We have power constraint of the form
  \[
  \frac{1}{n} \sum_i EX_i^2 = \frac{1}{n} \text{tr}(K_X) \leq P
  \]
- Let first form upper bound:
  \[
  I(X_{1:n}; Y_{1:n}) = h(Y_{1:n}) - h(Z_{1:n})
  \]
- and
  \[
  h(Y_{1:n}) \leq \frac{1}{2} \log \left( (2\pi e)^n |K_X + K_Z| \right) \text{ since } X \perp Z
  \]
Now, $K_Z$ is fixed, we wish to maximize $K_X$ in Equation 21.21

I.e., we want to do

$$\max_{K_X \succeq 0} |K_X + K_Z| \quad \text{subject to} \quad \frac{1}{n} \tr(K_X) \leq P \quad (21.22)$$

Note that $K_X \succeq 0$ is a positive semidefinite constraint

Now via singular value decomposition (spectral eigenvalue factorization) $K_Z = Q \Lambda Q^\top$ with $QQ^\top = I$ and det. $|Q| = 1$. Hence

$$|K_X + K_Z| = |K_X + Q \Lambda Q^\top| \quad (21.23)$$

$$= |Q||Q^\top K_X Q + \Lambda||Q^\top| \quad (21.24)$$

$$= |Q^\top K_X Q + \Lambda| \quad (21.25)$$

$$= |A + \Lambda| \quad (21.26)$$

So, $K_X = QAQ^\top = Q(Q^\top K_X Q)Q^\top$

But trace is preserved, since

$$\tr(A) = \tr(Q^\top K_X Q) = \tr(QQ^\top K_X) = \tr(K_X) \quad (21.27)$$

So the optimization problem we get is:

$$\max_{A \succeq 0} |A + \Lambda| \quad \text{subject to} \quad \tr(A) \leq nP \quad (21.28)$$

Recall, for any jointly Gaussian random vector $U_{1:n}$, we have

$$h(U_{1:n}) = \frac{1}{2} \log((2\pi e)^n |K_U|) \quad (21.29)$$

$$\leq \sum_i h(U_i) \quad (21.30)$$

$$= \frac{1}{2} \log((2\pi e)^n |\text{diag}(K_U)|) \quad (21.31)$$

where $|\text{diag}(K_U)| = \prod_{i=1}^n K_U(i,i)$
Colored Noise

Thus

\[ |A + \Lambda| \leq \prod_i (A_{ii} + \lambda_i) \quad (21.32) \]

again noting that \( \{\lambda_i\}_i \) are the eigenvalues of \( K_Z \).

This achieves equality when \( A = \text{diag}(A) \), reducing our optimization problem to:

\[
\text{maximize } \prod_i (A_{ii} + \lambda_i) \quad \text{subject to } \sum_i A_{ii} = nP \quad (21.33)
\]

Why can we say \( \sum_i A_{ii} = nP \) rather than \( \sum_i A_{ii} \leq nP \)?

Note, we use equality in the constraint since \( \lambda_i \geq 0 \) and also we have constraint \( A_{ii} \geq 0 \) in the maximum process.

Colored Noise - KKT

Like before, we have KKT conditions to get:

\[ A_{ii} = (\nu - \lambda_i)^+ \quad (21.34) \]

and

\[ \sum_i (\nu - \lambda_i)^+ = nP \quad (21.35) \]

And the total capacity becomes

\[ C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i}) \quad \text{bits per transmission} \quad (21.36) \]

So this is water filling again, but on the eigenvalues of \( K_Z \).

I.e., we are filling a diagonalized version of the problem. Once we get \( A_{ii} \), we can “re-correlate” using \( K_X = QAQ^T \) to get final constraint on \( X \).
So now we have positive semidefinite kernel functions $f_X$ and $f_Z$ such that (stationary case)

$$EX_iX_j = f_X(|i - j|) \text{ and } EZ_iZ_j = f_Z(|i - j|)$$ \hspace{1cm} (21.37)

so that the matrices are symmetric Toeplitz (i.e., symmetric same along each diagonal)

It can be shown that we have water filling in this case as well, but this time in the frequency domain.

Let $R_{ZZ}(e^{j\omega})$ be the power spectrum of the noise and signal discrete FT $X(e^{j\omega})$.

Then the capacity in this case becomes:

$$C = \int_{-\pi}^{\pi} \frac{1}{2} \log \left( 1 + \frac{(\nu - R_{ZZ}(e^{j\omega}))^+}{R_{ZZ}(e^{j\omega})} \right) d\omega$$ \hspace{1cm} (21.38)

where $\nu$ is chosen to make

$$\int_{-\pi}^{\pi} (\nu - R_{ZZ}(e^{j\omega}))^+ d\omega = P$$ \hspace{1cm} (21.39)
Parallel Channels

Colored Noise

Feedback

Matrix Inequalities by IT

**Spectral Water filling**

- Pictured:

![Diagram](image)

**Feedback**

- $C_n$ was the capacity without feedback, and $C_{n,FB}$ is the capacity with, and our goal here is to compare the two.
- For discrete memoryless channel, feedback does not increase capacity (since in finding $C$, you are already exploiting the randomness to the extent possible).
- For discrete non-memoryless channels, feedback can help (but not by too much as we will see).
- Here, the form of non-memorylessness is that the noise is correlated, i.e.,

$$Z_{1:n} \sim \mathcal{N}(0, K_Z^{(n)})$$ (21.40)

where again $K_Z^{(n)}$ is not a diagonal matrix.
- Also, we have $Y_{1:n} = X_{1:n} + Z_{1:n}$ but **not** $X_{1:n} \perp Z_{1:n}$ due to feedback (i.e., independence doesn’t hold in this case).
- Since noise is correlated, can use previous received values to help predict the noise.
### Feedback

**Definition 21.5.1**

We define a $(2^{nR}, n)$ code as a set of mappings of the form

$$x_i(w, Y_{1:i-1}) : \{1, 2, \ldots, 2^{nR}\} \times \mathbb{R}^{(i-1)} \rightarrow \mathbb{R}$$

where $w \in \{1, 2, \ldots, 2^{nR}\}$ is the input message and $Y_{1:i-1}$ is the sequence of past output values (feedback). We think of $x_i(w, \cdot)$ this as a “code function” rather than a mere codeword.

- We also have a power constraint of the form

$$E \left[ \frac{1}{n} \sum_{i=1}^{n} x_i^2(w, Y_{1:i-1}) \right] \leq P \text{ for all } w \in \{1, 2, \ldots, 2^{nR}\}$$

We already saw that

$$C_n = \frac{1}{2} \sum_{i=1}^{n} \log(1 + \frac{(\nu - \lambda_i)^+}{\lambda_i})$$

bits per transmission

is the capacity for colored noise channel w/o feedback.

- We want to see, in various different ways, how much better $C_{n, FB}$ (the capacity with colored noise and feedback) is going to be.
- Memory now since noise is correlated.
- Feedback useful if we have channel asymmetric communication costs or abilities. E.g.,
  - Cheap to communicate in one direction, expensive to communicate in the other direction (e.g., Mars rover, some sensor networks, internet service providers).
  - Feasible to consider case when feedback is partly noisy (but asymmetries exist in the noise in different directions).
Top: without feedback and without correlated noise
Bottom: without feedback and with correlated noise

Note, $Z_i = Y_i - X_i$ which means that

$$I(X_1:n; Y_1:n) = h(Y_1:n) - h(Y_1:n | X_1:n)$$

$$= h(X_1:n + Z_1:n) - h(Z_1:n)$$

(21.44) max when Gaussian

(21.45) Gaussian

So, max is achieved when $X_1:n$ is also Gaussian, and in fact can be jointly Gaussian (exercise: show this).
Knowing previous $X$ value does not help to code any better than knowing previous $Y$ values. I.e.,

$$X_i = f(w, Y_{1:i-1}) = f(w, X_{1:i-1}, Y_{1:i-1}) = f(Z_{1:i-1}) \quad (21.46)$$

which follows since everything is jointly Gaussian, and $Z = Y - X$, and since what is required to deduce $X_i$ from $X_{1:i-1}, Y_{1:i-1}$ when $Y = X + Z$ is this difference $Y - X$.

I.e., all $X_i$'s can be written w.l.o.g. as

$$X_i = \sum_{j=1}^{i-1} b_{ij} Z_j + V_i \text{ with } V_i \perp Z_j \forall i, j \quad (21.47)$$

Or $X_{1:n} = BZ_{1:n} + V_{1:n}$ with $B$ a strictly lower diagonal matrix with $B_{ii} = 0$ for all $i$.

Matrix factorization. If $S \sim N(0, K)$ then $S_i = \sum_{j<i} b_{ij} S_j + \epsilon_i$ with $K = UDU^\top = (B + I)D(B + I)^\top$, with $U$ having 0 on lower triangular, 1's on diagonal, and the $b$'s on the upper triangular.

Note that $B$ controls the feedback — if $B = 0$ then no feedback.

This is a general form of conditional Gaussian, where $V$ is also Gaussian and independent of $Z$.

So we can write this capacity as the following optimization:

$$C_{n,FB} = \max_{\frac{1}{n} \text{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_{X+Z}|}{|K_Z|} \quad (21.48)$$

where $K_{X+Z} = E[(X + Z)(X + Z)^\top]$

$$= \max_{K_V \succeq 0} \frac{1}{2n} \log \frac{|(B + I)K_Z(B + I)^\top + K_V|}{|K_Z|}$$

$$\text{pred}(B) = \text{true} \quad (21.49)$$

where $\text{pred}(B) = \text{tr}(BK_ZB^\top + K_V) \leq nP$. Justification on next slide.
feedback and correlated noise

This follows since when $X = BZ + V$ with $Z \perp V$ we have

$$K_X = EXX^T = E[(BZ + V)(BZ + V)^T] = BKZB^T + KV$$ (21.50)

and


$$= E[(B + I)ZZ^T(B + I)^T + (B + I)ZV^T + V[(B + I)Z]^T + VV^T]$$ (21.52)

$$= (B + I)KZ(B + I)^T + KV$$ (21.53)

Theorem 21.5.2

For a Gaussian channel with feedback, the rate $R_n$ for any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ satisfies $R_n \leq C_n, FB + \epsilon_n$ when $\epsilon_n \to 0$ as $n \to \infty$.

Proof.

- Assume $W$ is uniform over $2^{nR}$.
- We have Markov chain

$$W \to X_{1:n} \to Y_{1:n} \to \hat{W}$$ (21.57)

- Thus, Fano’s inequality still holds, i.e.

$$H(W|\hat{W}) \leq 1 + nR_nP_e^{(n)} = n\left(\frac{1}{n} + R_nP_e^{(n)}\right) = n\epsilon_n$$ (21.58)

where $\epsilon_n \to 0$ as $n \to \infty$ (due to $P_e^{(n)} \to 0$).
feedback and correlated noise

... proof continued.

\[ nR_n = H(W) = I(W; \hat{W}) + H(W|\hat{W}) \leq I(W; \hat{W}) + n\epsilon_n \] (21.59)
\[ \leq I(W; Y^n) + n\epsilon_n = \sum_i I(W; Y_i|Y_{1:i-1}) + n\epsilon_n \] (21.60)
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1})] + n\epsilon_n \] (21.61)
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1})] + n\epsilon_n \] (21.62)
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Y_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \] (21.63)
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|W, Y_{1:i-1}, X_{1:i-1}, Z_{1:i-1})] + n\epsilon_n \]
\[ = \sum_i [h(Y_i|Y_{1:i-1}) - h(Z_i|Z_{1:i-1})] + n\epsilon_n \]
\[ = h(Y_1:n) - h(Z_1:n) + n\epsilon_n \] (21.64)

...since \( h(Y_1:n) \) is maximized when \( Y_1:n \) is jointly Gaussian.

So

\[ R_n \leq \frac{1}{n} [h(Y_1:n) - h(Z_1:n)] + \epsilon_n \] (21.65)
\[ \leq \frac{1}{2n} \log \frac{|K_Y|}{|K_Z|} + \epsilon_n \] (21.66)
Matrix Inequalities

- Next, we are going to prove a number of very useful matrix inequality theorems that are important in their own right.
- These inequalities are also going to be useful to compare the capacity with and without feedback and we’ll find that feedback doesn’t help that much (we do this at the very end of this section, and the topic of feedback comes up every few slides).
- See also chapter 17 in the book (2nd edition).

**Theorem 21.6.1**

∀\(X, Z\) \(n\)-D random variables (not necessarily Gaussian), we have

\[
K_{X+Z} + K_{X-Z} = 2K_X + 2K_Z
\]  

(21.67)

**Proof.**

\[
K_{X+Z} = E(X + Z)(X + Z)^\top = EXX^\top + EXZ^\top + EZX^\top + EZZ^\top
\]

\[
= K_X + K_{XZ} + K_{ZX} + K_Z
\]  

(21.68)

\[
K_{X-Z} = E(X - Z)(X - Z)^\top = EXX^\top - EXZ^\top - EZX^\top + EZZ^\top
\]

\[
= K_X - K_{XZ} - K_{ZX} + K_Z
\]  

(21.69)

sum up both to get the result.
Matrix Inequalities

**Theorem 21.6.2**

If $A \succeq 0$, $B \succeq 0$, and $A - B \succeq 0$, then $|A| \geq |B|$

**Proof.**

- Let $C = A - B$, $X_1 \sim \mathcal{N}(0, B)$, $X_2 \sim \mathcal{N}(0, C)$, $X_1 \perp X_2$.
- If $Y = X_1 + X_2$ then $Y \sim \mathcal{N}(0, B + C) = \mathcal{N}(0, A)$.
- Then we get
  
  $$h(Y) \geq h(Y|X_2) = h(X_1|X_2) = h(X_1)$$  \hspace{1cm} (21.70)

  Then, using the entropy of Gaussian formula,

  $$\frac{1}{2} \log((2\pi e)^n |A|) \geq \frac{1}{2} \log((2\pi e)^n |B|)$$  \hspace{1cm} (21.71)

  and the result follows

**Theorem 21.6.3**

Let $X, Z$ be arbitrary $n$-D random variables. Then

$$|K_{X+Z}| \leq 2^n |K_X + K_Z|$$  \hspace{1cm} (21.72)

**Proof.**

$$2(K_X + K_Z) = K_{X+Z} + K_{X-Z}$$  \hspace{1cm} (21.73)

$$\Rightarrow 2(K_X + K_Z) - K_{X+Z} = K_{X-Z} \succeq 0$$  \hspace{1cm} (21.74)

$$\Rightarrow K_{X+Z} \leq 2(K_X + K_Z)$$  \hspace{1cm} (21.75)

$$\Rightarrow |K_{X+Z}| \leq |2(K_X + K_Z)| = 2^n |K_X + K_Z|$$  \hspace{1cm} (21.76)
Theorem 21.6.4

Let \( A \succeq 0, B \succeq 0, \) and \( 0 \leq \lambda \leq 1, \) then

\[
|\lambda A + (1 - \lambda)B| \geq |A|^\lambda |B|^{1-\lambda} \tag{21.77}
\]

or in other words, \( \log(|\lambda A + (1 - \lambda)B|) \geq \lambda \log |A| + (1 - \lambda) \log |B| \) or log-determinant is concave for non-negative definite matrices.

Proof.

- Let \( X \sim \mathcal{N}(0, A), Y \sim \mathcal{N}(0, B) \)
- Let \( Z = X \mathbf{1}(\theta = 1) + Y \mathbf{1}(\theta = 0) \) with \( X \perp \!\!\!\perp Y \) and where \( P(\theta = 1) = 1 - P(\theta = 0) = \lambda. \) Note \( Z \) is not necessarily Gaussian.

\[
K_Z = EZZ^\top = E[(X \mathbf{1}(\theta = 1) + Y \mathbf{1}(\theta = 0))^2] \tag{21.78}
\]

\[
= E[XX^\top \mathbf{1}(\theta = 1)\mathbf{1}(\theta = 1)] + E[XY^\top \mathbf{1}(\theta = 1)\mathbf{1}(\theta = 0)] \tag{21.79}
\]

\[
+ E[YX^\top \mathbf{1}(\theta = 0)\mathbf{1}(\theta = 1)] + E[YY^\top \mathbf{1}(\theta = 0)\mathbf{1}(\theta = 0)]
= \lambda A + (1 - \lambda)B \tag{21.80}
\]

Thus, we get

\[
\frac{1}{2} \log((2\pi e)^n |\lambda A + (1 - \lambda)B|) \geq h(Z) \geq h(Z|\theta) \tag{21.81}
\]

\[
= \lambda h(X) + (1 - \lambda)h(Y) = \frac{1}{2} \log((2\pi e)^n |A|^\lambda |B|^{1-\lambda}) \tag{21.82}
\]

Must we assume \( X \perp \!\!\!\perp Y \)?
Definition 21.6.5

We define a distribution that factorizes generatively if:

\[ f(x_1:n, z_1:n) = f(z_1:n) \prod_i f(x_i | x_{1:i-1}, z_{1:i-1}) \]  

(21.84)

- Note this is called “causally related” in the book, but it need not have anything to do with causality, which can be something entirely different (see the work of J. Pearl for discussions on and models of causality).
- We can view this pictorially:

![Diagram showing factorization](image)

Definition 21.6.6

We define a distribution that factorizes generatively if:

\[ f(x_1:n, z_1:n) = f(z_1:n) \prod_i f(x_i | x_{1:i-1}, z_{1:i-1}) \]  

(21.85)

called “causally related” in the book, but it need not have anything to do with causality, which can be something entirely different (see the work of J. Pearl for discussions on and models of causality).
- Pictorially:

![Diagram showing factorization](image)
Parallel Channels  Colored Noise  Feedback  Matrix Inequalities by IT

Factorization

- Since feedback codes at capacity are of the form $X_i = f(Z_{1:i-1}) = f(Z_{1:i-1}, X_{1:i-1})$ the above generative factorization model applies in this case as well.

**Theorem 21.6.7**

If $x_{1:n}$ and $z_{1:n}$ factorize generatively (not nec. Gaussian), then

$$h(x_{1:n} - z_{1:n}) \geq h(z_{1:n}) \quad (21.86)$$

and

$$|K_{X-Z}| \geq |K_Z| \quad (21.87)$$

**Proof.**

$$h(X_{1:n} - Z_{1:n}) = \sum_i h(X_i - Z_i | X_{1:i-1} - Z_{1:i-1}) \quad (chain rule) \quad (21.88)$$

$$\geq \sum_i h(X_i - Z_i | X_{1:i-1}, Z_{1:i-1}, X_i) \quad (conditioning) \quad (21.89)$$

$$= \sum_i h(Z_i | X_{1:i-1}, Z_{1:i-1}, X_i) \quad (h(X_i | X_i) = 0) \quad (21.90)$$

$$= \sum_i h(Z_i | Z_{1:i-1}) \quad (factorization) \quad (21.91)$$

$$= h(Z_{1:n}) \quad (chain rule) \quad (21.92)$$
factorization

... proof continued.

Next, let $\tilde{X}_{1:n}, \tilde{Z}_{1:n}$ be independent Gaussian r.v.’s with covariance $K_X, K_Z$ respectively and let factorization assumption hold. Then

$$\frac{1}{n} \log[(2\pi e)^n |K_X - Z|] = h(\tilde{X} - \tilde{Z})$$

(21.93)

$$\geq h(\tilde{Z}) = \frac{1}{2} \log((2\pi e)^n |K_Z|)$$

(21.94)

and the result follows due to monotonicity of the log

Now we are ready to start comparing $C_n$ (capacity without feedback) and $C_{n, FB}$ (capacity with feedback).

We’ll do this both with an additive bound and a multiplicative bound.

Additive Bound

**Theorem 21.6.8**

$$C_{n, FB} \leq C_n + 1/2$$

(21.95)

So at most 1/2 bit per channel use of gain!

**Proof.**

$$C_{n, FB} \leq \max_{\frac{1}{n} \operatorname{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_Y|}{|K_Z|}$$

(21.96)

$$\leq \max_{\frac{1}{n} \operatorname{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{2^n |K_X + Z|}{|K_Z|}$$

(21.97)

$$\leq \max_{\frac{1}{n} \operatorname{tr}(K_X) \leq P} \frac{1}{2n} \log \frac{|K_X + Z|}{|K_Z|} + \frac{1}{2}$$

(21.98)

$$\leq C_n + \frac{1}{n}$$

(21.99)
Theorem 21.6.9

\[ C_{n,FB} \leq 2C_n \]  \hspace{1cm} (21.100)

or equivalently

\[ \frac{1}{2n} \log \frac{|KX+Z|}{|KZ|} \leq 2 \frac{1}{2n} \log \frac{|KX+KZ|}{|KZ|} \]  \hspace{1cm} (21.101)

Proof.

\[ \frac{1}{2n} \log \frac{|KX+Z|}{|KZ|} = \frac{1}{2n} \log \frac{|KX+Z|^{1/2}}{|KZ|^{1/2}} \]  \hspace{1cm} (21.102)

\[ = \frac{1}{2n} \log \frac{|KX+Z|^{1/2}|KZ|^{1/2}}{|KZ|} \]  \hspace{1cm} (21.103)

\[ \leq \frac{1}{2n} \log \frac{|KX-Z|^{1/2}|KX-Z|^{1/2}}{|KZ|} \hspace{1cm} \text{by Thm 21.6.7} \]  \hspace{1cm} (21.104)

\[ \leq \frac{1}{2n} \log \frac{\frac{1}{2}KX+Z + \frac{1}{2}KX-Z}{|KZ|} \hspace{1cm} \text{by Thm 21.6.4} \]  \hspace{1cm} (21.105)

\[ = \frac{1}{2n} \log \frac{|KX+KZ|}{|KZ|} \hspace{1cm} \text{by Thm 21.6.1} \]  \hspace{1cm} (21.106)
Combined Bound for Feedback

Corollary 21.6.10

\[ C_{n,FB} \leq \min \{2C_n, C_n + 1/2\} \quad (21.108) \]

So unfortunately, feedback in this model is not as useful as we might think it would be.