ELEN E6717: Information Theory

Time: Tu 4:10-6:40pm.

Room: 1024 Mudd.

Text: T.M. Cover and J.A. Thomas. *Elements of Information Theory*, Wiley, 1991.

Prerequisite: ELEN 4815 (basic knowledge on probability and random process).

Instructor: X. Wang (Email: wangx@ee.columbia.edu, Phone: 854-6592).

Office Hours: Th 4:00pm – 5:00pm, 717 CEPSR.

Grading: Homework (bi-weekly) 20% + Midterm 40% + Final 40%.

Course Outline:

1. Preliminaries.
2. Lossless Source Coding.
3. Lossy Source Coding.
4. Discrete Memoryless Channels.
5. Gaussian Channels.
6. Information Theory and Portfolio Theory.

(Midterm covers chapters 1-3, and Final covers chapters 4-6.)

ELEN E6717: Information Theory

Chapter 1: Preliminaries

- Entropy of discrete sources
  - Definition of $H(X)$
  - Properties of entropy: nonnegativity, upper bound, concavity
- Joint entropy and conditional entropy
  - Definitions of $H(X,Y)$ and $H(X|Y)$
  - Chain rule of entropy
  - Conditioning reduces entropy
  - Independence bound on entropy
  - Entropy of a function of r.v.'s
- Relative entropy
  - Definitions of $D(p(x,y) \parallel q(x,y))$ and $D(p(x,y|z) \parallel q(x,y|z))$
  - Information inequality (nonnegativity of relative entropy)
  - Convexity of relative entropy
  - Chain rule of relative entropy
- Mutual information
  - Definitions of $I(X;Y)$ and $I(X;Y|Z)$
  - Mutual information and entropy
  - Chain rule of information
  - Nonnegativity of information
  - Concavity/Convexity of information
  - Data processing inequality
- Entropy rate
  - Two definitions of entropy rate: $H(\mathcal{X})$ and $H'(\mathcal{X})$
  - Equivalence of $H(\mathcal{X})$ and $H'(\mathcal{X})$ for stationary process

- Entropy rate of a stationary Markov chain
- Upper and lower bounds on entropy rate of functions of a stationary Markov chain

- Differential entropy of continuous sources
  - Definition of $h(X)$
  - Properties differential entropy: translation, scaling
  - Joint and conditional differential entropy: chain rule, conditioning, independence bound, entropy of a function of r.v.'s
  - Differential entropy of a linear transformation of a random vector
  - Differential entropy of multivariate Gaussian distribution
  - Relative differential entropy and mutual information: nonnegativity, mutual information and differential entropy, chain rules, data processing inequality
  - Gaussian distribution maximizes entropy
- Differential entropy rate
  - Definitions of $h(\mathcal{X})$
  - Differential entropy rate of a stationary process
  - Differential entropy rate of a stationary Gaussian process
  - Maximum entropy theorem

ELEN E6717: Information Theory

Chapter 2: Lossless Source Coding

- Shannon source coding theorems
  - Asymptotic equipartition property (AEP) for i.i.d. and stationary ergodic sources
  - Properties of typical sets
  - Fixed-length-to-fixed-length source coding theorem and its converse
  - Fixed-length-to-variable-length source coding theorem
- Properties of optimal (fixed-length-to-variable-length) codes
  - Non-singular codes, uniquely decodable codes, instantaneous codes
  - Kraft's inequality for instantaneous codes
  - Lower and upper bounds on optimal codeword length, large block encoding to approach entropy
  - Kraft's inequality for uniquely decodable codes
  - Shannon code, length of Shannon code using wrong distribution, competitive optimality of Shannon code
- Huffman coding
  - Huffman encoding algorithm
  - Optimality of Huffman code
- Arithmetic coding
  - Shannon-Fano-Elias code
  - Arithmetic codes
- Universal source coding
  - Definitions of type and type class
  - Properties of type for i.i.d. sequences
  - Existence of universal source code for i.i.d. sources
  - Lempel-Ziv encoding and decoding algorithms
  - Asymptotical optimality of Lempel-Ziv code

ELEN 6717: Information Theory

Chapter 3: Lossy Source Coding (Rate Distortion Theory)

- Rate distortion function
  - Hamming distortion and squared error distortion, rate distortion code, rate distortion function
  - Rate distortion function for binary source
  - Rate distortion function for Gaussian source
  - KKT conditions for constrained optimization
  - Rate distortion function for independent Gaussian sources - reverse water-filling
- Achievability of rate distortion theorem
  - Distortion typical sequence pair
  - Rate distortion theorem - achievability
- Converse of rate distortion theorem
  - Convexity of rate distortion function
  - Rate distortion theorem - converse
  - Properties of optimal rate distortion codes
- Computation of rate distortion function
  - Blahut-Arimoto algorithm - alternating minimization

ELEN 6717: Information Theory

Chapter 4: Discrete Memoryless Channels

- Channel capacity
  - Discrete memoryless channel, channel code, probability of error, achievability, channel capacity
  - Capacity of binary symmetric channel
  - Capacity of binary erasure channel
  - Capacity of (weakly) symmetric channel
- Property of channel capacity
  - $0 \leq C \leq \max\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$
  - Continuous and concave in $p(x)$.
  - Blahut-Arimoto algorithm for computing C: alternating maximization
- Achievability of channel coding theorem
  - Joint AEP
  - Proof of achievability
- Converse to channel coding theorem
  - Multiple use of DMC does not increase capacity per transmission
  - Zero-error code and properties of capacity achieving codes
  - Fano's inequality
  - Proof of converse
- Feedback capacity
  - Feedback code and feedback channel capacity
  - Capacity of DMC with feedback = Capacity of DMC without feedback

- Joint source-channel coding theorem
- Discrete memoryless multiple access channels
  - Definitions of multiple-access channel, code, and capacity region
  - Capacity of multiple-access channel
  - Convexity of capacity region
  - Interpretation of capacity region
- Distributed encoding of correlated sources
  - Distributed source coding problem
  - Slepian-Wolf theorem and interpretations
  - Duality between Slepian-Wolf coding and multiple access channel

ELEN 6717: Information Theory

Chapter 5: Gaussian Channels

- Discrete-time memoryless channel with input constraint
  - Definitions, achievability, channel capacity
  - Capacity of Gaussian channel
  - Gaussian channel is the worst additive noise channel
- Achievability of Gaussian channel capacity
  - Joint AEP
  - Achievability of capacity
  - Sphere packing interpretation
- Converse to Gaussian channel coding theorem
- Parallel Gaussian channels
  - Independent noise - water-filling interpretation
  - Colored noise - water-filling on eigenvalues
- Gaussian channels with feedback
  - Feedback code
  - Capacity of white Gaussian feedback channel - same as without feedback
  - Capacity of colored Gaussian feedback channel - increase by at most $\frac{1}{2}$ bit compared with no feedback
- Band-limited Gaussian channels
  - White Gaussian noise channel with idea filter
  - Colored Gaussian noise with arbitrary filtered input
- Gaussian multiple-access and broadcast channels
  - Gaussian multiple-access channel capacity and interpretation
  - Gaussian broadcast channel capacity and interpretation

ELEN 6717: Information Theory

Chapter 6: Information Theory and Portfolio Theory

- Definitions on stock market
  - Portfolio and growth rate
  - Properties of growth rate
- Properties of log-optimal portfolio
  - KTT characterization of log-optimal portfolio
  - Asymptotic optimality of log-optimal portfolio
  - Competitive optimality of log-optimal portfolio
  - Side information and growth rate
- Stationary markets
  - Chain rule
  - Growth rate and AEP
- Examples: Horse race
  - Doubling rate
  - Gambling and side information
  - Dependent horse races