

MODELING AND CODING OF SPEECH AND AUDIO SIGNALS

Bastiaan Kleijn

KTH School of Electrical Engineering

Stockholm

This talk is largely based on the book chapter:

W.B. Kleijn, "Principles of Speech Coding", in: J. Benesty, A. Huang, M. Sondhi, Eds., "Speech Processing", Springer, 2007 (in press).

KTH Sound and Image Processing

- KTH (14000 students, 250 Professors)
- School of Electrical Engineering
- Sound and Image Processing Lab
 - 2 professors
 - 1 visitor
 - 10 PhD students
 - 1 administrator
 - 5 undergraduate courses
 - Funding from
 - EU (HEARCOM, ACORNS, FLEXCODE)
 - Swedish Research Council (VR)
 - SSF
 - Swedish Industry (Ericsson, GN Resound, Global IP Sound)



Motivation for Topic

- Code length is useful measure of goodness for models
 - Use coding to find models for recognition/modification etc.
- Heterogeneous networks require continuously adaptive coding
 - Analytic solutions necessary
 - Replace data/codebooks with understanding (models)
 - Exploit knowledge of models to code efficiently

- Modeling of a signal
- Coding as a motivation for model selection
- Universal modeling
- Application to autoregressive coding
- The role of the distortion measure

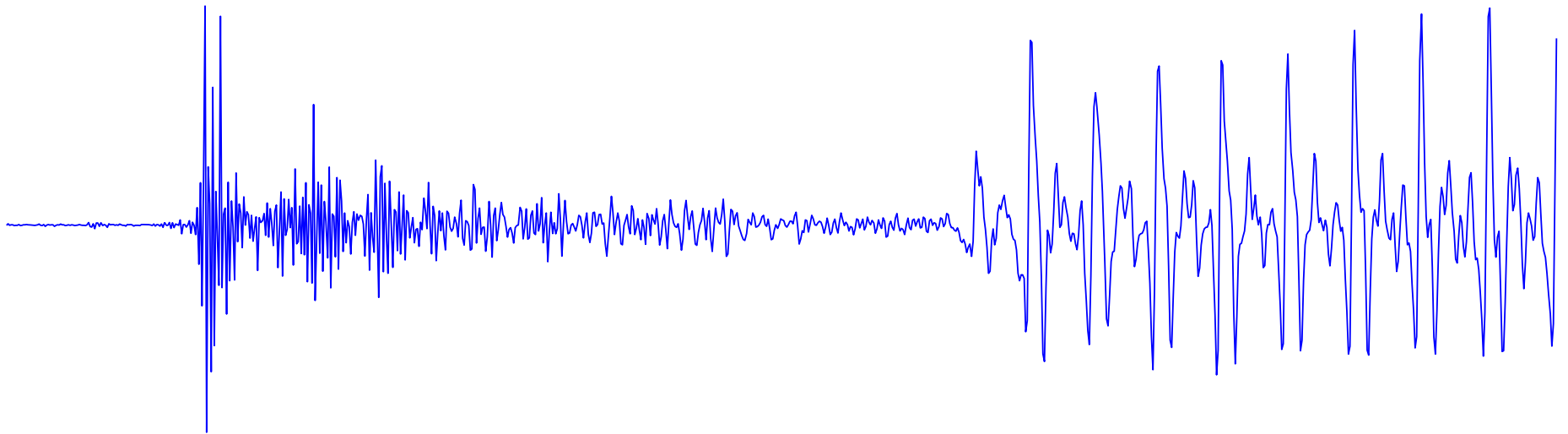
A Model

An abstract **model** (or conceptual model) is a theoretical construct that represents something, with a set of variables and a set of logical and quantitative relationships between them. Models in this sense are constructed to enable reasoning within an idealized logical framework about these processes and are an important component of scientific theories. *Idealized* here means that the model may make explicit assumptions that are known to be false (or incomplete) in some detail. Such assumptions may be justified on the grounds that they simplify the model while, at the same time, allowing the production of acceptably accurate solutions.

What is a Model for?

- Model enhances description efficiency: alternative is to store data (TTS)
 - Replaces data with knowledge
- Classify the signal based on model parameters and not the raw signal
 - Speech recognition
 - Speaker recognition
- Modify the structure of the signal by changing model parameters
 - Change speaker identity
- Code more efficiently by knowing the model:
 - Reduce dynamic range / reduce support
 - Facilitates scalar quantization through decorrelation
 - Reduce complexity through selection of one of many models

A Signal



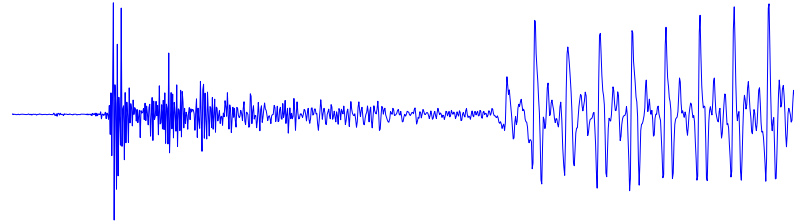
from a billion to one thousand
(30 bits to 10 bits)

What is a Model?

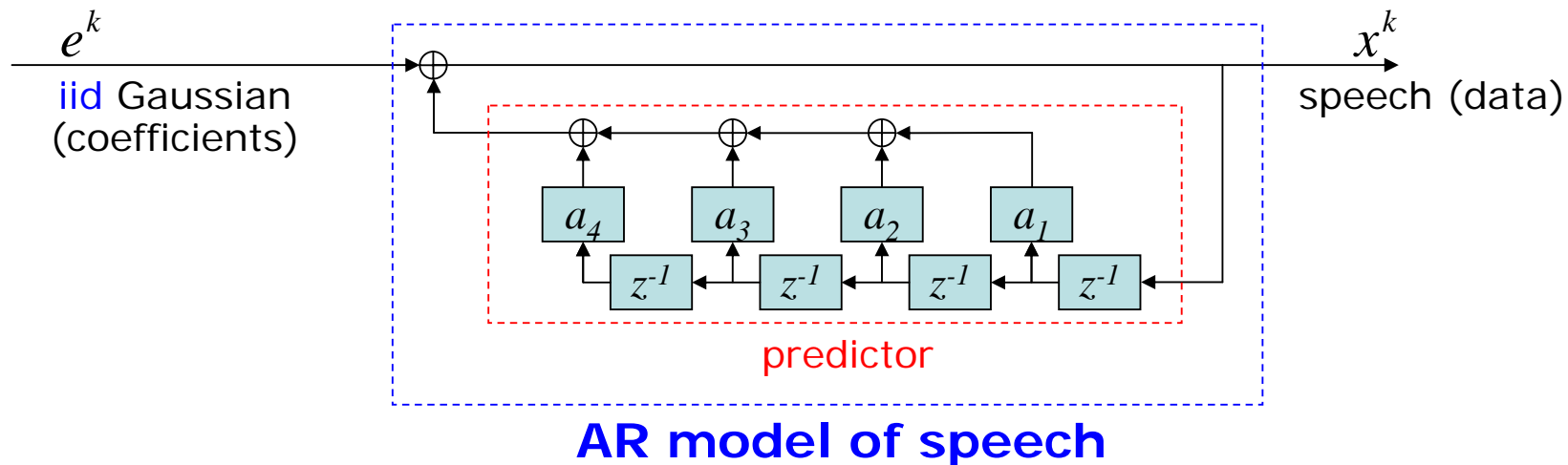
- Model characterizes *structure* in a signal segment
- Structure described by *model family* and model *parameters*
 - Parameters define a particular *model* within a *model family*
 - The class of moving average (MA) models of order 14
- Signal segment described by
 - model family, model parameters, and coefficients
 - model family, model parameters, and signal data

Speech Example I

- Speech processing: extracts/modifies/exploits *structure* of signal segment
 - Short-term correlations
 - Long-term correlations

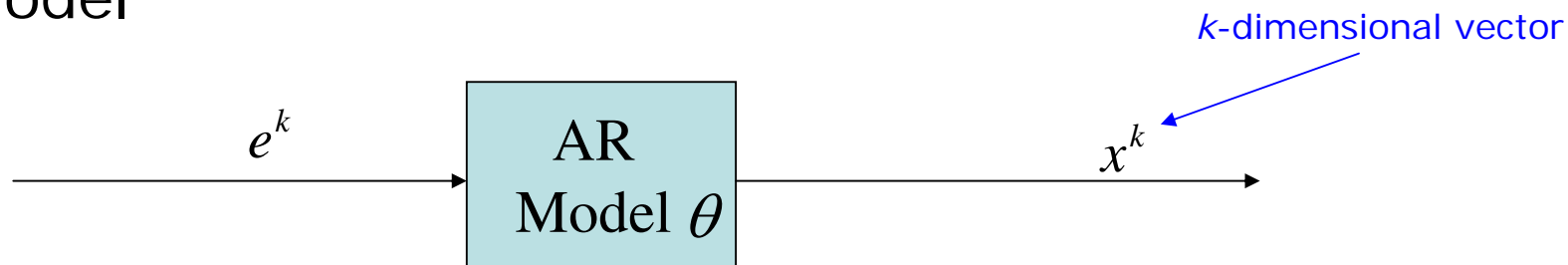


- Example: *particular* speech segment described by
 1. Autoregressive (AR) model of order 10: model class
 2. AR parameters: parameters: $\theta = [a_1, a_2, \dots, a_p]$
 3. excitation signal: coefficients (presumed *unstructured*)



Speech Example II

- AR model



- Implication: $A = A(\theta)$

$$R_X = E[x^k x^{kH}] = E[A e^k e^{kH} A^H] = \sigma^2 A A^H$$

- A is Toeplitz, since AR model is a linear filter power spectrum
 - Circulant approximation: $F^H R_{X^k} F = \text{diag}(P_A)$

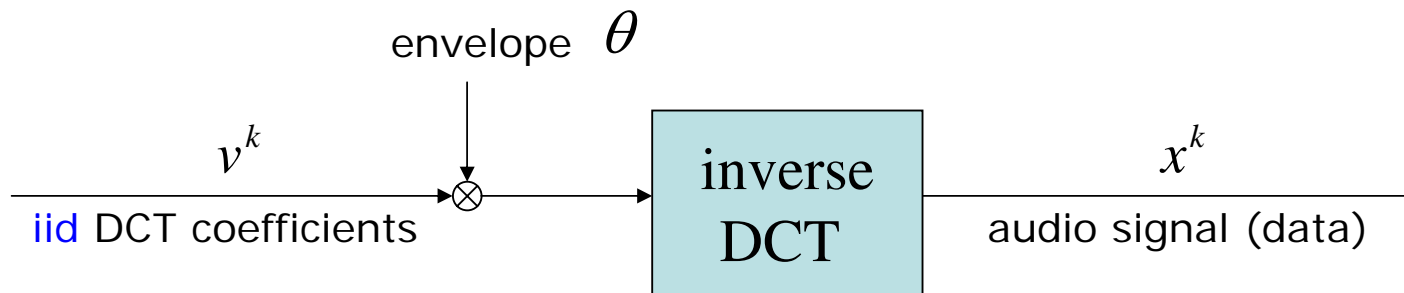
- Data distribution: $p_{X^k}(x^k) = \frac{1}{\sqrt{2\pi \det(R_{X^k})}} \exp\left(-\frac{1}{2} x^k R_{X^k}^{-1} x^k\right)$

Speech Example III

- How to describe particular speech segment:
 - AR model, AR parameters, coefficients
 - Coefficients Gaussian iid
 - Distortion measure modified by model; problem
 - Gaussian probability distribution, speech data
 - Probability distribution specified by AR model family and AR parameters
- $$p_{X^k}(x^k) = \frac{1}{\sqrt{2\pi \det(R_{X^k})}} \exp\left(-\frac{1}{2} x^k R_{X^k}^{-1} x^k\right)$$
- Distortion measure defined directly on speech

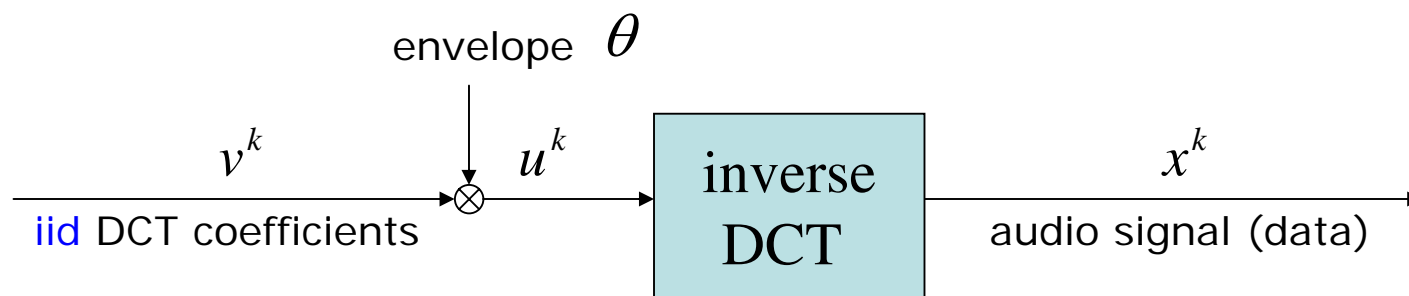
Audio Example I

- Audio processing: extracts/modifies/exploits *structure* of signal
 - Short term power spectrum
- Example: *particular* audio segment described by
 1. Envelope θ
 2. Normalized discrete cosine transform (DCT) coefficients



DCT with masking curve "model" of audio

Audio Example II



- Particular audio segment described by
 1. Envelope θ Model: DCT and envelope
 2. Normalized DCT coefficients

- Data distribution stationary signal

$$p_{X^k}(x^k) = p_{U^k}(u^k(x^k)) = \prod_{i=1, \dots, k} p_{U_i}(u_i(x^k)) = \prod_{i=1, \dots, k} p_{V_i}(\alpha(\theta) u_i(x^k))$$

Audio Example III

- How to describe audio segment:
 - Envelope, scaled DCT coefficients
 - Coefficients Gaussian iid
 - Distortion measure modified by model; problem
 - Envelope, *unscaled* DCT coefficients
 - Coefficients Gaussian and independent
 - Distortion equivalent to speech domain
 - Envelope, speech vector
 - Speech data dependent

$$\{p_{U_i}(u_i)\}_{i=\{1,\dots,k\}}$$

A Signal Model is:

- Model characterizes structure in a *particular* signal segment by
 - Model class
 - Model parameters

- Signal structure \longleftrightarrow probability density function
 - Model \longleftrightarrow Model class + parameters
 - \longleftrightarrow probability density function $p_{X^k}(x^k)$

- Signal described by
 - Model class, model parameters, and coefficients / signal data
 - Model makes description of signal data “more efficient”

use efficiency to select model?

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection
- Universal modeling
- Application to autoregressive coding
- The role of the distortion measure

- What is a *good* model class for speech?
 - Current approach: evaluate performance in application
 - Coding
 - Modification
 - Recognition
 - Not unified / generic
 - Model probability density should be good “fit”
 - How to measure “fit”
 - Prefer signal-domain distortion measure
 - Minimum Description Length (MDL) approach:
 - Optimal code depends on probability density
 - Code length is a measure of goodness for probability density
 - A good model for specific signal segment facilitates efficient coding

How to Select a Model

- Minimum Description Length (MDL)* approach: efficient description for specific signal segment (sequence) x^k
 - Select segment/sequence (and distortion criterion + threshold)
 - For each class, select model that minimizes mean code length for sequence
 - Select model class with best performing model

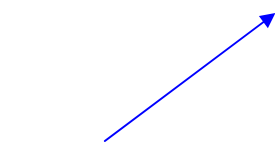
* J.Rissanen, Universal coding, information, prediction, and estimation, *IEEE*

Trans. Information Theory, vol. 30 (1984), pp. 629-636

Code Length (No Distortion/Lossless)

- Assume constrained-entropy coding
 - “rate” = average rate
- Kraft inequality is condition on uniquely decodable code; shows that shortest mean code length requires codeword length $-\log(P(x^k))$
- Example: $P(x^k) = 0.625 \Leftrightarrow -\log_2(0.625) = 4$
- If the distribution is not right, then the average rate is too high, but the code is still unique

$$-\sum_{x^k} P(x^k) \log(P(x^k | \theta)) = \underbrace{\sum_{x^k} P(x^k) \log\left(\frac{P(x^k)}{P(x^k | \theta)}\right)}_{\substack{\text{Kulback-Leibler dist} \\ > 0}} - \underbrace{\sum_{x^k} P(x^k) \log(P(x^k))}_{H(X^k)}$$


 model

Rudimentary MDL Example

- *Particular* sequence: $x^k = [0, 1, 3, 1, 1, 1, 1, 2]$ (“the speech block”)
 - Model A: $P(x_i | A) = 0.25$
 - Code length = $2 * 8 = 16$ bits
 - Model B: $P(0 | B) = 0.125$ $P(2 | B) = 0.125$
 $P(1 | B) = 0.5$ $P(3 | B) = 0.125$
 - Code length = $3 * 3 + 5 * 1 = 14$ bits
- Conclusion: model B is a better model
 - If model family contains A and B, then B is the *maximum likelihood* solution

$$P(x^k | A) = 0.25^8 = 0.000015$$

$$P(x^k | B) = 0.5^5 * 0.125^3 = 0.000061$$

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection **coding = max likelihood**
- **Universal modeling**
 - Lossless
 - With distortion
 - Distribution of rate between model and data
 - Fixed-rate coders
- Application to autoregressive coding
- The role of the distortion measure

Two-Stage MDL

- Consider sequence x^k
- Two-stage MDL

$$L_A = \underbrace{L(\tilde{\theta}(x^k))}_{\text{quantized parameter code length}} - \underbrace{\log(P(x^k | \tilde{\theta}(x^k)))}_{\text{code length using quantized parameters}}$$

quantized
parameter
code length

code length using
quantized parameters

maximum likelihood model

$$= L(\tilde{\theta}(x^k)) + \underbrace{\log\left(\frac{P(x^k | \hat{\theta}(x^k))}{P(x^k | \tilde{\theta}(x^k))}\right)}_{\text{regret=excess code length not a Kulback-Leibler distance}} - \underbrace{\log(P(x^k | \hat{\theta}(x^k)))}_{\text{optimal coefficient code length}}$$

regret=excess code length
not a Kulback-Leibler distance

optimal coefficient
code length

index of resolvability

Rudimentary Lossless Example

- *Particular* sequence: $x^k = [0, 1, 3, 1, 1, 1, 1, 2]$ (“the speech block”)

- Model A: $P(x_i | A) = 0.25$

likelihood of A

– Code length: $-\log(P(x^k | A)) = -\log\left(\prod_{i=1, \dots, 8} P(x_i | A)\right) = 16$

- Model B: $P(0 | B) = 0.125$ $P(2 | B) = 0.125$
 $P(1 | B) = 0.5$ $P(3 | B) = 0.125$

– Code length: $-\log(P(x^k | B)) = -\log\left(\prod_{i=1, \dots, 8} P(x_i | B)\right) = 14$

- *Regret* for model A: $\log\left(\frac{P(x^k | B)}{P(x^k | A)}\right) = 2$

- Should also include cost of coding the model (e.g., 1 bit for each)

Coding an Ensemble of Sequences

- Consider distribution of sequences $p_{X^k}(x^k)$, $x^k \in \mathbb{Z}^k$
 - Each sequence satisfies one of a set of distributions
- Objective
 - Select model family with lowest *mean* rate
 - Select model with lowest rate for each x^k
 - We must select a (sub)set of models for each model family
- Coding benefit of decomposing into *mixture of components/models*:
 - Reduced codebook size
 - Simple component distributions
 - Insight in rate distribution model and sequence-given-model
- Modeling benefit
 - Select best model family for “speech”, “audio”, “jazz”
 - Provides insight in selection subset of models
 - Leads to faster design

$$p_{X^k}(x^k) = \sum_{\xi} p_{\Xi}(\xi) p_{X^k|\Xi}(x^k | \xi)$$

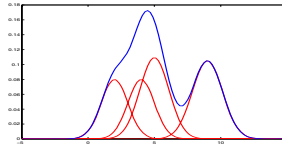
prior probability model family

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection **coding = max likelihood**
- **Universal modeling**
 - Lossless **split into simple models**
 - **With distortion**
 - Distribution of rate between model and data
 - Fixed-rate coders
- Application to autoregressive coding
- The role of the distortion measure

Coding an Ensemble of Sequences

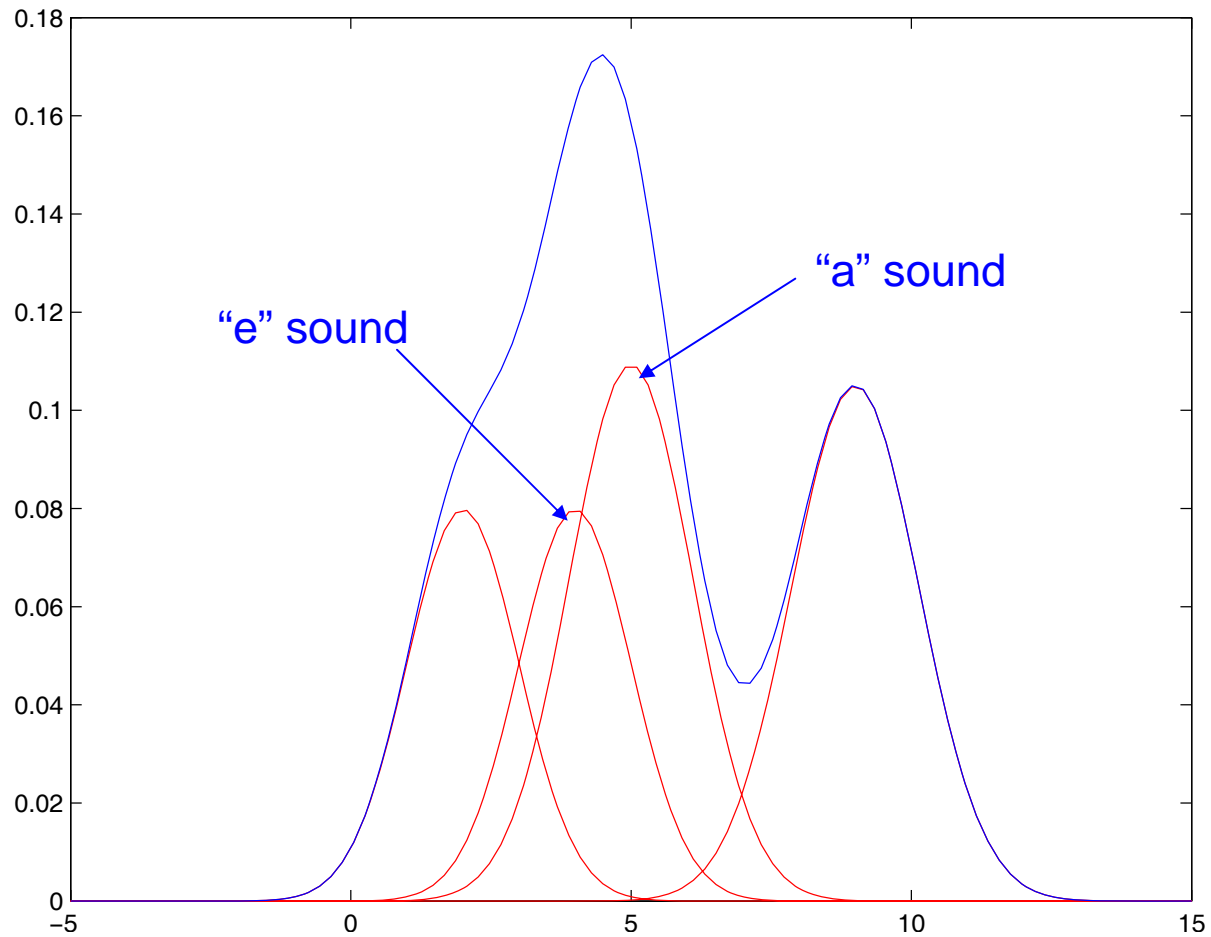
- Consider distribution of sequences $p_{X^k}(x^k)$, $x^k \in \mathbb{R}^k$
 - Each sequence satisfies one of a set of distributions



$$p_{X^k}(x^k) = \int p_{\Xi}(\xi) p_{X^k|\Xi}(x^k | \xi) d\xi$$

- Objective: lowest mean rate over distribution, *given a distortion*
 - Select model family with lowest mean rate given distortion
 - Select model with lowest rate for each x^k
 - Must have a **countable set** of models for each model family
- Coding benefit of decomposing into *mixture of components/models*:
 - Reduced codebook size
 - Simple component distributions
 - Insight in rate distribution model and sequence-given-model
- Modeling benefit
 - Select best model family for “speech”, “audio”, “jazz”
 - Provides insight in selection subset of models
 - Leads to faster design

Mixture Model



Relation to Reality: Speech

- Consider distribution of 20 ms speech segments $p_{X^k}(x^k)$, $x^k \in \mathbb{R}^k$
 - Each segment corresponds to a speech sound
 - Each sound has a distribution $p_{X^k|\Xi}(x^k | \xi)$
 - Each sound has a probability density $p_{\Xi}(\xi)$
 - Speech segment has distribution

$$p_{X^k}(x^k) = \int p_{\Xi}(\xi) p_{X^k|\Xi}(x^k | \xi) d\xi$$

- We neglect all the dependencies between speech segments
- Objective: lowest mean rate over distribution, *given a distortion*
 - Select model family with lowest mean rate given distortion
 - Select model with lowest rate for each x^k and average
 - Must have a **countable set** of models for each model family
- Modeling benefit
 - Select best model family for speech
 - Provides insight in selection subset of models
 - Leads to faster design

Coding with Distortion

- We *assume* optimal quantization of the sequence
 - Constrained resolution -> density to power $k/(k+2)$ (normalized)
 - Constrained entropy -> uniform
- Quantization cell size determines distortion; for squared error:

$$D = CV^{\frac{2}{k}}$$

- Codeword length for quantization cell of volume V is

$$-\log(Vp_{X^k}(x^k))$$

Two-Stage MDL with Distortion

- Consider sequence x^k
- Two-stage MDL for discrete data (code length)

$$L_A = \underbrace{-\log(p_{\Theta}(\tilde{\theta}(x^k)))}_{\text{code length using quantized parameters}} - \underbrace{\log(V p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k)))}_{\text{signal code length}}$$

code length using
quantized parameters

signal code length

maximum likelihood model

$$= \underbrace{-\log(p_{\Theta}(\tilde{\theta}(x^k))) + \log\left(\frac{p_{X^k|\Theta}(x^k | \hat{\theta}(x^k))}{p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k))}\right)}_{\text{regret=excess code length}} - \underbrace{\log(V p_{X^k|\Theta}(x^k | \hat{\theta}(x^k)))}_{\text{signal code length for optimal model}}$$

index of resolvability

signal code length
for optimal model

Universal Coding

- Consider distribution of sequences

$$p_{X^k}(x^k) = \int p_{\Xi}(\xi) p_{X^k|\Xi}(x^k | \xi) d\xi$$

- Consider model family and associated countable set of models
- Select model family that minimizes cost

$$E[L_A] = -E[\log(p_{\Theta}(\tilde{\theta}(x^k)))] - E[\log(V p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k)))]$$

$$= -E[\log(p_{\Theta}(\tilde{\theta}(x^k)))] + E\left[\log\left(\frac{p_{X^k|\Theta}(x^k | \hat{\theta}(x^k))}{p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k))}\right)\right] - E[\log(V p_{X^k|\Theta}(x^k | \hat{\theta}(x^k)))]$$

mean index of resolvability

only term relating to distortion

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection **coding = max likelihood**
- **Universal modeling**
 - Lossless **split into simple models**
 - With distortion **model set not related to distortion**
 - **Distribution of rate between model and data**
 - Fixed-rate coders
- Application to autoregressive coding
- The role of the distortion measure

Finding the Model Family

- Select model family that minimizes minimum cost

$$\begin{aligned}
 E[L_A] &= -E[\log(p_{\Theta}(\tilde{\theta}(x^k)))] - E[\log(V p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k)))] \\
 &= -E[\log(p_{\Theta}(\tilde{\theta}(x^k)))] + E\left[\log\left(\frac{p_{X^k|\Theta}(x^k | \hat{\theta}(x^k))}{p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k))}\right)\right] - E[\log(V p_{X^k|\Theta}(x^k | \hat{\theta}(x^k)))]
 \end{aligned}$$

- Select countable model set = minimize mean index of resolvability

$$\eta = -E[\log(p_{\Theta}(\tilde{\theta}(x^k)))] + E\left[\log\left(\frac{p_{X^k|\Theta}(x^k | \hat{\theta}(x^k))}{p_{X^k|\Theta}(x^k | \tilde{\theta}(x^k))}\right)\right]$$

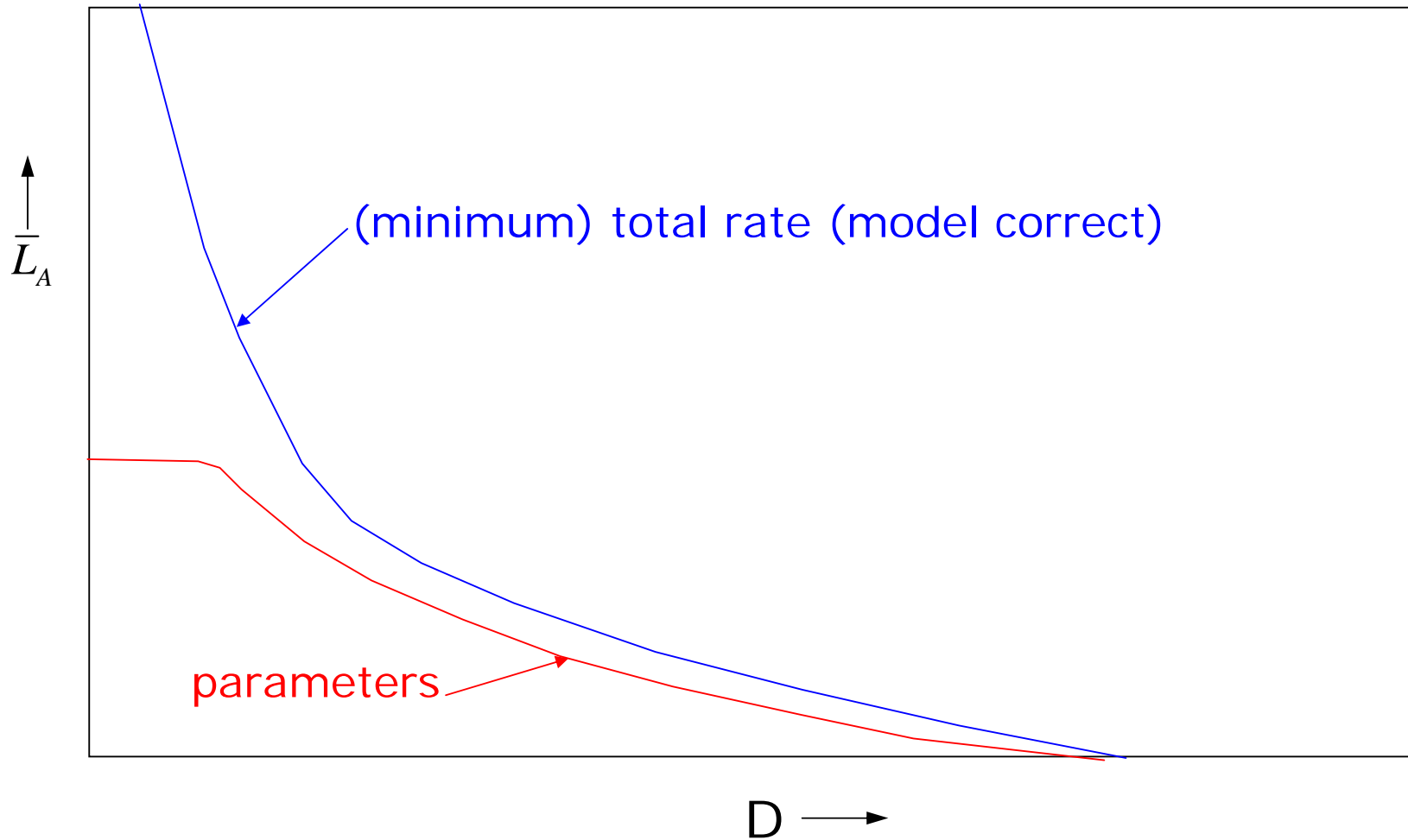
increases with model precision

decreases with model precision

Approach to Universal Coding

- Select a set of model families
 - Autoregressive
 - MLT with envelope
 - Sinusoidal
- Design optimal set of models in each family
 - Is codebook of model parameters
 - Not dependent on overall distortion
- Coding approach as before:
 - Select model from family subset
 - index of resolvability
 - Encode signal vector using model

R(D) and Modeling



From low distortion to high distortion is
 from hybrid coding to parametric coding

Notes on Universal Coding

- The model family does *not* have to contain the *generating* model
- Theory was for variable-rate (entropy-constrained) coding
- Most coders are still fixed-rate (resolution-constrained)

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection **coding = max likelihood**
- **Universal modeling**
 - Lossless **split into simple models**
 - With distortion **model set not related to distortion**
 - Distribution of rate between model and data **fixed rate for model**
 - **Fixed-rate coders**
- Application to autoregressive coding
- The role of the distortion measure

Fixed Rate

- Coding a sequence x^k with fixed-rate allocation for sequence and for model:

$$\begin{aligned}
 L &= L_m + L(x^k) \\
 &= L_m + \log(N) \\
 &= L_m - \mathbb{E} \left[\log \left(\frac{p_{X^k|\Theta}(X^k | \tilde{\theta})^{\frac{k}{k+2}}}{p_{X^k|\Theta}(X^k | \hat{\theta})^{\frac{k}{k+2}}} \right) \right] - \frac{k}{2} \log \left(\frac{D_{CR}}{C} \right) \\
 &= L_m + \underbrace{\mathbb{E} \left[\log \left(\frac{p_{X^k|\Theta}(X^k | \hat{\theta})^{\frac{k}{k+2}}}{p_{X^k|\Theta}(X^k | \tilde{\theta})^{\frac{k}{k+2}}} \right) \right]}_{\text{regret=excess code length}} - \underbrace{\mathbb{E} \left[\log \left(\frac{p_{X^k|\Theta}(X^k | \hat{\theta})^{\frac{k}{k+2}}}{p_{X^k|\Theta}(X^k | \tilde{\theta})^{\frac{k}{k+2}}} \right) \right] - \frac{k}{2} \log \left(\frac{D_{CR}}{C} \right)}_{\text{signal code length for optimal model}}
 \end{aligned}$$

mean index of resolvability

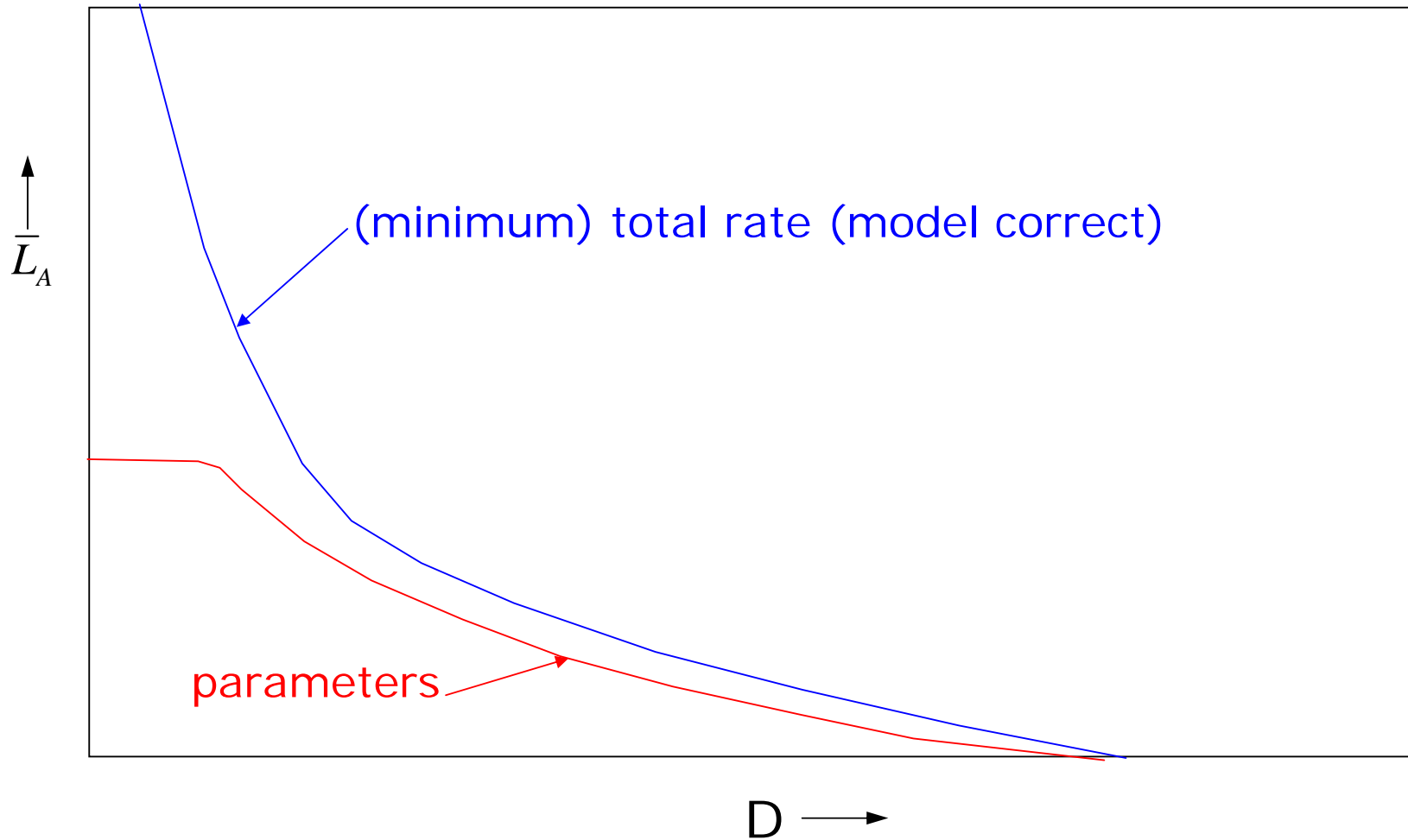
Fixed Rate for “Sufficiently” Large k

- Coding a sequence x^k with fixed-rate allocation for sequence and for model:

$$\begin{aligned}
 L &= L_m + L(x^k) \\
 &= L_m + \mathbf{E} \left[\log \left(\frac{p_{X^k|\Theta}(X^k | \hat{\theta})}{p_{X^k|\Theta}(X^k | \tilde{\theta})} \right) \right] - \mathbf{E} \left[\log \left(p_{X^k|\Theta}(X^k | \hat{\theta}) \right) \right] - \frac{k}{2} \log \left(\frac{D_{CR}}{C} \right)
 \end{aligned}$$

- Expression for regret is identical as for variable rate (constrained-entropy) case
- Rate for model depends only on first two terms:
 - Independent of distortion = independent of overall rate

R(D) and Modeling



From low distortion to high distortion is
 from hybrid coding to parametric coding

A Practical Coder: AMR-WB*

	Rate, kb/s	6.6	8.85	12.65	14.25	15.85	18.25	19.85	23.05
Model Parameters	AR model	36	46	46	46	46	46	46	46
	pitch	23	26	30	30	30	30	30	30
	gains	24	24	28	28	28	28	28	28
	LTP flag	0	0	4	4	4	4	4	4
	VAD flag	1	1	1	1	1	1	1	1
Coefficients	excitation	48	80	144	176	208	256	288	352

* AMR-WB coder uses 20 ms blocks

AMR Wide-Band

- Adaptive multi-rate codec (ETSI, 3GPP, 2001)
 - Used in GSM and WCDMA 3 systems
 - *Model optimized for one criterion, excitation for another*; heuristic balancing
- Low-rate coders:
 - Emphasize parameters over coefficients
 - Suggests: *proper criterion* and *good model*: no specification coefficients might be needed (variances are parameters)

Overview

- Modeling of a signal **model = probability distribution**
- Coding as a motivation for model selection **coding = max likelihood**
- Universal modeling
 - Lossless **split into simple models**
 - With distortion **model set not related to distortion**
 - Distribution of rate between model and data **fixed rate for model**
 - Fixed-rate coders **no change in conclusions**
- **Application to autoregressive coding**
- The role of the distortion measure

Coding with Autoregressive Models

- Autoregressive models are used in essentially all mobile telephones
- Interesting application of the theory
 - What does the index of resolvability correspond to?
- Our model assumption is that the signal is Gaussian
 - from slide 6:

$$p_{x^k|\Theta}(x^k) = \frac{1}{\sqrt{2\pi \det(R_\Theta)}} \exp\left(-\frac{1}{2} x^k R_\Theta^{-1} x^k\right)$$

- For large k , we can show:

$$\log\left(p_{x^k|\Theta}(x^k | \theta)\right) \approx \frac{1}{2} \log(2\pi) - \frac{k}{4\pi} \int_0^{2\pi} \log(R_\theta(e^{j\omega})) d\omega - \frac{k}{4\pi} \int_0^{2\pi} \frac{R_\theta(e^{j\omega})}{R_X(e^{j\omega})} d\omega$$

I of Resolvability Autoregressive Models

- Index of resolvability (33):

$$\psi = L_m + \mathbf{E} \left[\log \left(\frac{p_{X^k|\Theta}(X^k | \hat{\theta})^{\frac{k}{k+2}}}{p_{X^k|\Theta}(X^k | \tilde{\theta})^{\frac{k}{k+2}}} \right) \right]$$

$$= L_m + \underbrace{\frac{k}{4\pi} \int_0^{2\pi} -\log \left(\frac{R_{\hat{\theta}}(e^{j\omega})}{R_{\theta}(e^{j\omega})} \right) + \left(\frac{R_{\hat{\theta}}(e^{j\omega})}{R_{\theta}(e^{j\omega})} - 1 \right) R_W(e^{j\omega}) d\omega}_{\text{Itakura-Saito criterion if } R_W=1}$$

$R_W=1$ and small spectral error

$$\approx L_m + \underbrace{\frac{k}{8\pi} \int_0^{2\pi} \left(\log(R_{\hat{\theta}}(e^{j\omega})) - \log(R_{\theta}(e^{j\omega})) \right)^2 d\omega}_{\text{mean square log spectral error}}$$

Criterion as Before, but with Threshold

- Index of resolvability:

$$\begin{aligned}\psi &\approx L_m + \frac{k}{8\pi} \int_0^{2\pi} \left(\log(R_{\hat{\theta}}(e^{j\omega})) - \log(R_{\theta}(e^{j\omega})) \right)^2 d\omega \\ &= L_m + D(R_{\hat{\theta}}, R_{\theta})\end{aligned}$$

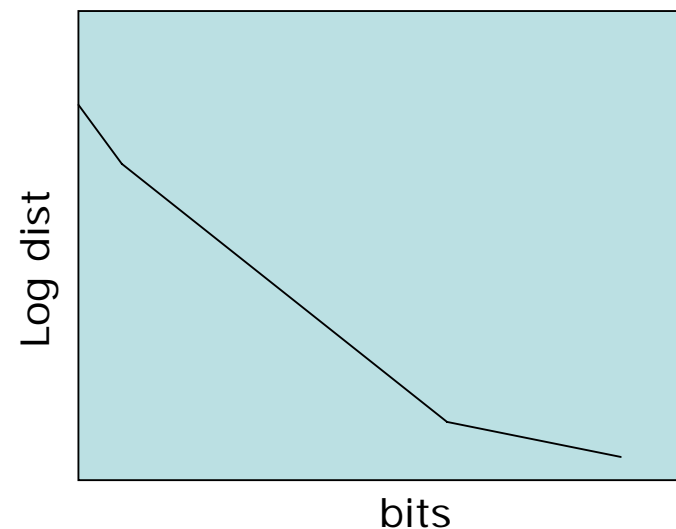
- Minimizing index of resolvability is like minimizing model rate L_m under model distortion constraint (like Lagrange multipliers) = minimizing model parameter rate under constraint on parameter distortion = *precisely what is aimed for in existing linear-predictive coding!*
- But we have more: we *know* the optimal parameter rate; where derivative of ψ to L_m vanishes. Contrasts with existing methods, which select a distortion on an empirical ("it sounds good") basis

Resulting Threshold

- Index of resolvability:

$$\psi \approx L_m + \frac{k}{8\pi} \int_0^{2\pi} (\log(R_{\hat{\theta}}(e^{j\omega})) - \log(R_{\theta}(e^{j\omega})))^2 d\omega$$

- Second term known in literature (Paliwal-Kleijn 1995)
- Threshold 2.5 dB = 10 bits
- Conclusion
 - *Outliers are most important!*



Squared Error

- Simple squared error

$$\eta = (x^k - \hat{x}^k)^H (x^k - \hat{x}^k)$$

- Weighted squared error

$$\eta = (x^k - \hat{x}^k)^H H^H H (x^k - \hat{x}^k)$$

- Filtering (neglect edges) $H = FWF^H$

diagonal

$$\begin{aligned} \eta &= (x^k - \hat{x}^k)^H F^H W^H F F^H W F (x^k - \hat{x}^k) \\ &= (x^k - \hat{x}^k)^H F^H W^2 F (x^k - \hat{x}^k) \end{aligned}$$

The Sensitivity Matrix

- If for \hat{x} near x

$$\eta = D(x - \hat{x}) = \alpha + \beta(x - \hat{x}) + \gamma^H |x - \hat{x}|^2$$

- and minimum distortion is zero at $x - \hat{x} = 0$

- Then $\alpha = 0$ and $\beta = 0$

- And so we can write

$$\eta = D(x - \hat{x}) \approx \gamma^H |x - \hat{x}|^2$$

- Generalizes to

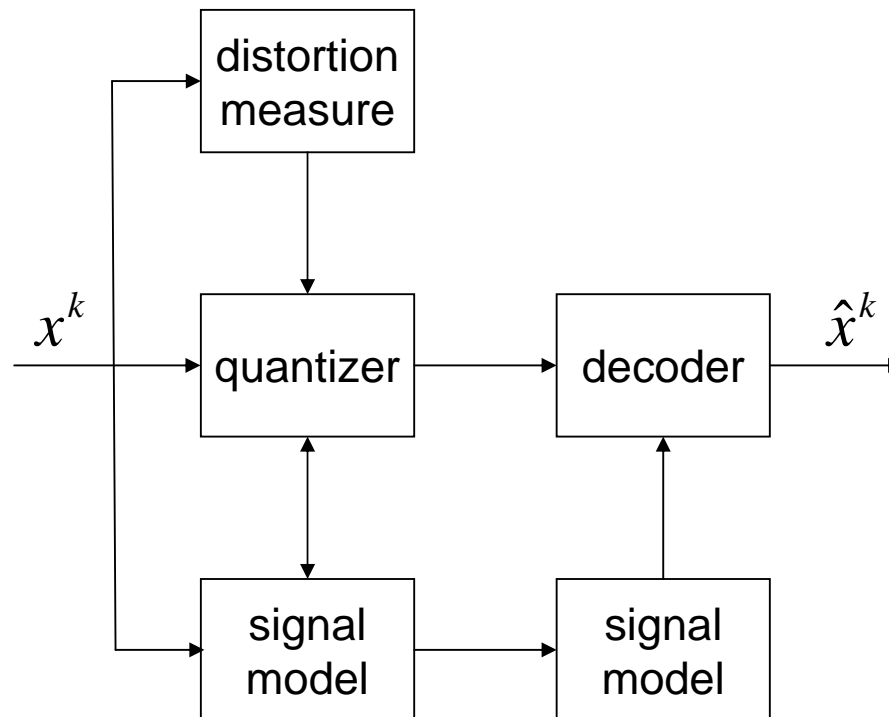
$$\eta = D(x^k - \hat{x}^k) \approx (x^k - \hat{x}^k)^H H(x^k)^H H(x^k) (x^k - \hat{x}^k)$$

Straightforward Usage of Distortion

- Weighted squared error approximates general error:

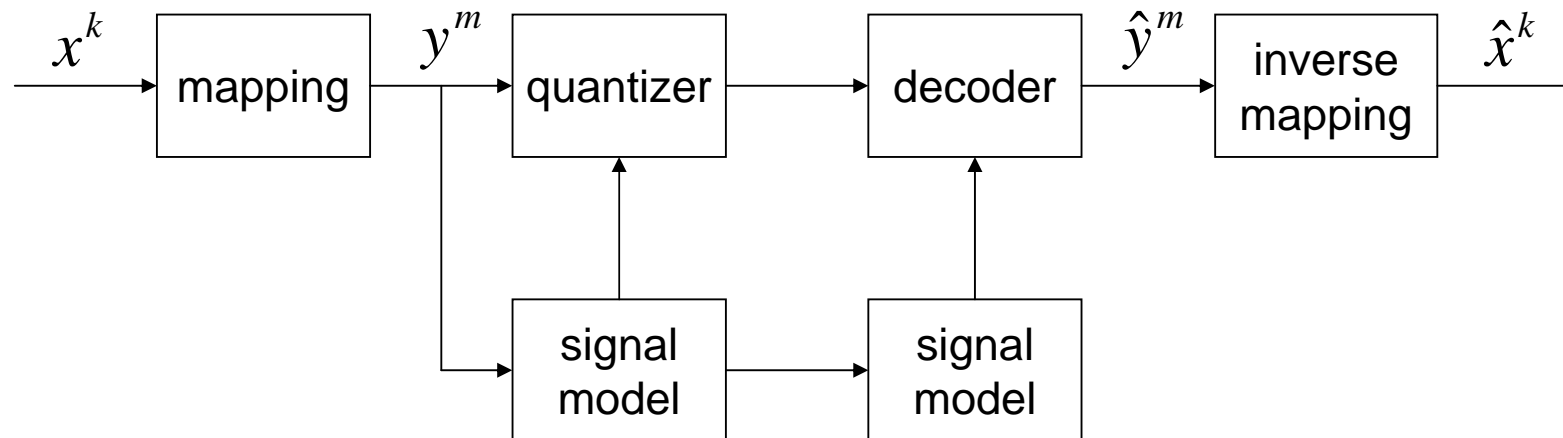
$$\eta = D(x^k - \hat{x}^k) \approx (x^k - \hat{x}^k)^H H(x^k)^H H(x^k) (x^k - \hat{x}^k)$$

- Unfortunate: criterion time varying; analysis, optimization difficult
- Common in speech coding (but not derived from sophisticated criteria)



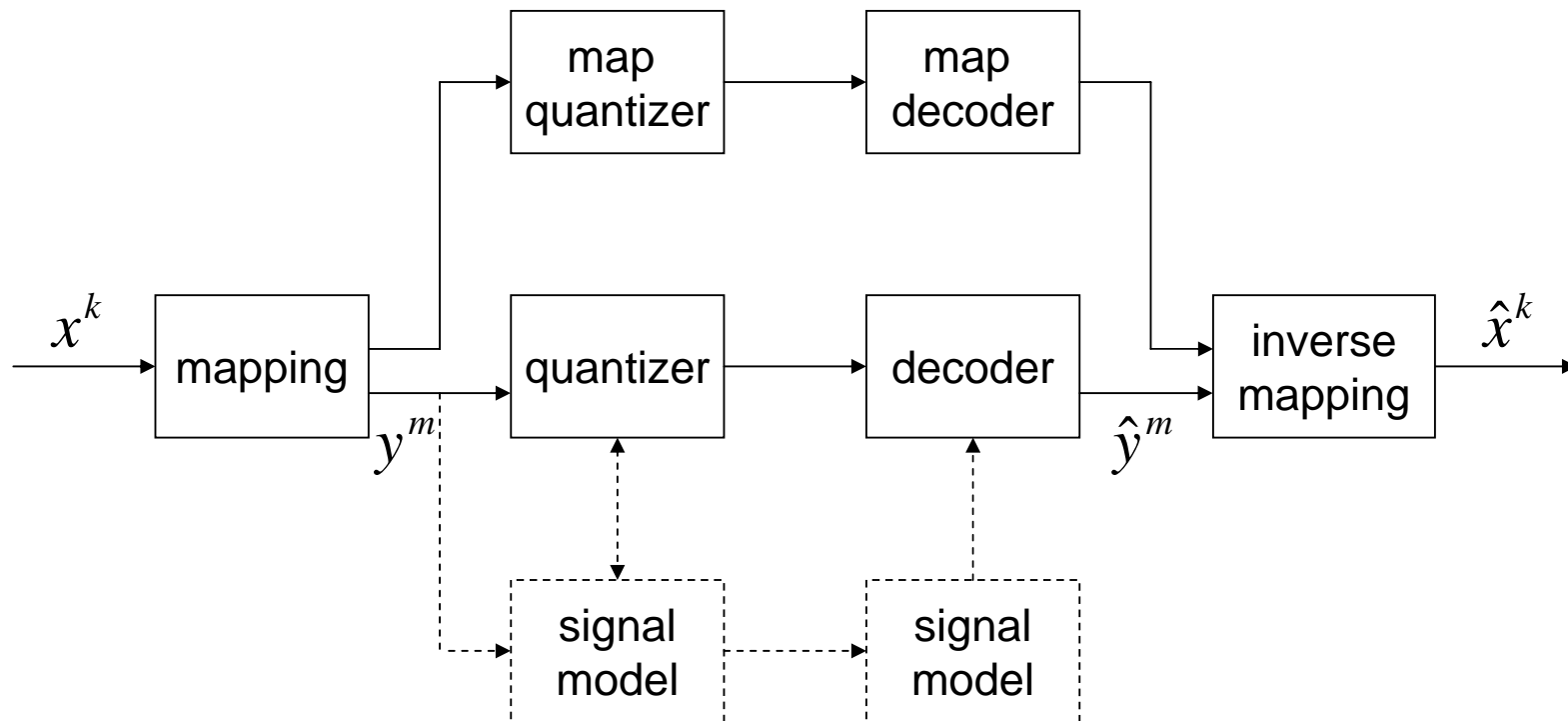
Ideal Perception in Coding

- Mapping represents auditory periphery
 - Must invert auditory periphery mapping
 - Mapping signal-dependent and *not* one-to-one (injective)
 - May require long delay
- Unweighted squared error in perceptual domain
- Dimensionality in perceptual domain an issue (data on manifold)



Perception in Audio Coding

- Mapping represents auditory periphery
 - Common: masking curve in frequency domain
 - Can be: index for sophisticated criterion
- Squared error reasonable after auditory periphery
- Dimensionality of perceptual domain is practical problem (data on manifold)
- Signal model is distribution of frequency domain coefficients



Distortion Measures in Coding

- Squared error reasonable
 - Mapping to perceptual domain
 - With mapping information (practical)
 - Without mapping information (problems)
 - Time-varying weighted squared-error criterion
 - Difficult to analyze

Recall Motivation

- Code length is useful measure of goodness for models
 - Use coding to find models for recognition/modification etc.
- Heterogeneous networks require continuously adaptive coding
 - Analytic solutions necessary
 - Replace data/codebooks with understanding (models)
 - Exploit knowledge of models to code efficiently

Conclusions

- Generic modeling = shortest mean code length for ensemble of signal segments, using proper criterion
- Adaptive coding: optimal bit rate for model parameters independent of overall rate (high rate)
- Existing speech coders satisfy expected general behavior
- Estimated model rate much lower than practice:
 - Mean distortion not important -> outliers are
- Distortion measure somewhat problematic; ideal is invertible auditory periphery, practical is to send information about mapping