ELE539: Optimization of Communication Systems
Lecture 20: Optimal Spectrum Management for DSL

Professor M. Chiang
Electrical Engineering Department, Princeton University

April 7, 2004
Lecture Outline

- Spectrum management in DSL systems
- Convexity of rate region
- Optimal spectrum management algorithm
- Proof of convergence
- Performance

Acknowledgement: Paper by Raphael Cendrillon, Wei Yu, Marc Moonen, Jan Verlinden, and Tom Bostoen, To appear in IEEE ICC
Spectrum Management in DSL

Cross talk is the performance bottleneck

- Crosstalk cancellation
- Static spectrum management
- Dynamic spectrum management

Power control in DSM:

- Iterative waterfilling
- Centralized operation
**Key Problem and Idea**

Problem: power constraints coupled across $K$ tones

$K = 256$ for ADSL and $K = 4096$ for VDSL

Exponential complexity for a non-convex problem

Solution: dual decomposition to decouple the coupling constraint

Special structures makes dual decomposition work for non-convex problem

Linear complexity in $K$
System Model

\[ y_k = H_k x_k + z_k, \quad k = 1, 2, \ldots, K \]

\( x^n_k \): signal transmitted onto line \( n \) at tone \( k \)

\( H_k \): \( N \times N \) channel transfer matrix for tone \( k \)

\[ \sigma^2_k \triangleq \mathbb{E} \{|z^n_k|^2\} : \text{Noise PSD} \]

\[ s^n_k \triangleq \mathbb{E} \{|x^n_k|^2\} : \text{Transmit PSD} \]

\[ s_n \triangleq [s^n_1, \ldots, s^n_K] \]

\( \Gamma \): SNR-gap to capacity (function of BER, coding gain, margin)

Bit loading on line \( n \) at tone \( k \):

\[ b^n_k \triangleq \left\lfloor \log_2 \left( 1 + \frac{1}{\Gamma} \frac{|h^{n,n}_k|^2 s^n_k}{\sum_{m \neq n} |h^{n,m}_k|^2 s^m_k + \sigma^n_k} \right) \right\rfloor \]

Data rate on line \( n \):

\[ R_n = \sum_k b^n_k \]
Bitloading to Powerloading

\[
\begin{bmatrix}
  s^1_k \\
  s^2_k \\
\end{bmatrix} = (I_N - \Lambda_k A_k)^{-1} \Lambda_k \sigma_k
\]

where

\[
A \triangleq \begin{bmatrix}
  0 & \alpha_{k}^{1,2} \\
  \alpha_{k}^{2,1} & 0 \\
\end{bmatrix}
\]

\[
\alpha_{k}^{n,m} \triangleq \Gamma |h_{k}^{n,m}|^2 |h_{k}^{n,n}|^{-2}
\]

\[
\sigma_k \triangleq \Gamma[\sigma_k^1, \sigma_k^2]^T
\]

\[
\Lambda_k \triangleq \text{diag}\{2^{b_k^1} - 1, 2^{b_k^2} - 1\}
\]
Spectrum Management

Problem formulation for 2 lines:

maximize \( R_1 \)

subject to

\[ R_1 \geq R_1^{target} \]
\[ \sum_k s_{nk}^n \leq P_n, \quad n = 1, 2 \]
\[ s_{nk}^n \leq s_{nk,\text{max}}, \quad \forall k, \quad n = 1, 2 \]

Pareto optimization formulation:

maximize \( wR_1 + (1 - w)R_2 \)

subject to

\[ \sum_k s_{nk}^n \leq P_n, \quad n = 1, 2 \]
\[ s_{nk}^n \leq s_{nk,\text{max}}, \quad \forall k, \quad n = 1, 2 \]
Convexity of Rate Region

ADSL and VDSL: 4.3125 kHz tone spacing

For fine-enough tone spacing (as $K \to \infty$), achievable rate region is convex

Therefore, by varying $w \in [0, 1]$, all Pareto optimal point will be obtained on the optimal tradeoff curve between $R_1$ and $R_2$

Problems:

- Non-convex objective function
- Total power constraint coupled
- Exhaustive search takes exponential amount of time in $K$
Dual Decomposition

Lagrangian:

\[ L \triangleq wR_1 + (1 - w)R_2 + \lambda_1(P_1 - \sum_k s^1_k) + \lambda_2(P_2 - \sum_k s^2_k) \]

Lagrangian on tone \( k \):

\[ L_k \triangleq wb^1_k + (1 - w)b^2_k - \lambda_1 s^1_k(b^1_k, b^2_k) - \lambda_2 s^2_k(b^1_k, b^2_k) \]

Decomposition of Lagrangian:

\[ L = \sum_k L_k + \lambda_1 P_1 + \lambda_2 P_2 \]

First, maximize \( L(w, \lambda_1, \lambda_2, s^1_k, s^2_k) \) over \( s_1, s_2 \) in a decoupled way

Then find the optimal \( \lambda_1, \lambda_2 \)
Main Function

for $w = 0...1$

\[ s_1, s_2 = \text{optimize}_1 \lambda_1(w) \]

end

Function $s_1, s_2 = \text{optimize}_1 \lambda_1(w)$

\[ \lambda_{1\text{max}} = 1, \; \lambda_{1\text{min}} = 0 \]

while $\sum_k s_k^1 > P_1$

\[ \lambda_{1\text{max}} = 2 \lambda_{1\text{max}} \]

\[ s_1, s_2 = \text{optimize}_2 \lambda_2(w, \lambda_{1\text{max}}) \]

end

repeat

\[ \lambda_1 = (\lambda_{1\text{max}} + \lambda_{1\text{min}})/2 \]

\[ s_1, s_2 = \text{optimize}_2 \lambda_2(w, \lambda_1) \]
if $\sum_k s^1_k > P_1$, then $\lambda_1^{\text{min}} = \lambda_1$, else $\lambda_1^{\text{max}} = \lambda_1$

until convergence

Function $s_1, s_2 = \text{optimize}_{\lambda_2}(w, \lambda_1)$

$\lambda_2^{\text{max}} = 1$, $\lambda_2^{\text{min}} = 0$

while $\sum_k s^2_k > P_2$

$\lambda_2^{\text{max}} = 2\lambda_2^{\text{max}}$

$s_1, s_2 = \text{optimize}_s(w, \lambda_1, \lambda_2^{\text{max}})$

end

repeat

$\lambda_2 = (\lambda_2^{\text{max}} + \lambda_2^{\text{min}})/2$

$s_1, s_2 = \text{optimize}_s(w, \lambda_1, \lambda_2)$

if $\sum_k s^2_k > P_2$, then $\lambda_2^{\text{min}} = \lambda_2$, else $\lambda_2^{\text{max}} = \lambda_2$

until convergence
Function $s_1, s_2 = \text{optimize}_s(w, \lambda_1, \lambda_2)$

for $k = 1 \ldots K$

$$b_{k}^1, b_{k}^2 = \arg \max_{b_{k}^1, b_{k}^2} L_k(b_{k}^1, b_{k}^2, w, \lambda_1, \lambda_2)$$

$$s_{k}^1 = s_{k}^1(b_{k}^1, b_{k}^2), s_{k}^2 = s_{k}^2(b_{k}^1, b_{k}^2)$$

end

• Computational complexity: linear in $K$
• Each inner-cycle: non-convex optimization over 2 variables is easy
• Theorem: converges to optimal spectrum management
Near-Far Problem in Upstream VDSL

![Diagram showing the Near-Far Problem in Upstream VDSL](image)

**Graph**

- **Axes**:
  - X-axis: 600m line (Mbps)
  - Y-axis: 1200m line (Mbps)

- **Legend**:
  - Flat PBO
  - Ref. PSD
  - Lt. W.J.
  - OSM

- **Data Points**:
  - Various data points indicating performance metrics for different scenarios.

**Diagram**

- **Network Setup**:
  - CO/RT
  - 1200m line
  - 600m line

- **Key Metrics**:
  - Performance degradation due to distance and configuration factors.
Near-Far Problem in Upstream VDSL
RT Downstream ADSL

![Diagram showing CO, RT, and ADSL rates across different distances.]

- CO to RT: 5 km
- RT to CO: 4 km
- CO to RT: 3 km

Graph showing RT ADSL Rate (Mbps) vs. CO ADSL Rate (Mbps) with different markers for Flat PBO, It. W.f., and OSM.
RT Downstream ADSL
Proof

We first define the cost function

\[ L(s_n, \lambda_n) \triangleq f(s_n) + \lambda_n (P_n - \sum_k s^n_k) \]  \hspace{1cm} (1)

Denote the optimal power allocation for a given \( \lambda_n \)

\[ s_n(\lambda_n) \triangleq \arg \max_{s_n} L(s_n, \lambda_n) \]

with \( s^n_k(\lambda_n) \triangleq [s_n(\lambda_n)]_k \). The routine for user \( n \) is then
Routine for user n

\[ \lambda_n^{\text{max}} = 1, \lambda_n^{\text{min}} = 0 \]

while \( \sum_k s_k^n > P_n \)

\[ \lambda_n^{\text{max}} = 2 \lambda_n^{\text{max}} \]

\[ s_n = \arg \max_{s_n} f(s_n) + \lambda_n^{\text{max}}(P_n - \sum_k s_k^n) \]

end

repeat

\[ \lambda_n = (\lambda_n^{\text{max}} + \lambda_n^{\text{min}})/2 \]

\[ s_n = \arg \max_{s_n} f(s_n) + \lambda_n(P_n - \sum_k s_k^n) \]

if \( \sum_k s_k^n > P_n \), then \( \lambda_n^{\text{min}} = \lambda_n \), else \( \lambda_n^{\text{max}} = \lambda_n \)

until convergence

To prove the convergence of this routine we make use of the following Lemma.

**Lemma 1** \( \sum_k s_k^n(\lambda_n) \) is monotonic decreasing in \( \lambda_n \).

**Proof.** Consider two Lagrangian multipliers \( \lambda_n^a \) and \( \lambda_n^b \) and their corresponding optimal PSDs \( s_n^a \triangleq s_n(\lambda_n^a) \) and \( s_n^b \triangleq s_n(\lambda_n^b) \). Denote the elements of these PSDs as \( s_k^n,a \) and
\( s_k^{n,b} \) respectively. Let

\[
\lambda_n^b \geq \lambda_n^a
\]

(2)

Define

\[
A \triangleq f(s_n^b) + \lambda_n^b (P_n - \sum_k s_k^{n,b})
\]

\[
B \triangleq f(s_n^a) + \lambda_n^b (P_n - \sum_k s_k^{n,a})
\]

\[
C \triangleq f(s_n^a) + \lambda_n^a (P_n - \sum_k s_k^{n,a})
\]

\[
D \triangleq f(s_n^b) + \lambda_n^a (P_n - \sum_k s_k^{n,b})
\]

Now \( L(s_n^b, \lambda_n^b) \geq L(s_n^a, \lambda_n^b) \) by the optimality of \( s_n^b \) in \( L(s_n, \lambda_n^b) \).

Hence \( A \geq B \). Similarly the optimality of \( s_n^a \) in \( L(s_n, \lambda_n^a) \)

implies \( C \geq D \). Consider 3 cases:

In the first case let \( P_n - \sum_k s_k^{n,a} \geq 0 \). Combining this with (2)

implies \( B \geq C \). Now \( A \geq B \geq C \geq D \) implies \( A - D \geq B - C \).

Hence
\[(\lambda_n^b - \lambda_n^a)(P_n - \sum_k s_{k, n}^{n,b}) \geq (\lambda_n^b - \lambda_n^a)(P_n - \sum_k s_{k, n}^{n,a})\]

which implies

\[\sum_k s_{k, n}^{n,a} \geq \sum_k s_{k, n}^{n,b}\] (3)

In the second case let \(P_n - \sum_k s_{k, n}^{n,b} \leq 0\). Combining this with (2) implies \(D \geq A\). Now \(C \geq D \geq A \geq B\) implies \(C - B \geq D - A\). Hence

\[(\lambda_n^a - \lambda_n^b)(P_n - \sum_k s_{k, n}^{n,a}) \geq (\lambda_n^a - \lambda_n^b)(P_n - \sum_k s_{k, n}^{n,b})\]

which again implies (3).

In the third case let \(P_n - \sum_k s_{k, n}^{n,a} < 0\) and \(P_n - \sum_k s_{k, n}^{n,b} > 0\). This implies \(P_n - \sum_k s_{k, n}^{n,b} > P_n - \sum_k s_{k, n}^{n,a}\) and again leads to (3).

So in all cases a larger \(\lambda_n\) leads to a smaller \(\sum_k s_{k}^{n}\). This implies that \(\sum_k s_{k}^{n}\) is monotonic decreasing in \(\lambda_n\).
Lemma 2  Routine for user $n$ converges. At convergence

$$s_n = \arg \max_{s_n} f(s_n) \text{ s.t. } \sum_k s_k^n \leq P_n$$

Proof. The routine consists of two stages: a preamble that determines $\lambda_{n}^{\max}$ and the actual routine itself.

The preamble clearly converges since $\sum_k s_k^n(\lambda_n) \to 0$ as $\lambda_n \to \infty$.

The convergence of the main part of routine for user $n$ can be shown as follows: $\lambda_{n}^{\max} - \lambda_{n}^{\min}$ decreases by half in each iteration. Thus, $\lambda_n$ converges to a fixed value. Let’s now consider two cases, depending on whether $\sum_k s_k^n(\lambda_{n}^{\min}) > P_n$ or not.

Suppose that $\sum_k s_k^n(\lambda_{n}^{\min}) > P_n$ at $\lambda_{n}^{\min} = 0$, then since the preamble ensures that $\sum_k s_k^n(\lambda_{n}^{\max}) \leq P_n$, throughout the algorithm it is always the case that $\sum_k s_k^n(\lambda_{n}^{\min}) > P_n$ and $\sum_k s_k^n(\lambda_{n}^{\max}) \leq P_n$. Since $\lambda_{n}^{\max} \geq \lambda_n \geq \lambda_{n}^{\min}$, $\lambda_{n}^{\min}$ and $\lambda_{n}^{\max}$ converge to a fixed value, and since $\sum_k s_k^n(\lambda_n)$ is monotonic in $\lambda_n$, this implies that $\sum_k s_k^n(\lambda_n)$ must converge to $P_n$. On
the other hand, suppose that \( \sum_k s_k^n(\lambda_n^{\text{min}}) \leq P_n \) at \( \lambda_n^{\text{min}} = 0 \). Then, \( \lambda_n \) will converge to 0.

Hence the algorithm will converge and at convergence either \( \lambda_n = 0 \) or \( \sum_k s_k^n(\lambda_n) = P_n \). So at convergence

\[
L(s_n, \lambda_n) = f(s_n)
\]

In the routine

\[
s_n = \arg \max_{s_n} L(s_n, \lambda_n)
\]

\[
= \arg \max_{s_n} f(s_n) \text{ s.t. } \sum_k s_k^n \leq P_n
\]

To see this clearly \( s_n \) satisfies the constraint. Further, if there is some other feasible \( s'_n \) that does better than \( s_n \) for the objective function \( f(s_n) \) then \( s'_n \) should do better than \( s_n \) for the objective \( L(s_n, \lambda_n) \) also. This is contradicted by the optimality of \( s_n \) in \( L(s_n, \lambda_n) \). Hence \( s_n \) must be optimal in \( f(s_n) \).

**Lemma 3** The function optimize\(_s\) yields PSDs \( s_1 \) and \( s_2 \) which maximize the Lagrangian.
Proof. From function optimize_s

\[ s_k^1, s_k^2 = \arg \max_{s_k^1, s_k^2} L_k(w, \lambda_1, \lambda_2, s_k^1, s_k^2) \]

Since \( L = \sum_k L_k + \lambda_1 P_1 + \lambda_2 P_2 \), and since optimisation of the Lagrangian is unconstrained (recall that the constraints are incorporated into the cost function and need not be explicitly enforced) this implies

\[ s_1, s_2 = \arg \max_{s_1, s_2} L(w, \lambda_1, \lambda_2, s_1, s_2) \]

Define the rates of user 1 and user 2 with the PSDs \( s_1 \) and \( s_2 \) as \( R_1(s_1, s_2) \) and \( R_2(s_1, s_2) \).

Corollary 1 The function optimize_\( \lambda_2 \) converges. At convergence

\[ s_2 = \arg \max_{s_2} \max_{s_1} wR_1(s_1, s_2) + (1 - w)R_2(s_1, s_2) + \lambda_1(P_1 - \sum s_k^1) \]

\[ + \sum_k s_k^2 \leq P_2 \]
Proof. Let $n = 2$ and
\[ f(s_2) \triangleq \max_{s_1} wR_1(s_1, s_2) + (1 - w)R_2(s_1, s_2) + \lambda_1(P_1 - \sum_k s^1_k). \]
Lemma 3 implies that optimize $\lambda_1$ and the routine are equivalent. Hence Lemma 2 implies optimize $\lambda_2$ converges, and that at convergence (4) is satisfied.

Corollary 2 The function optimize $\lambda_1$ converges. At convergence

\[ s_1 \quad = \quad \arg \max_{s_1} \max_{s_2} wR_1(s_1, s_2) + (1 - w)R_2(s_1, s_2) \]

s.t. \[ \sum_k s^1_k \leq P_1, \quad \sum_k s^2_k \leq P_2 \] (5)

Proof. Let $n = 1$ and
\[ f(s_1) = \max_{s_2} wR_1(s_1, s_2) + (1 - w)R_2(s_1, s_2) \text{ s.t. } \sum_k s^2_k \leq P_2. \]
Then Lemma 2 and Corollary 1 imply optimize $\lambda_1$ converges, and that at convergence (5) is satisfied.
Lecture Summary

- Centralized optimal spectrum management for DSL
- Find optimal rate pairs and improve from iterative waterfilling
- Dual decomposition still works for non-convex problem
- Scalable in number of tones, may not be scalable in number of lines

Toward Last Module of the Course

In the last three lectures, we have seen
• Applications of convex optimization to signal processing and resource management algorithms in physical layer for wireless MIMO and DSL systems
• How to deal with non-convexity

In the next three lectures, we will see
• How to optimize multiple objectives
• How to optimize over time
• How to optimize over discrete optimization variables
And applications: detection, decoding, routing, topology