OUTLINE

♦ Motivation: fundamental performance limits of analog-to-digital compression

♦ Combined sampling and source coding

♦ Optimal sampling structure

♦ Optimal sampling frequency under bitrate constraint
  - Below the Nyquist rate
MOTIVATION:

- Analog to digital (A/D) conversion:

Main goal: measuring and minimizing distortion in A/D

\[
\begin{align*}
D & \quad \text{error} \\
R & \quad \text{bitrate (memory) [bit/sec]}
\end{align*}
\]
COMBINED SAMPLING AND SOURCE CODING

Analog source coding:

\[ X(0 : T) \rightarrow \text{Encoder} \rightarrow W \in \{0, 1\}^{RT} \rightarrow \text{Decoder} \rightarrow \hat{X}(0 : T) \]

\[ X(0 : T) \rightarrow f_s \rightarrow \text{Quantizer} \rightarrow \{0, 1\}^{RT} \]

Distortion is due to:
- Time discretization / sampling (next slide)
- Finite bit representation (lossy compression)

System model: combined sampling and source coding

\[ X(\cdot) \rightarrow f_s \rightarrow Y[\cdot] \rightarrow R \rightarrow \text{Decoder} \rightarrow \hat{X}(\cdot) \]
WHY SUB-SAMPLING?

- High sampling rate bloats memory
- Sampling rate is limited by technology

![Graph showing SNR in bits vs. F_sample (Hz) for data from 1978 to 1999 and 2000 to 2005. Analog Devices: 24 bit 2.5MS/s, 16 bit 100 MS/s.](image)

**Analog vs Digital**

- What's that all over the floor?
- Oh, the data is coming in faster than our computers can absorb.

*Thomas A. Finney*
WHY SUB-SAMPLING?

Sampling theorem: \( f_s > f_{Nyq} \triangleq 2f_B \)

\[
X(t) = \sum_{n \in \mathbb{Z}} X \left( \frac{n}{f_s} \right) \frac{\sin (f_s t - n)}{f_s t - n}
\]

Shannon [1948]:
“we are not interested in exact transmission when we have a continuous source, but only in transmission to within a given tolerance”

- \( D(R) \) is the minimal distortion possible using bitrate \( R \)
- Can we achieve \( D(R) \) by sampling below \( f_{Nyq} \)?
INFORMATION THEORETIC REPRESENTATION

Remote (indirect) source coding

[Dobrushin, Wolf & Ziv ’70, Berger ’71, Witsenhausen ’80, CEO…]

\[ D(f_s, R) = \inf_{T} \inf_{Y \rightarrow \hat{X}} \frac{1}{2T} \int_{-T}^{T} \mathbb{E}d(x(t), \hat{x}(t)) \, dt \]
Remote Source Coding

\[ X^n_0 \xrightarrow{P_{Y|X}} Y^{m(n)}_0 \xrightarrow{\text{Enc}} \{1, \ldots, 2^{nR}\} \xrightarrow{\text{Dec}} d(X_n, \hat{X}_n) \]

- Reduced distortion measure [Witsenhausen '80]

\[ \hat{d}(y^m_0, \hat{y}^m_0) \triangleq \mathbb{E} \left[ d(X^n_0, \hat{X}^n_0|Y^m_0 = y^m_0, \hat{Y}^m_0 = y^m_0) \right] \]

\[ \mathbb{E}d_n \left( X^n_0, \hat{X}^n_0 \right) = \mathbb{E}\hat{d}_m \left( Y^m_0, \hat{Y}^m_0 \right) \]

\[ Y^{m(n)}_0 \xrightarrow{\text{Enc}} \{1, \ldots, 2^{nR}\} \xrightarrow{\text{Dec}} \hat{Y}^m_0 \]

- Corollary: source coding theorem:

\[ D_{X|Y}(R) = \inf_{n} \inf_{Y \rightarrow \hat{Y}} \mathbb{E}d_n \left( X^n_0, \hat{X}^n_0(\hat{Y}^m_0) \right) \]

\[ I(Y, \hat{Y}) \leq R \]
ASSUMPTIONS

diamond quadratic distortion \( d(x(t), \hat{x}(t)) = (x(t) - \hat{x}(t))^2 \)

[Wolf & Ziv ‘70]

\[ D(f_s, R) = \text{mmse}_{X|Y}(f_s) + \inf_{Y \rightarrow \hat{X}} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{T} \left( \mathbb{E}[X(t)|Y] - \hat{X}(t) \right)^2 dt \]

\( I(Y; \hat{X}) \leq R \)

\[ D(f_s, R) = \text{mmse}_{X|Y}(f_s) + \inf_{Y \rightarrow \hat{X}} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{T} \left( \mathbb{E}[X(t)|Y] - \hat{X}(t) \right)^2 dt \]

diamond \( X(\cdot) \) is stationary Gaussian with PSD \( S_X(f) \)

bullet \( S_X(f) \) is unimodal

diamond pointwise uniform sampler \( Y[n] = X(n/f_s) \)
SPECIAL CASE: QUADRATIC GAUSSIAN DISTORTION-RATE $f_s > f_{Nyq}$

\[ D(f_s, R) = D(R) \]

**Encoder**

**Decoder**

\[ R(\theta) = \frac{1}{2} \int_{-\infty}^{\infty} \log^+ \left[ \frac{S_X(f)}{\theta} \right] df \]

\[ D(\theta) = \int_{-\infty}^{\infty} \min \{ S_X(f), \theta \} df \]
RESULT I: SAMPLING-RATE-DISTORTION FUNCTION

Theorem [K., Goldsmith, Weissman, Eldar 2014]

\[
R(f_s, \theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left[ \tilde{S}_{X|Y}(f)/\theta \right] df
\]

\[
D(f_s, \theta) = \text{mmse}_{X|Y}(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\{\tilde{S}_{X|Y}(f), \theta\} df
\]

\[
\text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \tilde{S}_{X|Y}(f) df
\]

\[
\tilde{S}_{X|Y}(f) = \frac{\sum_{k \in \mathbb{Z}} S_X^2(f - f_s k)}{\sum_{k \in \mathbb{Z}} S_X(f - f_s k)}
\]
WATERFILLING INTERPRETATION

\[ R(f_s, \theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left[ \frac{\tilde{S}_{X|Y}(f)}{\theta} \right] df \]

\[ D(f_s, \theta) = \text{mmse}_{X|Y}(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\{\tilde{S}_{X|Y}(f), \theta\} df \]

\[ D(f_s, R) = \text{mmse} + \text{waterfilling} \]

Distortion due to sampling

\[ \sum_{k \in \mathbb{Z}} S_X(f - f_s k) \]

\[ \tilde{S}_{X|Y}(f) \]

\[ \sigma_X^2 = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \sum_{k \in \mathbb{Z}} S_X(f - f_s k) df \]

\[ \text{mmse}_{X|Y}(f_s) = \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \tilde{S}_{X|Y}(f) df \]
RESULT 1

EXAMPLE

\[ D(f_s, R) = \begin{cases} 
1 - \frac{f_s}{2W} + \frac{f_s}{2W} 2^{-\frac{2R}{f_s}} & f_s < 2W \\
2^{-R/W} & f_s \geq 2W 
\end{cases} \]
RESULT II

EXTENDED MODEL: PRE-SAMPLING FILTER

\[ X(\cdot) \xrightarrow{H(f)} Y[\cdot] \xrightarrow{f_s} \text{Encoder} \xrightarrow{R} \text{Decoder} \xrightarrow{\hat{X}(\cdot)} \]

Theorem [K. Goldsmith, Weissman, Eldar 2014]

\[
R(f_s, \theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left[ \frac{\tilde{S}_{X|Y}(f)}{\theta} \right] \, df
\]

\[
D(f_s, \theta, H) = \text{mmse}_{X|Y}(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min \left\{ \tilde{S}_{X|Y}(f), \theta \right\} \, df
\]

\[
\tilde{S}_{X|Y}(f) = \frac{\sum_{k \in \mathbb{Z}} S_X^2(f - f_s k) |H(f - f_s k)|^2}{\sum_{k \in \mathbb{Z}} S_X(f - f_s k) |H(f - f_s k)|^2}
\]

What is \( D^*(R, f_s) \triangleq \inf_H D(R, f_s, H) \)?
RESULT II: OPTIMAL PRE-SAMPLING FILTER

Theorem [K. Goldsmith, Eldar, Weissman 2014]

Optimal pre-sampling filter maximizes passband energy under an aliasing-free constraint. For unimodal PSD:

\[
R(\theta) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left[ S_X(f)/\theta \right] df
\]

\[
D^*(f_s, \theta) = \text{mmse}^*_X|Y(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min \{ S_X(f), \theta \} df
\]

- Suppresses lower energy bands
- When PSD is non-unimodal filter-bank sampling must be used
WHY ANTI-ALIASING IS OPTIMAL?

\[ X(\cdot) \xrightarrow{H(f)} Y[\cdot] \]

\[ X_1 \sim \mathcal{N}(0, \sigma_1^2) \xrightarrow{h_1} + \]

\[ Y = h_1 X_1 + h_2 X_2 \]

\[ X_2 \sim \mathcal{N}(0, \sigma_2^2) \xrightarrow{h_2} \]

\[ \text{mmse} \]

\[ \tilde{X}_1 \triangleq \mathbb{E}[X_1|Y] \]

\[ \tilde{X}_2 \triangleq \mathbb{E}[X_2|Y] \]

\[ \text{mmse}_{(X_1, X_2)|Y} = \left( X_1 - \tilde{X}_1 \right)^2 + \left( X_2 - \tilde{X}_2 \right)^2 \]

\[ = \frac{h_1^2 \sigma_1^4 + h_2^2 \sigma_2^4}{h_1^2 \sigma_1^2 + h_2^2 \sigma_2^2} \]

\[ \inf_{h_1, h_2} \text{mmse}_{(X_1, X_2)|Y} = \min \{ \sigma_1^2, \sigma_2^2 \} \]

\[ \begin{cases} 
  h_1 = 0, \ h_2 = 1, \ \sigma_1^2 < \sigma_2^2 \\
  h_1 = 1, \ h_2 = 0, \ \sigma_1^2 > \sigma_2^2
\end{cases} \]
RESULT II

OPTIMAL PRE-SAMPLING FILTER - EXAMPLE

\[ X(\cdot) \xrightarrow{H(f)} Y[\cdot] \xrightarrow{f_s} \text{Encoder} \xrightarrow{R} \text{Decoder} \xrightarrow{\hat{X}(\cdot)} \]

\[ D^*(R, f_s) \text{ and } D(R, f_s) \text{ vs } f_s \]

Distortion vs \( f_s \):

- \( D^*(R, f_s) \)
- \( D(R, f_s) \)

\[ \frac{f_s}{2} \]

\[ \frac{f_s}{2} \]

\[ H^*(f) \]

\[ \theta \]
COROLLARY:

OPTIMAL SAMPLING RATE

\[ D^*(R, f_s) \] vs \( f_s \)

\[ D^*(f_s, R) = D(R) \text{ for } f_s \geq f_{DR}(R) \ (\!\!\!) \]
RESULT III: OPTIMAL SAMPLING RATE IN ANALOG-TO-DIGITAL COMPRESSION

- A new sampling theorem for Gaussian stationary processes:

Theorem [K. Goldsmith, Eldar 2015]

\[ D^*(f_s, R) = D(R) \quad f_s \geq f_{DR}(R) \]

- Extends Shannon-Nyquist-Kotelinkov-Whittaker sampling theorem:
  - Incorporates lossy compression
  - Valid when \( X() \) is not bandlimited
  - Holds under non-uniform sampling

Interpretation: lossy compression reduces degrees of freedom
RESULTS III

OPTIMAL SAMPLING RATE - EXAMPLE

\[ D^*(f_s, R) \text{ as a function of } \frac{f_s}{f_{Nyq}} \]

\[ f_{DR}(R) \]

\[ \sigma_X^2 \]

\[ D^*(f_s, R) \text{ in } \text{[dB]} \]

\[ R = 0.5 \]

\[ R = 1 \]

\[ R = 2 \]

\[ R = 20 \]
RESULT III

OPTIMAL SAMPLING RATE - EXAMPLE

\[ D(R, f_s) = D(R) \text{ for } f_s \geq f_{DR}(R)! \]

\[ f_{DR} \text{ vs } R \]
SAMPLING NON-BANDLIMTED SIGNALS

\[ f_{Nyq} = \infty \]
\[ f_{DR}(R) < \infty \]

\[ MSE [\text{dB}] \]

\[ f_{s}/2 \]

\[ \frac{f_{s}}{2} \]

\[ f \]

\[ H^*(f) \]

\[ \theta \]

\[ f_{Nyq} = \infty \]

\[ f_{DR}(R) < \infty \]

\[ f_{Nyq} = \infty \]

\[ f_{DR}(R) \rightarrow \infty \]

Sampling a Brownian motion? ISIT 2016 ....
Constraining sampling frequency in the analog source coding problem leads to $D(\text{fs}, R) = \text{mmse} + \text{`water-filling’}$

Optimal pre-sampling filter eliminates aliasing

New critical sampling frequency $f$ performance $D(R)$ under bitrate constraints — typically below the Nyquist rate
THE END!

\[ D(R, f_s) \]
OPEN QUESTIONS AND FUTURE WORK

- Optimal pre-sampling filter with different loss criterion
- Unknown source statistic
- Effect of lossy compression on DoF in other signal model (e.g. sparse signals)
- Optimal sampling rate - bit allocation strategy in existing A/D schemes
- Linear pre-processing reduces sampling distortion. Can non-linear (time-preserving)
A. Kipnis, A. J. Goldsmith, T. Weissman and Y. C. Eldar, “Rate distortion of Gaussian stationary processes” 2015, available online

A. Kipnis, A. J. Goldsmith and Y. C. Eldar, “Sub-Nyquist sampling achieves optimal rate-distortion”

A. Kipnis, A. J. Goldsmith and Y. C. Eldar, “Distrotion-rate function under sub-Nyquist nonuniform samplig”

A. Kipnis, A. J. Goldsmith and Y. C. Eldar, “Rate-Distortion Function of Cyclostationary Gaussian Processes”

A. Kipnis, A. J. Goldsmith and Y. C. Eldar, “Optimal tradeoff between sampling rate and quantizer resolution in A/D conversion”
EXTENSION TO GENERAL PSD

\[ I(Y; \hat{X}) \leq R \]

\[ P_{Y|\hat{X}} \]

\[ \hat{X}(\cdot) \]

\[ Y[\cdot] \]

\[ X(\cdot) \]

\[ f_s \]

\[ H_1(f) \]

\[ H_P(f) \]

\[ D^*(f_s, R) \triangleq \lim_{P \to \infty} \inf_{H_1, \ldots, H_P} D_P(f_s, R) \]
Theorem [K. Goldsmith, Weissman, Eldar 2014]

\[
D^*(f_s, \theta) = \int_{-\infty}^{\infty} S_X(f) df - \int_{A^*} [S_X(f) - \theta]^+ df
\]

\[
R = \frac{1}{2} \int_{A^*} \log^+ [S_X(f)/\theta] df
\]

\[
\int_{A^*} S_X(f) df = \inf_{\lambda(A) \leq f_s} S_X(f) df
\]
**NONUNIFORM SAMPLING**

\[ \Lambda \triangleq \{t_1, t_2, \ldots \} \]

\[ d^-(\Lambda) = \lim_{T \to \infty} \frac{n^-(\Lambda)}{T} \]

\[ n^-(\Lambda) \triangleq \text{minimal no. of elements of } \Lambda \text{ in an interval of length } T \]

\[ D_\Lambda(R) \triangleq \text{minimal expected distortion using rate } R \text{ codes on } Y[n] \]

**Theorem [K. Goldsmith, Eldar 2014]**

\[ D_\Lambda(R) \geq D(d^-(\Lambda), R) \]

- Nonuniform sampling does not provide theoretical gain
- May still reduce pre-processing complexity
WHY ANTI-ALIASING IS OPTIMAL?

Analogy: aliasing in sub-Nyquist sampling

\[ S_Y(f) = \sum_k H^*(f - f_s k) S_X(f - f_s k) \]

PASS ONLY THE HIGHEST FREQUENCIES, SUPPRESS THE REST