Chirplets:
Multiscale Recovery and Detection of Chirps

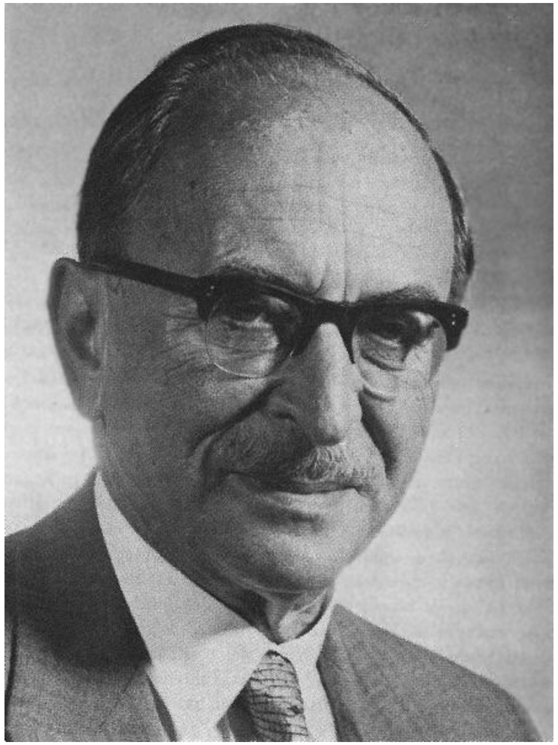
Emmanuel Candès, California Institute of Technology

Multiscale Geometric Analysis in High Dimensions: Workshop # 4
IPAM, UCLA, November 2004

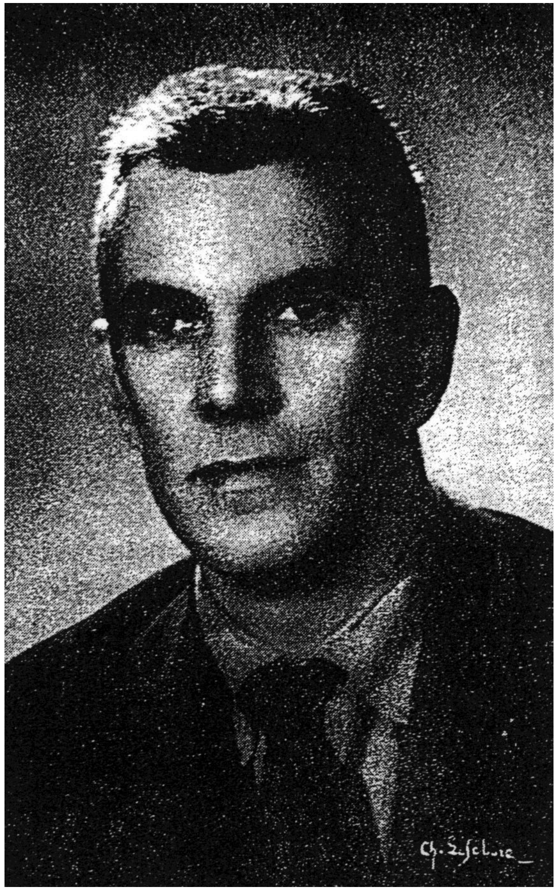
Collaborators: Hannes Helgason (Caltech), Philip Charlton (LIGO (Caltech) & LaTrobe University, Australia)

Agenda

Time-Frequency Analysis: Origin



Gabor



Ville

Wigner

Connected with the development of Quantum Mechanics (Weyl-Heisenberg, Von Neumann)

Time-Frequency Analysis: Philosophy

Problem of developing a mixed representation of a signal in terms of a double sequence of elementary signals, each of which occupies a certain domain in the time-frequency plane

"One is interested, in communication theory, in representing an oscillating signal as a superposition of elementary wavelets, each of which has a rather well defined frequency and position in time. Indeed, useful information is often conveyed by both the emitted frequencies and the signal's temporal structure (music is a typical example). The representation of a signal as a function of time provides a poor indication of the spectrum of frequencies in play, while, on the other hand, its Fourier analysis masks the point of emission and the duration of each signal's elements. An appropriate representation ought to combine the advantages of these two complementary descriptions; at the same time, it should be discrete so that it is better adapted to communication theory." (Roger Balian)

Time-Frequency Plane

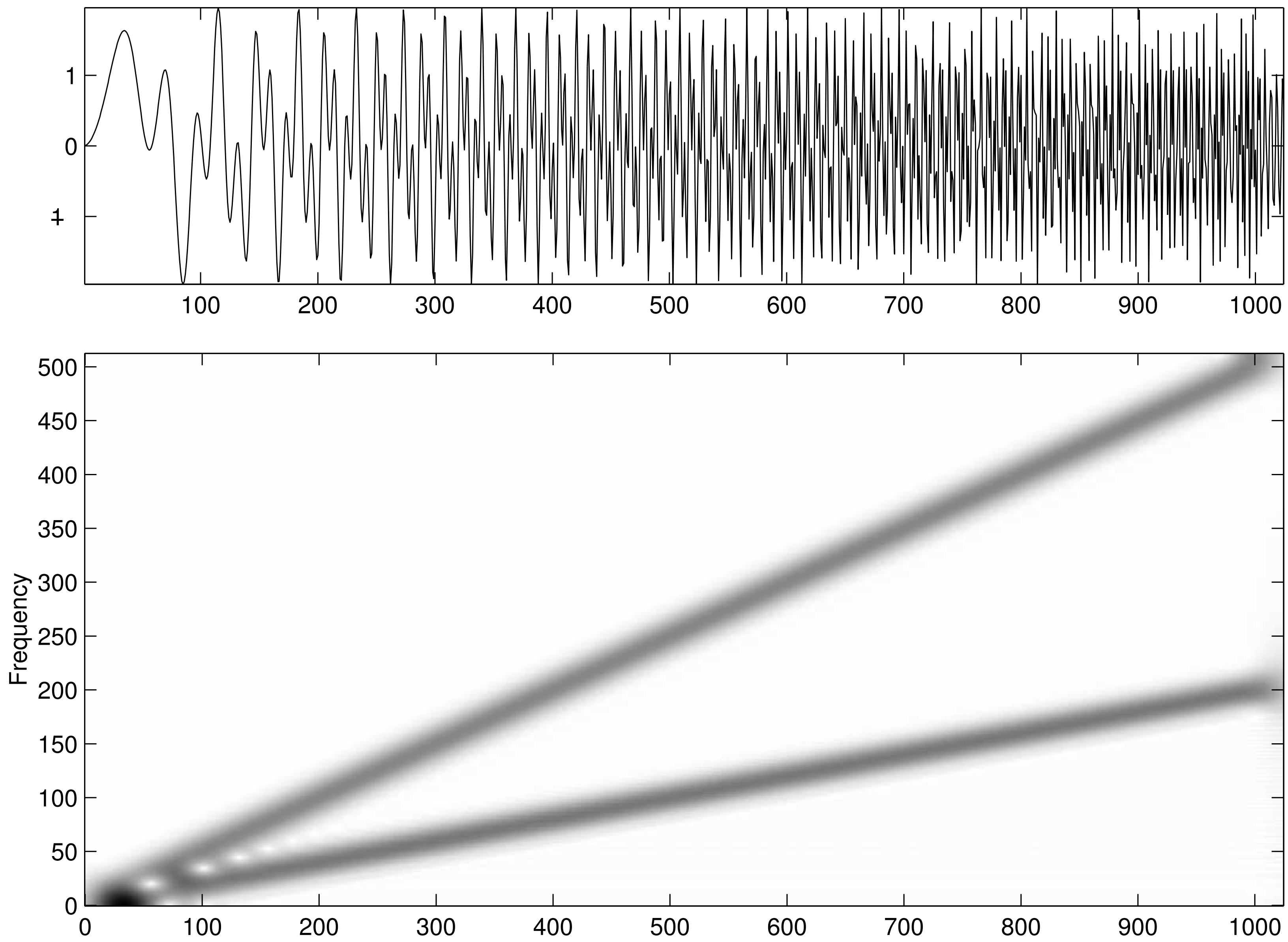
- Many time-frequency representations
- Short-time Fourier transform: window w

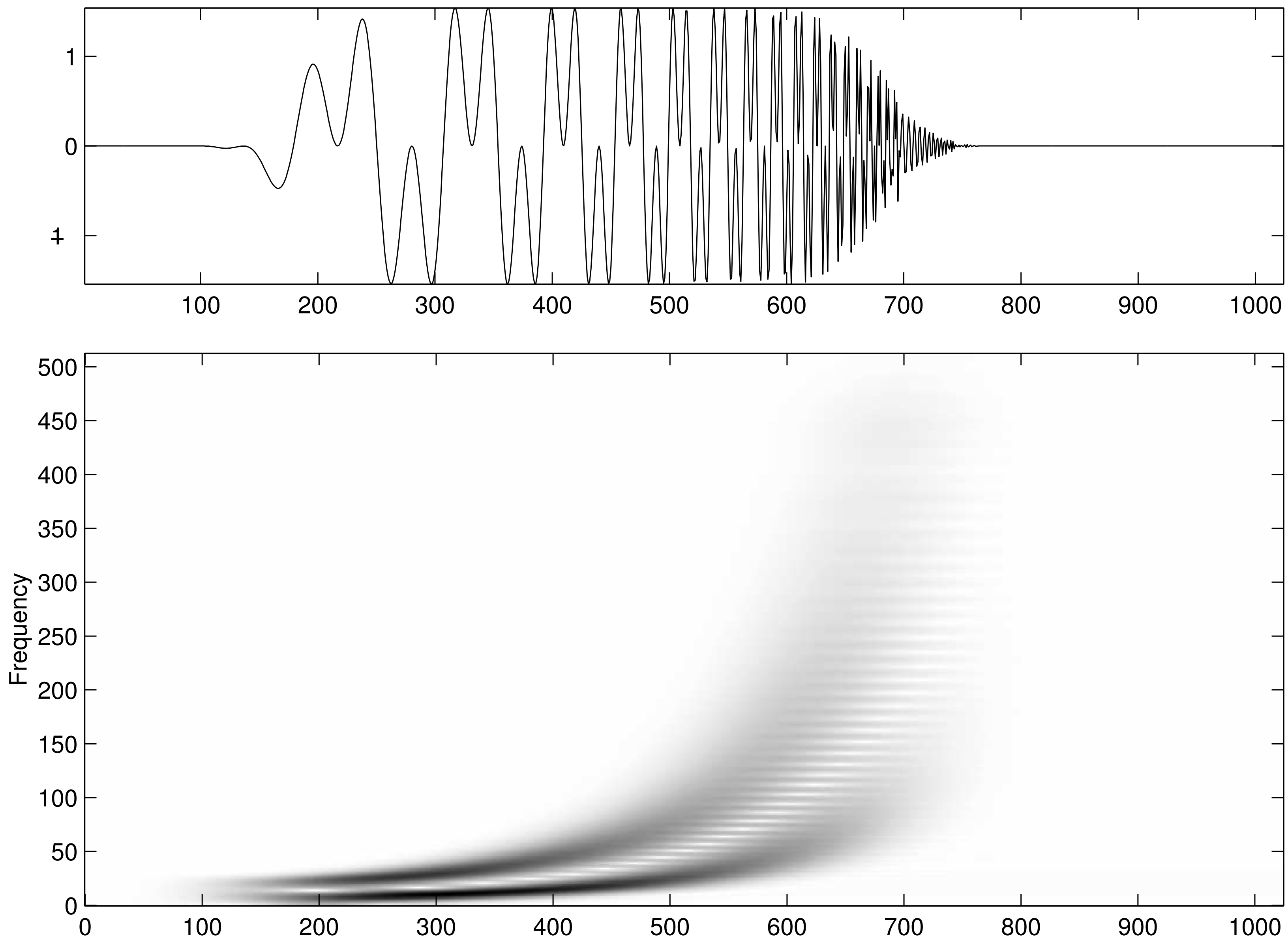
$$S(\tau, \omega) = \int f(t) w(t-\tau) e^{-\mathrm{i}\omega t}\, dt$$

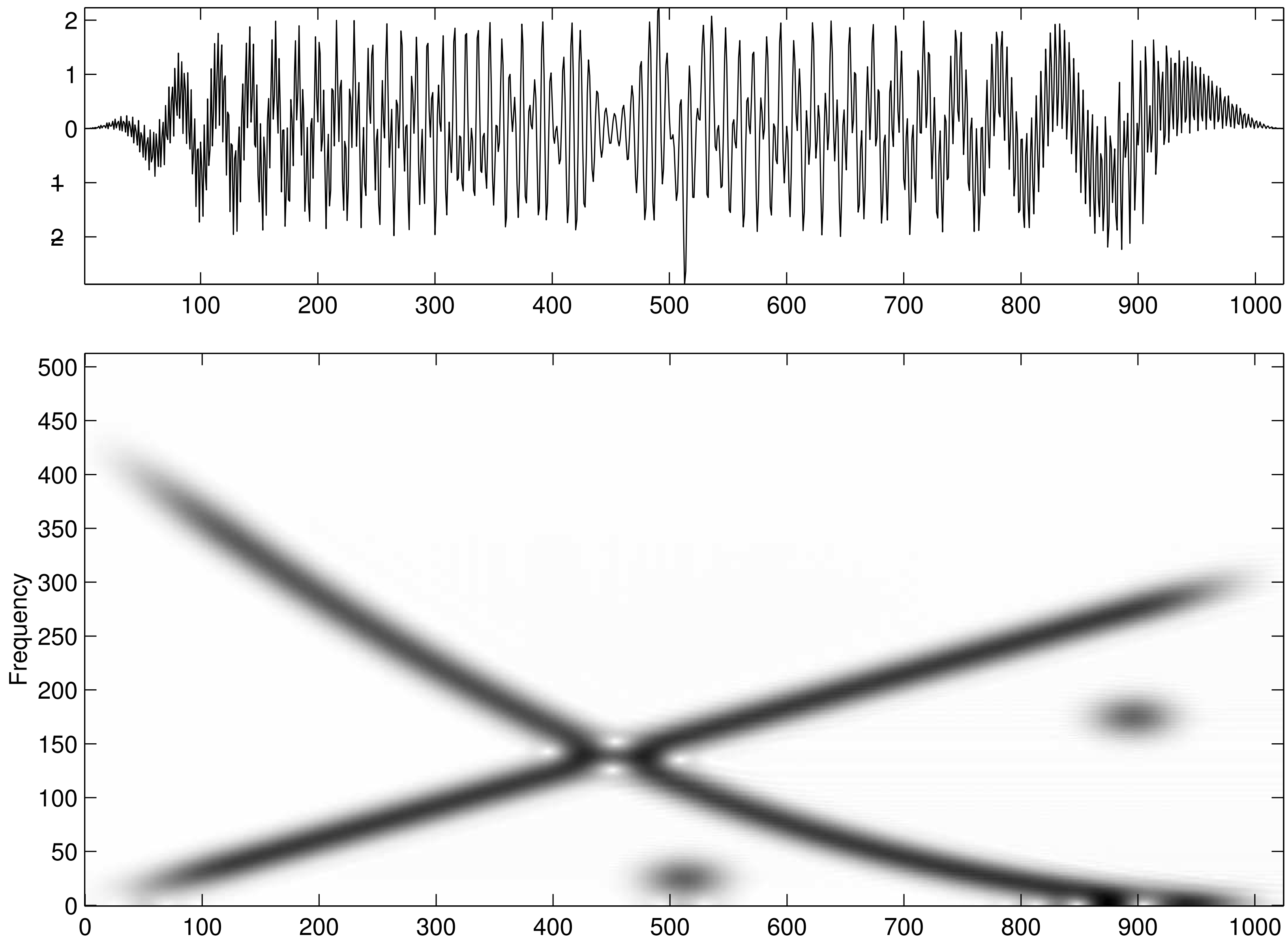
- Unfold the signal in the time-frequency plane (τ, ω) which leads to a mixed representation in time-frequency atoms (Ville)
- Synthesis with time-frequency atoms of the form (Gabor)

$$w(t) = h^{-1/2} e^{\mathrm{i}\omega t} g\left(\frac{t-\tau}{h}\right)$$

h fixed


1
0
-1
100
200
300
400
500
600
700
800
900
1000
Frequency
500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800
900
1000


1
0
-1
100
200
300
400
500
600
700
800
900
1000
Frequency
500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800
900
1000



2
1
0
-1
-2
100
200
300
400
500
600
700
800
900
1000
Frequency
500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800
900
1000

Development

- Theory
- Tools
- General applications

Very active fields over the last 20 years

- Journals
- Conferences

Tools

- Linear and quadratic time-frequency representations, which provide time-frequency phase portraits of a signal, e.g.
  - Short-time Fourier transform (STFT)
  - Wigner Ville Distribution (WVD)
- New orthonormal bases and frames which can efficiently represent certain kinds of time-frequency phenomena, e.g.
  - Wilson Bases
  - Cosine Packets
  - Gabor Frames

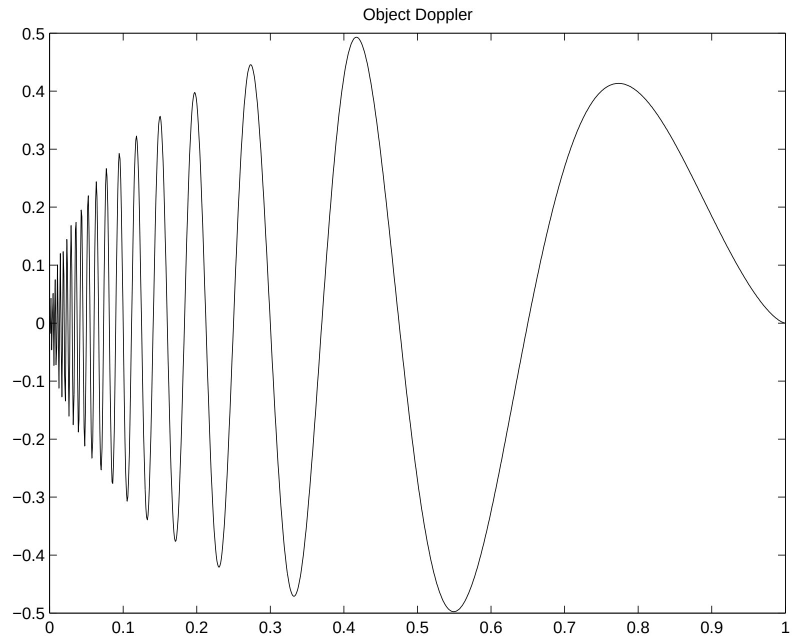
Applications

- Compression of audio signals
- Restoration of old audio recordings, e.g. Caruso (R. Coifman)
- Synthesis of a purely numerical (superhuman) voice (X. Rodet)

Chirps

$$f(t) = A(t)\cos(N\varphi(t))$$

- Amplitude A (smooth)
- Phase φ (smooth)
- Oscillation degree λ is large



Chirps arise in a number of important scientific disciplines

- Echolocation in bats and other mammals
- Gravitational waves
- Doppler effect, etc.

Gravitational Waves I

- GWs are predicted by the theory of relativity
- GWs have never been observed experimentally
- Current scientific projects of a very large scale aim at detecting GWs
  - LIGO
  - VIRGO

Gravitational Waves II

- A promising source of gravitational waves: coalescing binaries
- First order of approximation

$$f(t) = A(t_0 - t)_+^{-1/4} \cos(B(t_0 - t)^{5/8})$$

- Instantaneous frequency $\omega(t) \sim t^{-3/8}$ increases with time like a power-law
- Signal is expected to be burried in a “sea” of noise

The Problem of Chirps

Challenging problems

- Efficient representations of chirps
- Recovery of chirps from noisy data
- Detection of chirps from noisy data

The Detection Problem

$$y_t = \alpha\, S_t + z_t, \quad t = 0, 1, \ldots, N-1$$

- y: data
- S: unknown chirping signal.
- $\alpha > 0$ is the "signal strength"
- z is white noise, z_t $i.i.d.$ $N(0,1)$

To test

$$H_0: \; \alpha = 0 \quad vs \quad H_1: \; \alpha \neq 0$$

Models for Chirping Signals

$$f(t) = A(t)\cos(N\varphi(t)), \quad 0 \le t \le 1$$

Smoothness assumption on phase and amplitude (+ identifiability condition on the phase φ):

$$\textsc{Chirp}(s; R) = \{f, \quad \|A\|_{C^s} \le R, \ \|\varphi\|_{C^s} \le R\}$$

Hölder class C^s: $m = \lfloor s \rfloor$

$$|D^m\varphi(t) - D^m\varphi(t')| \le C \cdot |t - t'|^{s-m}$$

III. Detection Strategies: Background

Matched Filters

- Assume unknown signal $\boldsymbol{S}$ belongs to a parametric family $\boldsymbol{S_\theta}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- Test Statistic

$$\boldsymbol{Z^* = \max_{\theta} Z[\theta]}$$

where

$$\boldsymbol{Z[\theta] = \frac{\langle y, S_\theta \rangle}{\|S_\theta\|_2}}$$

- Compare with threshold

Notations: $\boldsymbol{\langle x, y \rangle = \frac{1}{N} \sum_{t=0}^{N-1} x_t y_t}$

Criticism

- Grid-size must be very fine
- Computational complexity is very high
- Not robust (often practically unrealistic)
- Not flexible

Ridge Detection, I

- Instantaneous frequency

$$f(t) = A(t)\, e^{\mathrm{i}\lambda\varphi(t)}$$

  Locally,

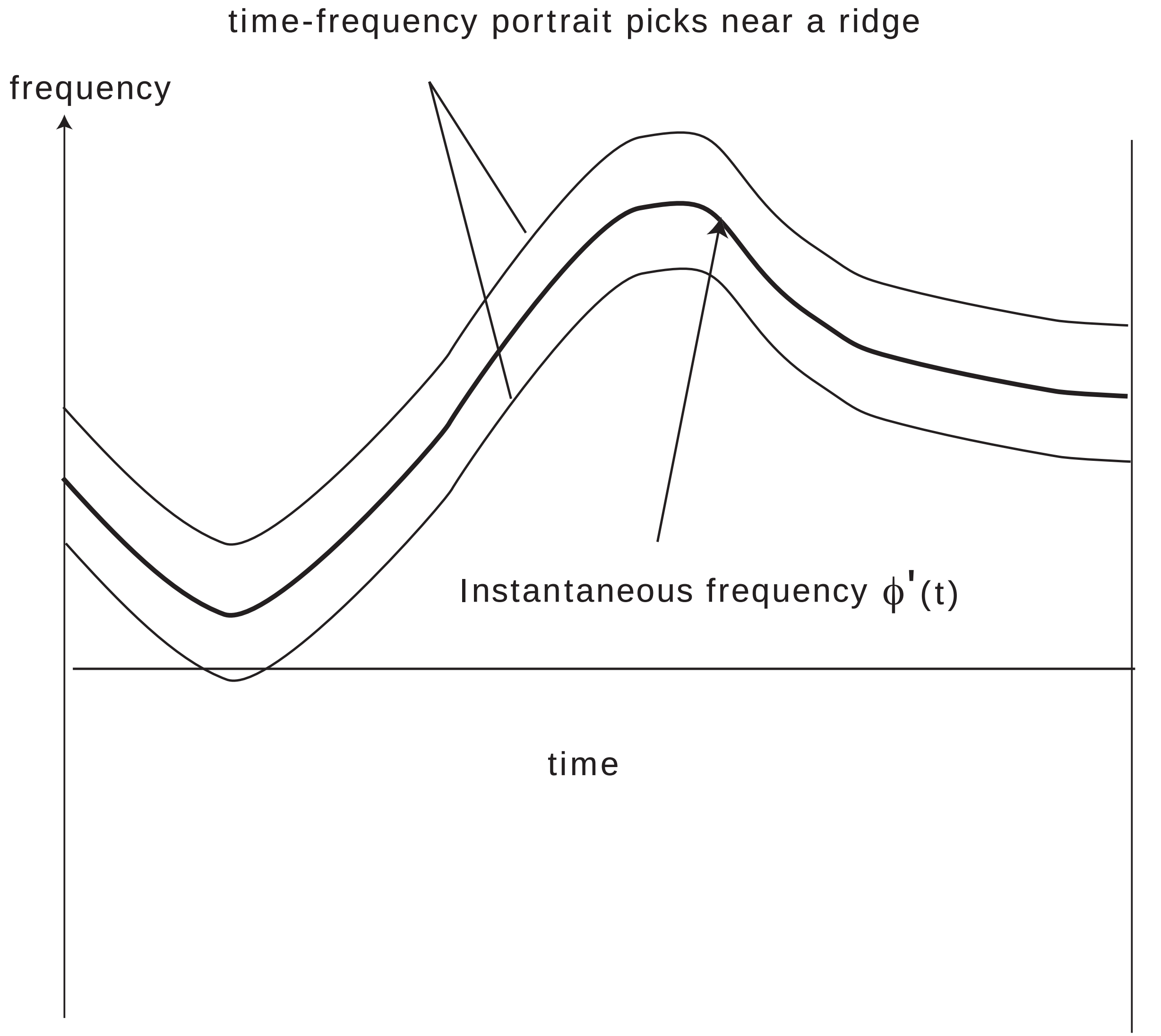
$$\varphi(t) = \varphi(t_0) + \varphi'(t_0)(t - t_0) + O(t - t_0)^2$$

  Instantaneous frequency: $\lambda\varphi'(t)$

- Short-time Fourier transform

$$F(\tau, \omega) = \int f(t) w(t - \tau) e^{-\mathrm{i}\omega t}\, dt.$$

Ridge Detection, II



Idea: detect this ridge from noisy data

Criticism

Problem: Short-time Fourier transform is highly oscillatory

Proposal: find curve such that

$$\int |F(t, \gamma(t))| \, dt \quad \text{is maximum}$$

- Not constructive
- How to choose the windowing? Resolution problem (Heisenberg)
- Ignore the thickness: loss of efficiency
- Noise does not cancel!

Our Goals

- Identify the optimal detection threshold, i.e. the signal strength below which no method of detection can be successful for large dataset size N.
- Focus on the computational complexity of a near-optimal detector, i.e. the complexity required to detect signals slightly exceeding the detection threshold.
- Design adaptive strategies, which do not require information about the smoothness of the phase and amplitude.

Higher goal: connect statistical theory and time-frequency analysis.

Connections

- Nonparametric detection: Ingster, Spokoiny.
- Beamlet analysis and detection of line segments: Arias-Castro, Donoho, Huo.

IV. Chirplets

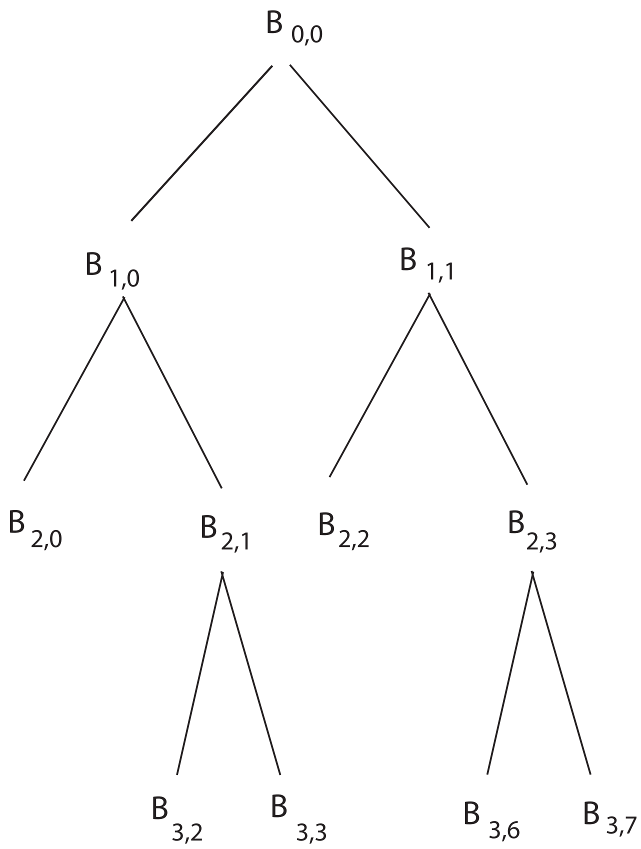
Recursive Dyadic Partitions (RDP)

Definition. An RDP is any partition obeying

- $I = [0, 1]$ is an RDP
- If $P = \{I_1, \ldots, I_i, \ldots, I_m\}$ is an RDP, then the partition obtained by splitting any interval $I_i \subset P$ is also an RDP.

Examples:

- $[0, 1/4), [1/4, 1/2), [1/2, 1)$ is an RDP
- $[0, 1/4), [1/4, 3/4), [3/4, 1)$ is not an RDP



Chirplet Dictionary

- I dyadic interval $I = [k2^{-j}, (k+1)2^{-j})$
- On each interval, discrete set of offsets and slopes: $a_{I,\mu}$, $b_{I,\mu}$
- Chirplets at scale 2^{-j} and location $t_I = k2^{-j}$

$$f_{I,\mu}(t) = |I|^{-1/2}\, e^{\mathrm{i}(a_{I,\mu}t^2/2 + b_{I,\mu}t)} 1_I(t), \qquad \|f_{I,\mu}\|_{L_2} = 1$$

- Instantaneous frequency is linear: $a_{I,\mu}t + b_{I,\mu}$
- Discretization of instantaneous frequency is scale-dependent

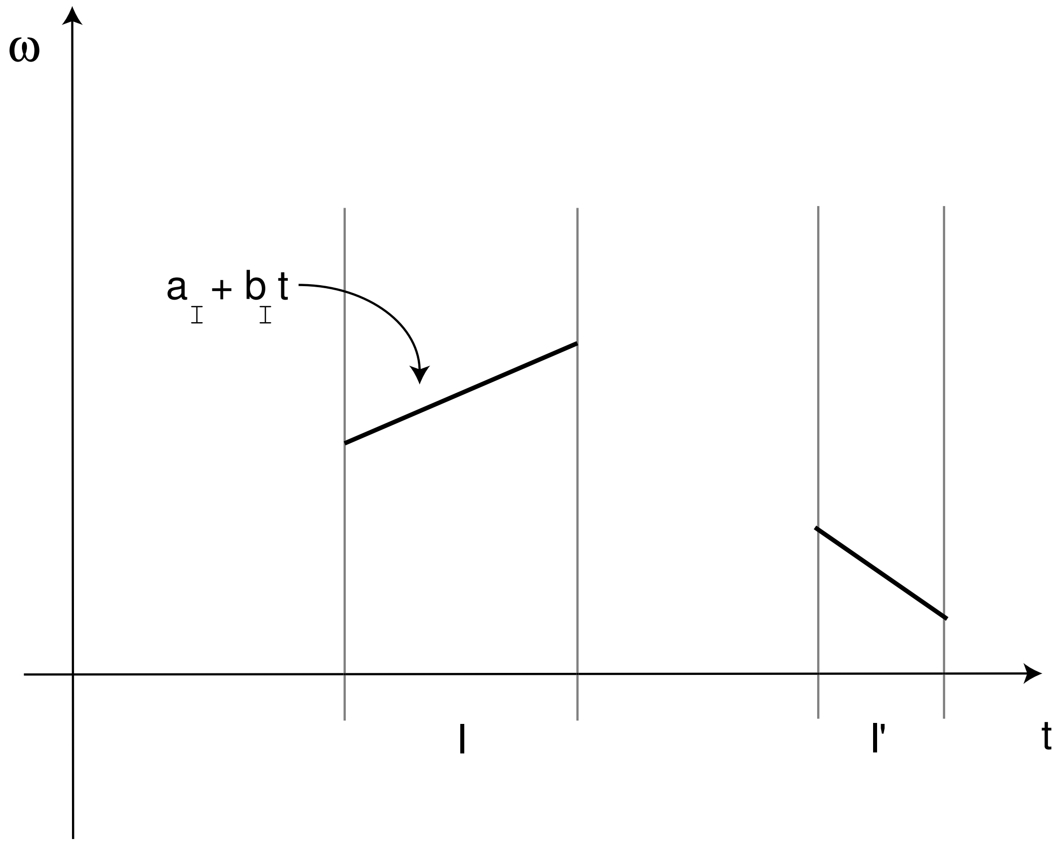


Figure 1: Schematic representation of two chirplets

Chirplet Graph

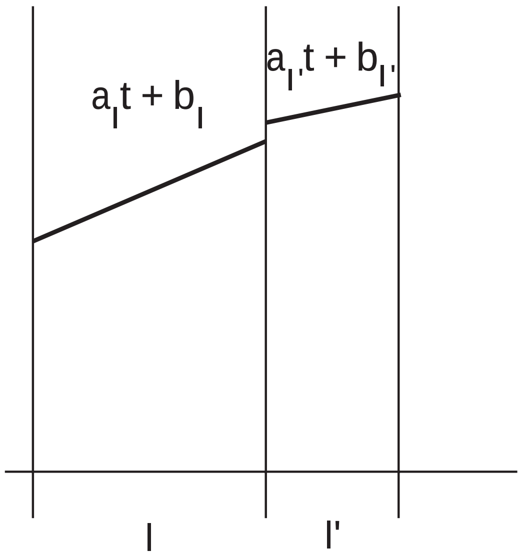
Chirplet graph

- Vertices: chirplets
- Edges: connectivities

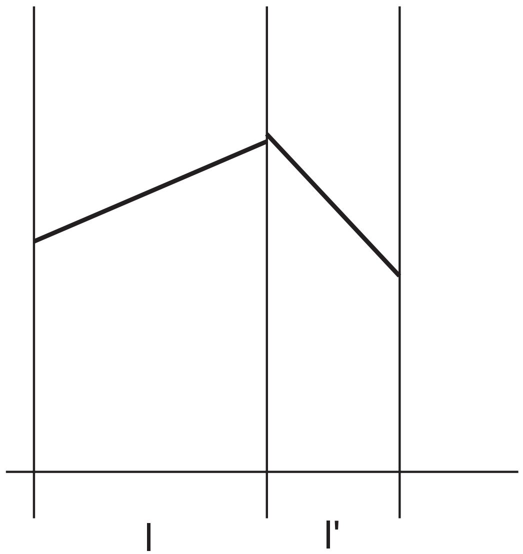
Connectivities

- 'Live' on a pair of adjacent intervals
- 'Continuity' of the instantaneous frequency

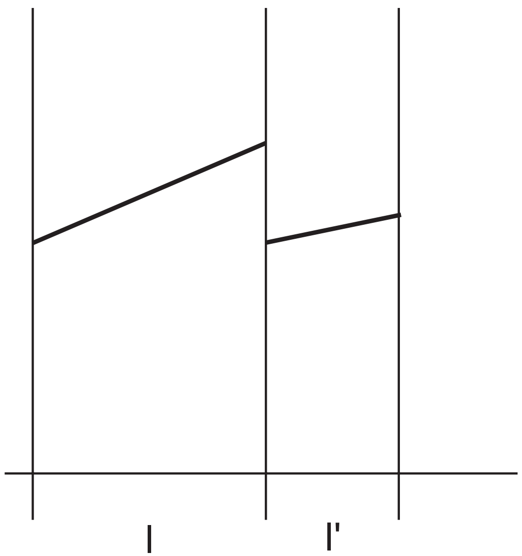
Many vertices and few connectivities/vertex, i.e. $O(\log N)$, N sample size.



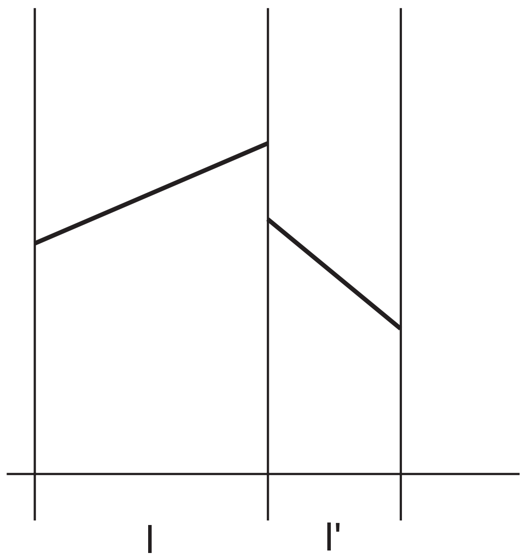
Admissible Continuation



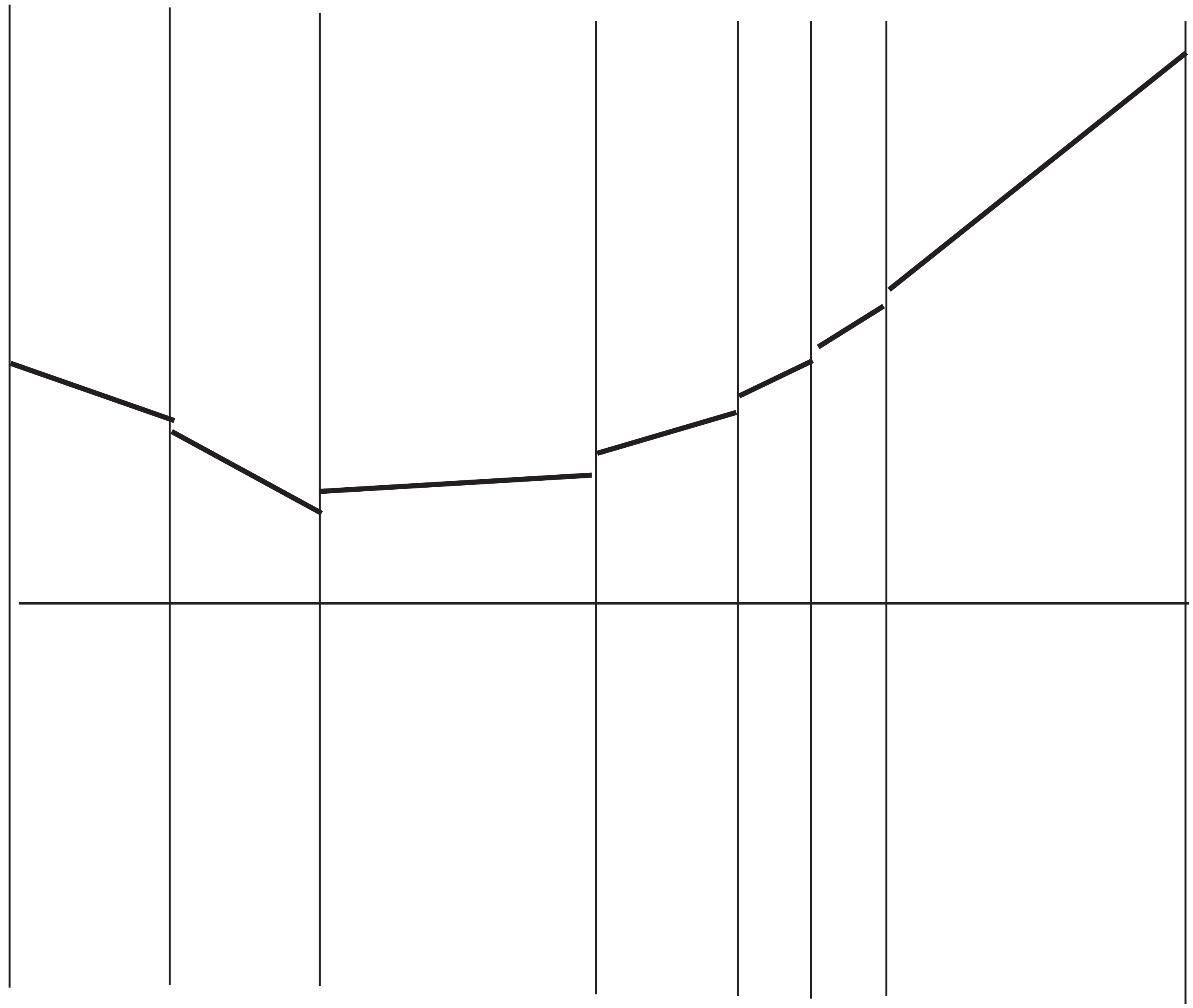
Inadmissible Continuation



Inadmissible Continuation



Inadmissible Continuation



Example of a path in the chirplet graph

Multiscale Chirplet Detection

Multiscale Detection Strategy

$$T^* = \sup_{\text{all paths}} \frac{\sum_{v \in V} |\langle y, f_v \rangle|^2}{\ell(V)}$$

If exceeds threshold A^*, reject H_0.

- $C(V) = -\sum_{v \in V} |\langle y, f_v \rangle|^2$: cost of a path
- $\ell(V)$: length of a path

Where Does This Come From?

Behavior of T^* under H_0; $y = z$, z white noise.

$$z^*(\ell) = \max_{(\#V)=\ell} \sum_I |\langle y, f_v \rangle|^2.$$

1. Size of $y^*(\ell)$: $E[y^*(\ell)] = \gamma \cdot \ell(1 + o(1))$, $\gamma > 1$.

2. Tools
   - Large deviations for chi-squares (upper bound)
   - Talagrand's majorizing measures (lower bound)

3. Fluctuations are negligible compared to the size of the expectation, $SD[y^*(\ell)] = O(\ell^{1/2})$.

Tools I: Talagrand's Majorizing Measures

- Z_f zero-mean Gaussian process, e.g.

$$Z_f = \langle z, f \rangle$$

- Distance

$$d(f,g) = \left[E(Z_f - Z_g)^2\right]^{1/2} = n \cdot \|f - g\|_2$$

- Entropy

$$H(\delta, \mathcal{F}) = \log_2 N(\delta, \mathcal{F}), \quad N(\delta, \mathcal{F}) = \min_i d(f, f_i) \le \delta$$

Then $Z^* = \sup_{f \in \mathcal{F}} Z_f$ obeys

$$K_1 \ \sup_\delta [\delta \sqrt{H(\delta, \mathcal{F})}] \le EZ^* \le K_2 \int_0^\infty \sqrt{H(\delta, \mathcal{F})}\, d\delta.$$

Tools II: Moderate deviations for χ^2's

$$P(Y_d - d \geq \lambda \sigma_d) \leq \lambda^{-1} \cdot e^{-\lambda^2/2}, \quad Y_d \sim \chi^2_d.$$

Statistical Theory: Detectability Threshold

$$y_t = \alpha\, S(t/N) + z_t, \quad S(t) = A(t)\cos(N\varphi(t))$$

$S \in$ CHIRP$(s; R)$; S unknown but s and R known.

Assume signal strength $\alpha \le t_0(s, R) \cdot N^{-1/2+1/2s}$, then for all tests

$$P_{H_0}(\text{reject } H_0) + \sup_{H_1} P_{H_1}(\text{accept } H_0) \to 1, \quad N \to \infty;$$

i.e. $\alpha_N^* \sim N^{-1/2+1/2s}$ is the level of detectability.

Statistical Theory: Near-Optimal Detection

H_1: composite alternative. $S \in \textsc{Chirp}(s; R)$

- regularity s is unknown
- modulus of smoothness is unknown

Reject when $T^* = \max \sum_{v \in V} |\langle y, f_v \rangle|^2 / \ell(V) \geq A^*$.

Suppose that $S \in \textsc{Chirp}(s; R)$ and $\alpha > t_1(s, R) \cdot N^{-1/2+1/2s}$. For this test

$$P_{H_1}(\text{reject } H_0) \to 1, \quad n \to \infty,$$

and

$$P_{H_0}(\text{accept } H_0) \to 1 \quad n \to \infty.$$

Near-optimality and adaptivity

Important Message

- On each dyadic interval, signal is not detectable; $|\langle \boldsymbol{y}, \boldsymbol{f}_v \rangle|^2$ is statiscally nonsignificant.
- Chain of connected intervals is statistically significant.

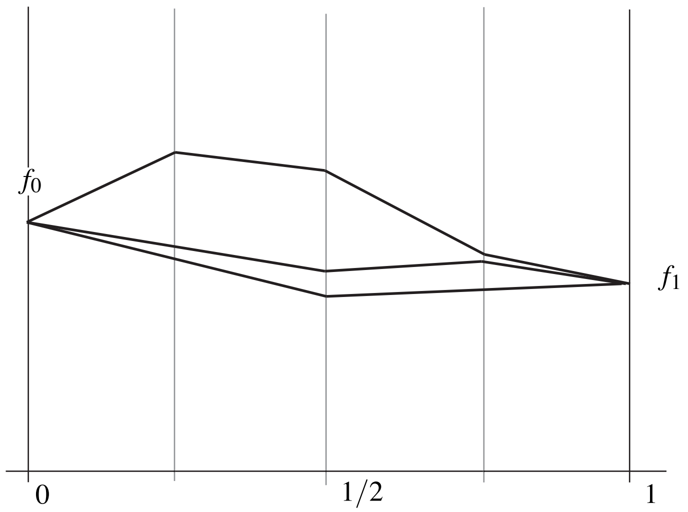
Computational Issues

- Many paths in the chirplet graph: exponential in the sample size N
- Test statistic is designed to be rapidly computable:

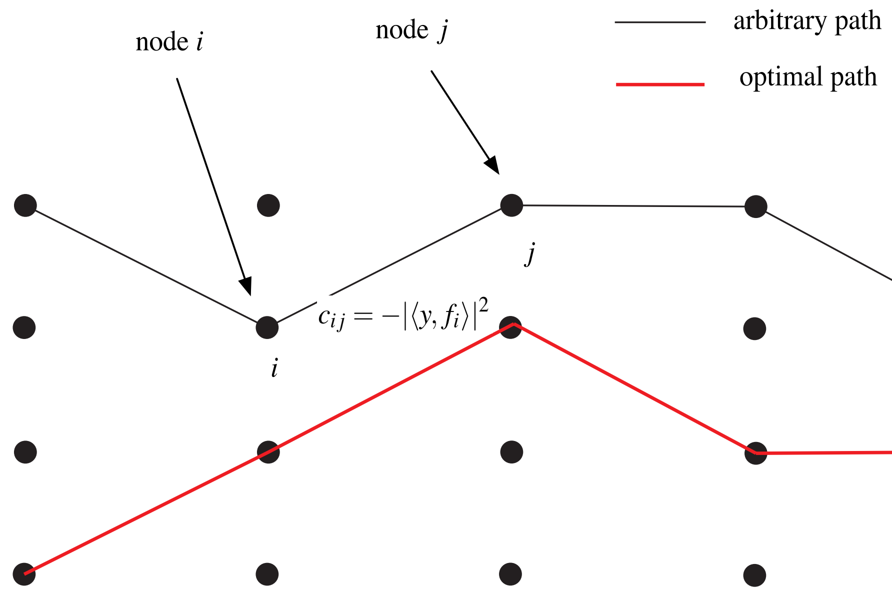
$$T^* = \min \frac{\sum_v C(v)}{\sum_v 1}$$

  - Minimum cost-to-time ratio
  - Can use dynamic programing
  - Complexity of the search is of the order of $O(P \log P)$; P, $\#$ of chirplets (more later).

Paths in the Chirplet Graph



Time-frequency graph



Chirplet graph

Shortest Path in a Graph

- Goal: find path V^* in G which minimizes the cost $\sum_{v \in V} C(v)$ (over all paths)

- Topological ordering: nodes labeled $i = 1, \ldots, P$, are said to be in *topological order* iff

$$(i, j) \in A \qquad \Leftrightarrow \qquad i < j$$

- Distance function $d(\cdot)$; $d(i)$ shortest distance from a source node s to node i.

- Shortest path for acyclic graphs: algorithm
  - *Initialize*: $d(s) = 0$, and $d(i) = \infty$ for all nodes (but the source node).
  - *Loop:* Examine nodes in topological order. For $i = 1 : P$,
    * scan the set of arcs going out from node i
    * For $(i, j) \in A(i)$, if $d(j) > d(i) + c_{ij}$, set $d(j) = d(i) + c_{ij}$ and pred$(j) = i$.
  - *Terminate:* Shortest path is that path such that $d(i)$ is minimum, i terminal node.

Decision Rule as a Shortest Path Problem

- Want to compare test statistic $T^* = \min \frac{\sum_v C(v)}{\sum_v 1}$ with a threshold α_c.
- Solve SP with modified costs

$$S^* = \min \frac{\sum_v C(v) - \alpha_c}{\sum_v 1}$$

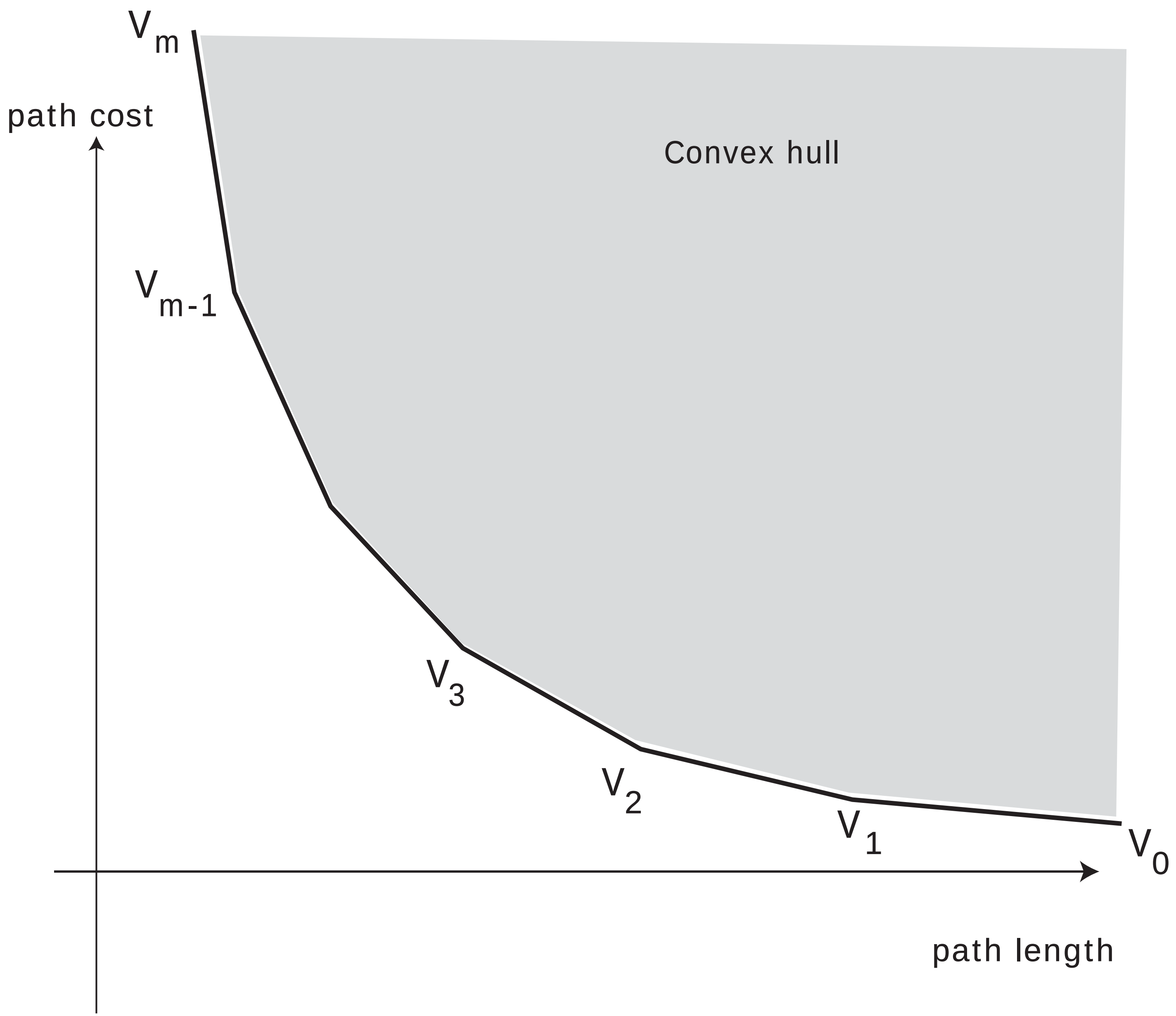
  - If $S^* < 0$, reject null
  - Otherwise, accept.
- Complexity of the search is of the order of $O(P \log P)$; P, $\#$ of chirplets.

Minimum Cost to Time Ratio (MCTTR)

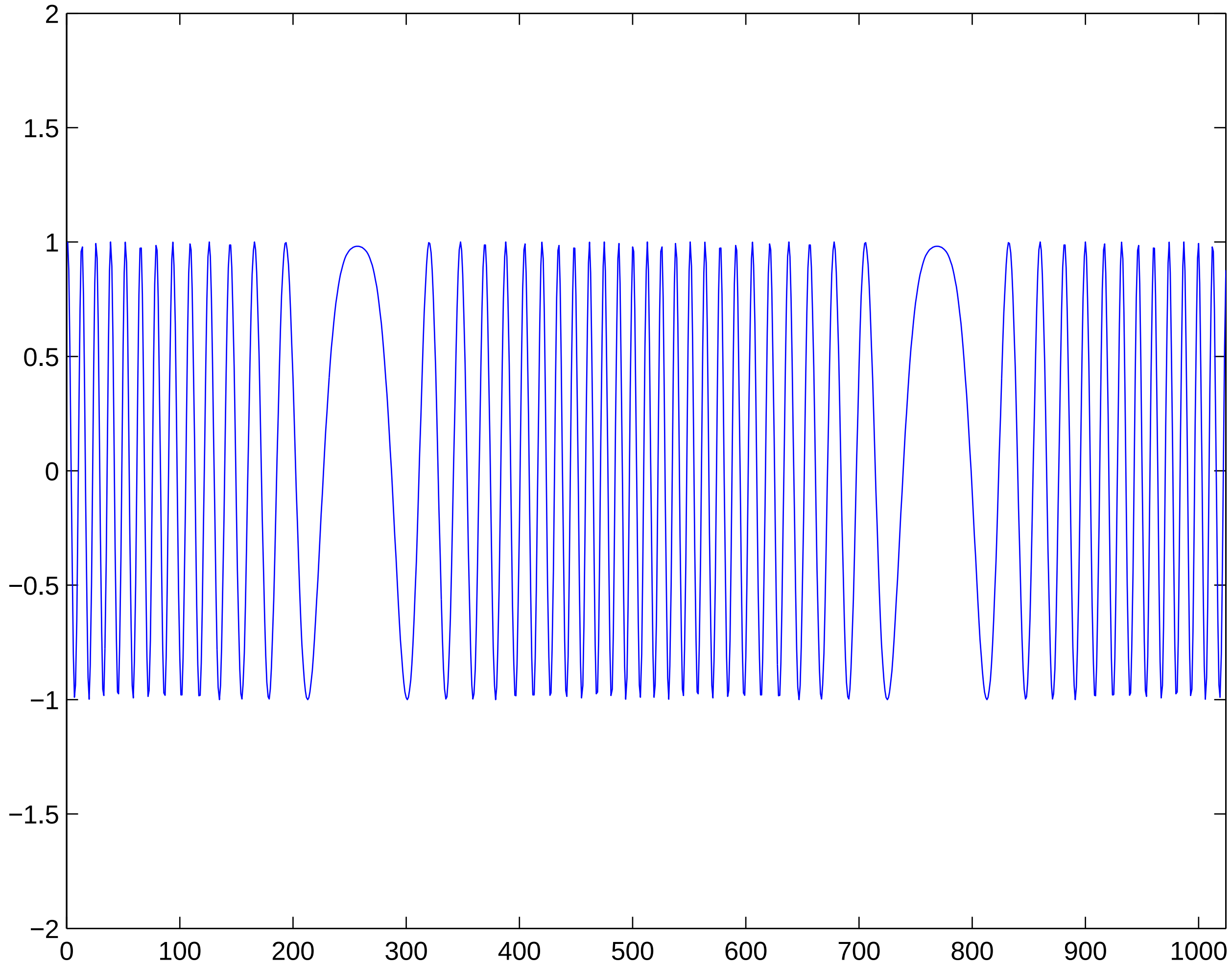
Interested in the value of the test statistic $t^* = \min \frac{\sum_v C(v)}{\sum_v 1}$ (t^* optimal value of MCTTR)

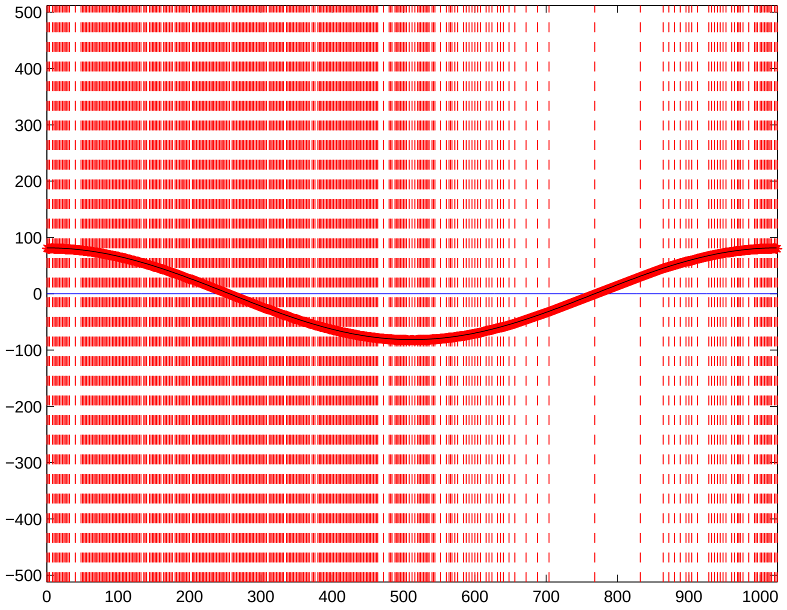
Solve a sequence of SP problems:

- Current upper bound: t_0
- Solve SP with modified cost $\ell_v = c_v - t_0$
  1. If $SP = 0$, we hold the MCTTR
  2. IF $SP < 0$, repeat with better upper-bound $\alpha_0 = (\sum_{V^*} c_v) / \sum_{V^*} 1$.
- Number of steps at most equal to the maximum length of a path.

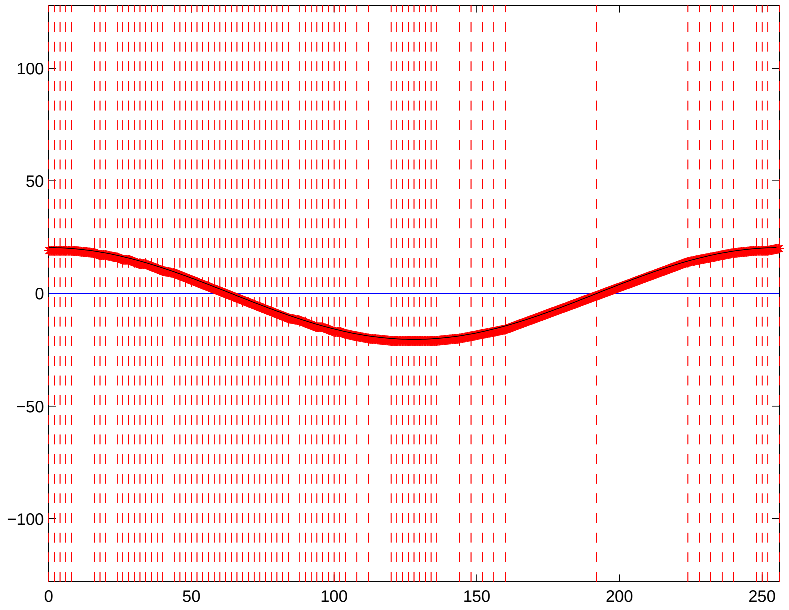

V
m
path cost
Convex hull
V
m-1
V
3
V
2
V
1
V
0
path length

Clean Chirp

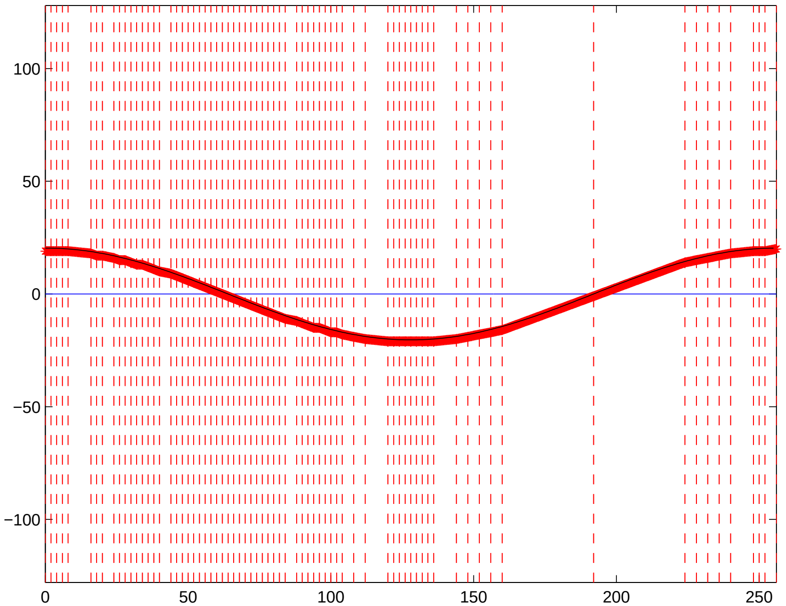




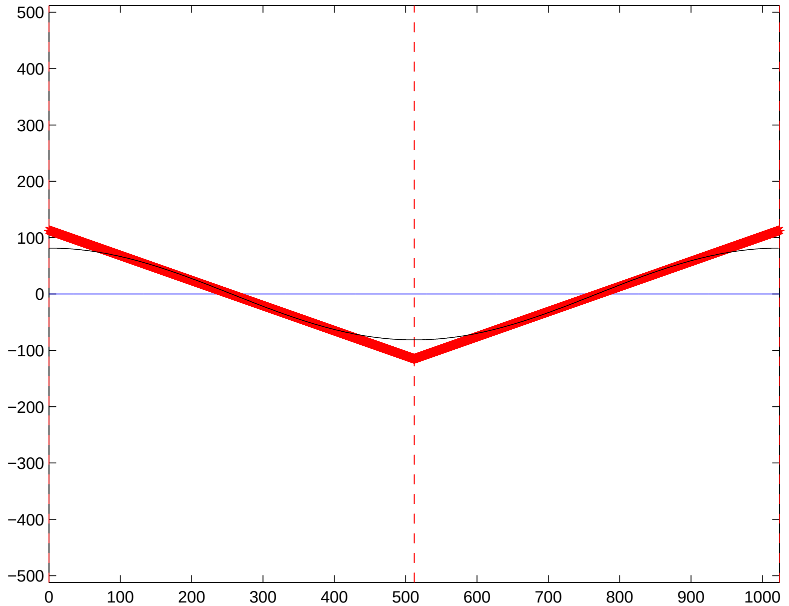
Best path, $\boldsymbol{N = 1024}$



Best path, $\boldsymbol{N = 256}$

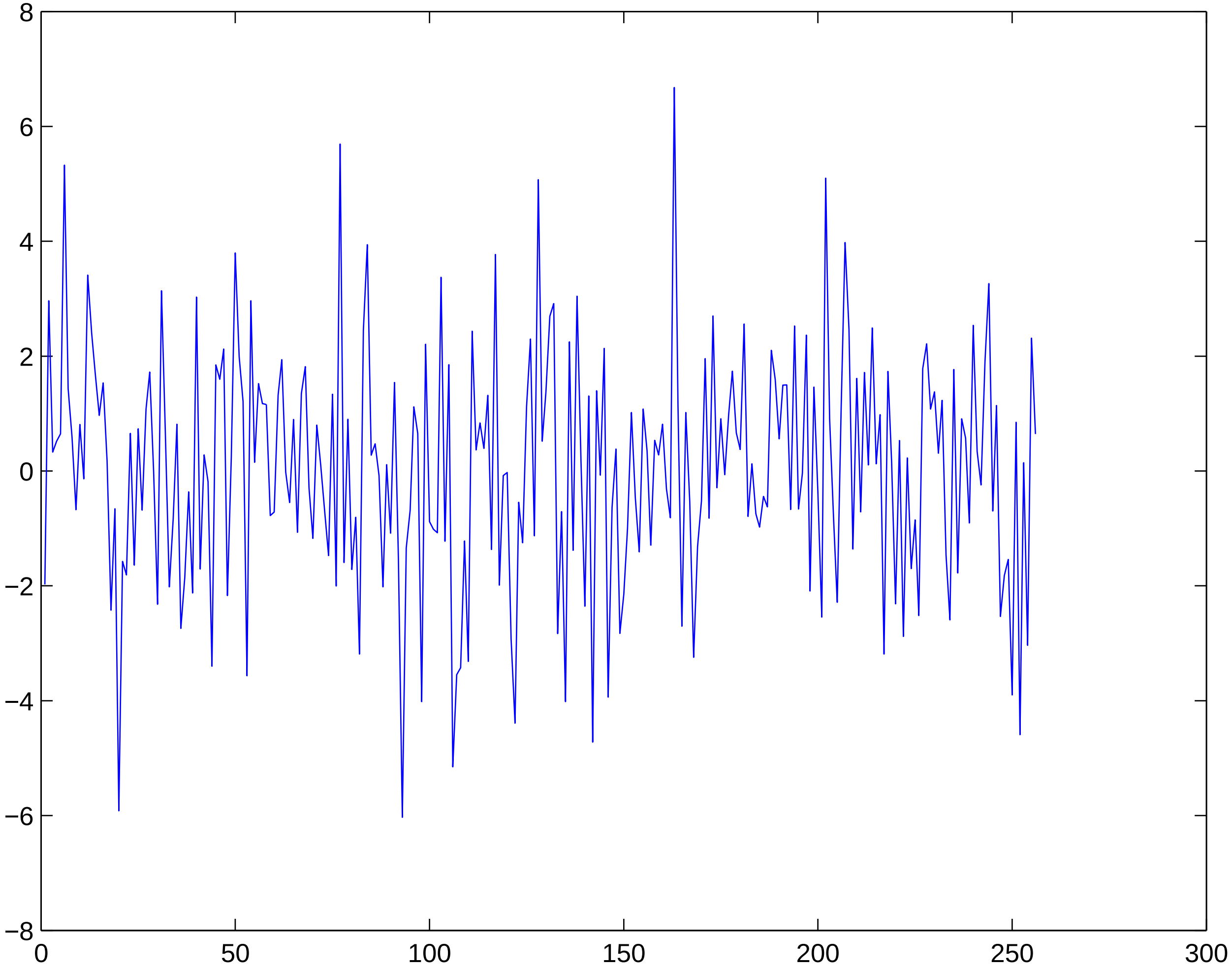


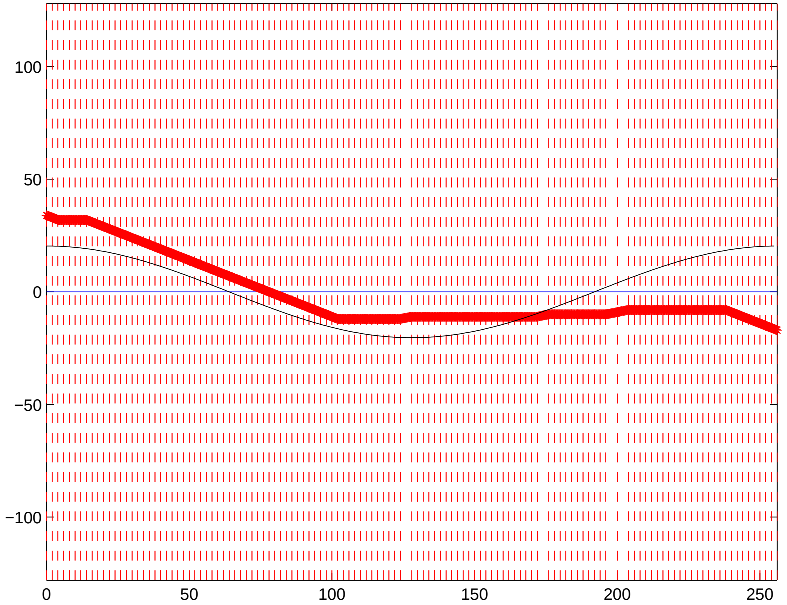
Best (unconstrained) path



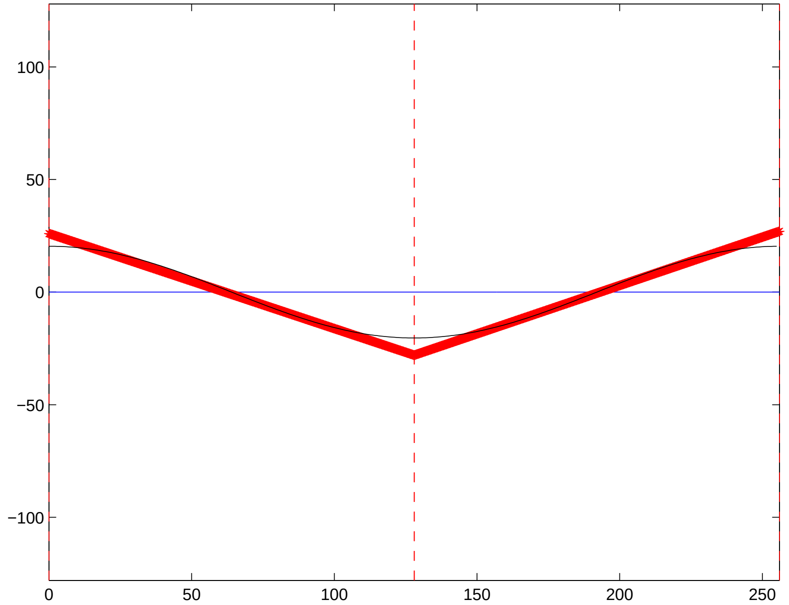
MCTTR path

Noisy Chirp?





Best (unconstrained) path

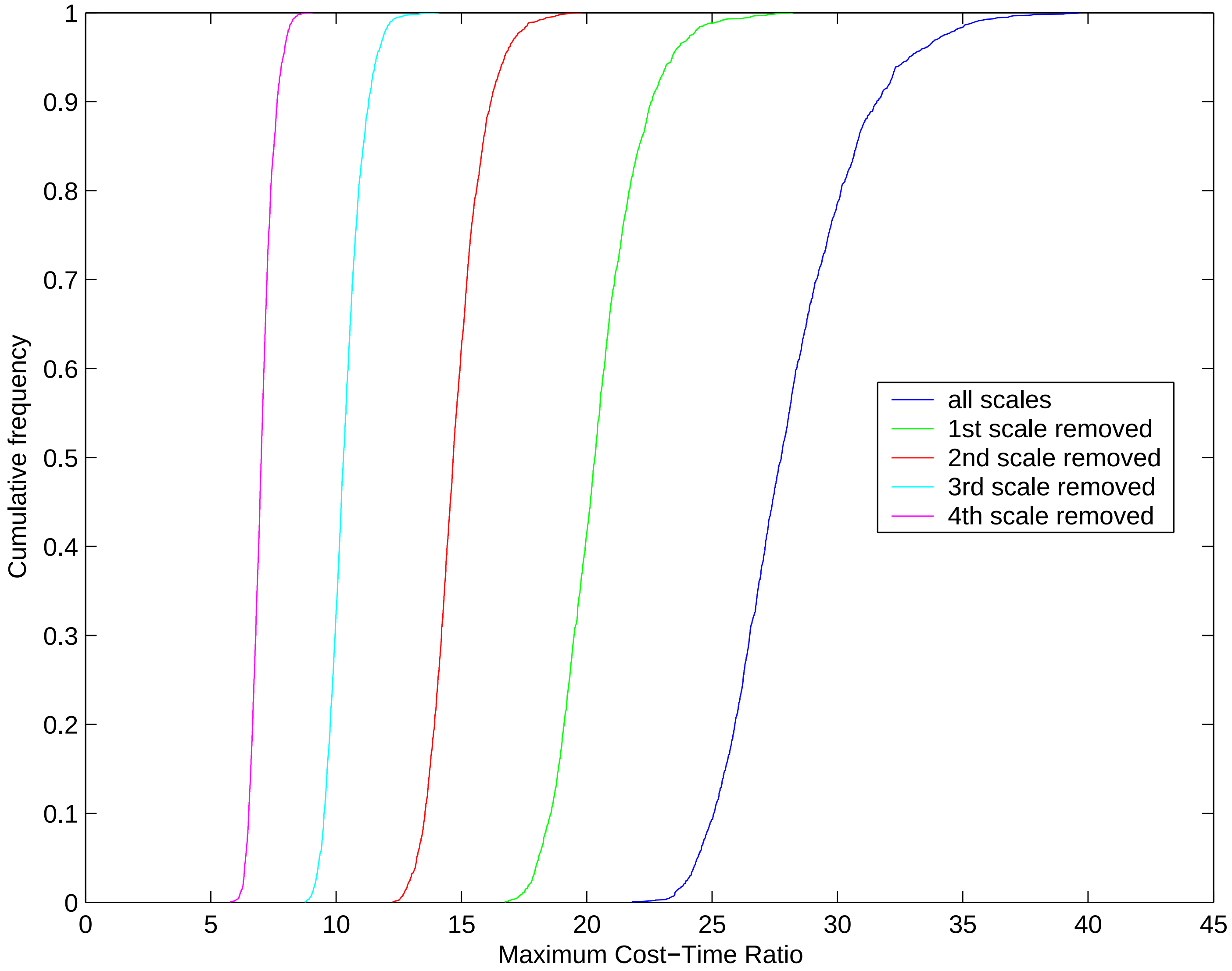


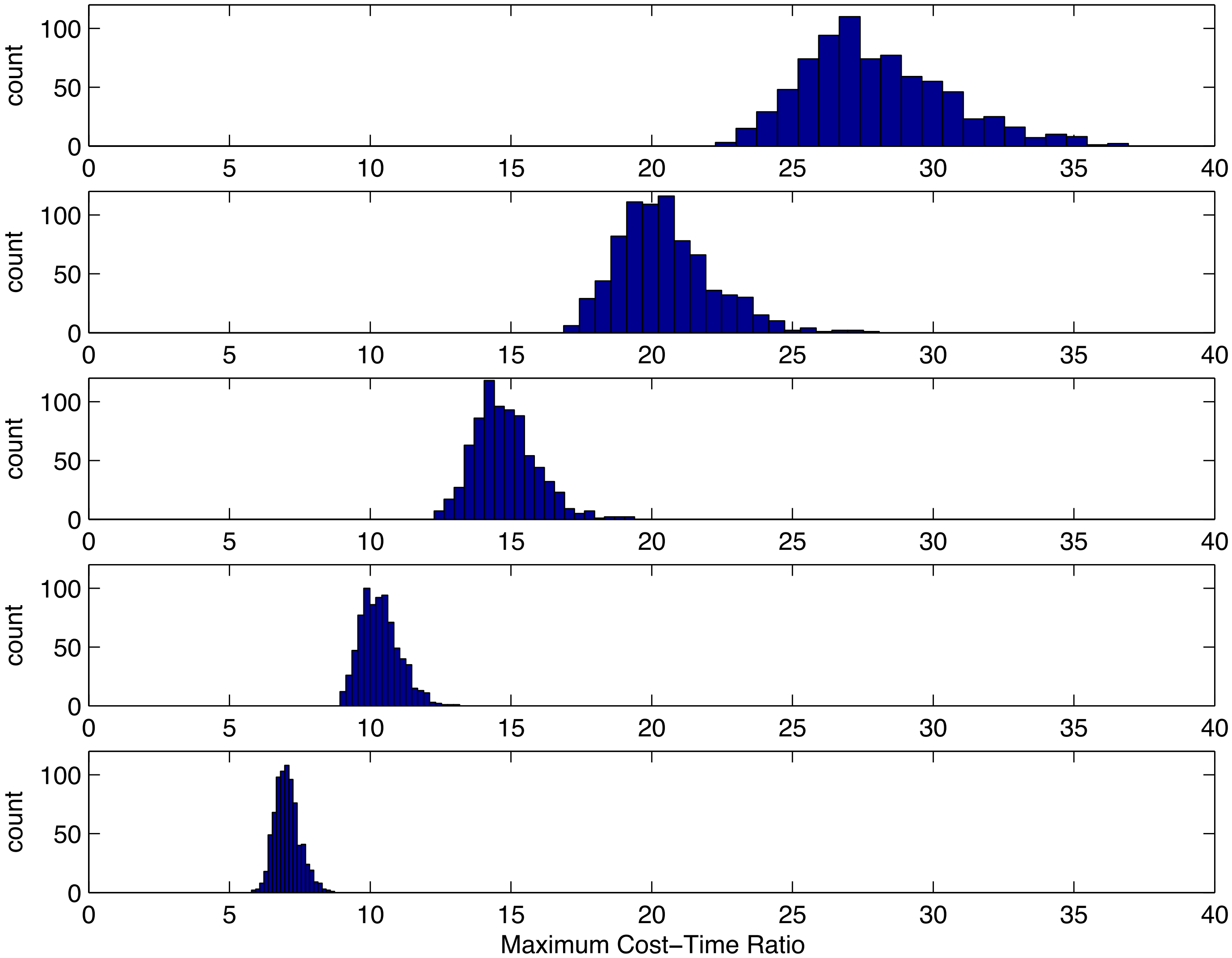
MCTTR path

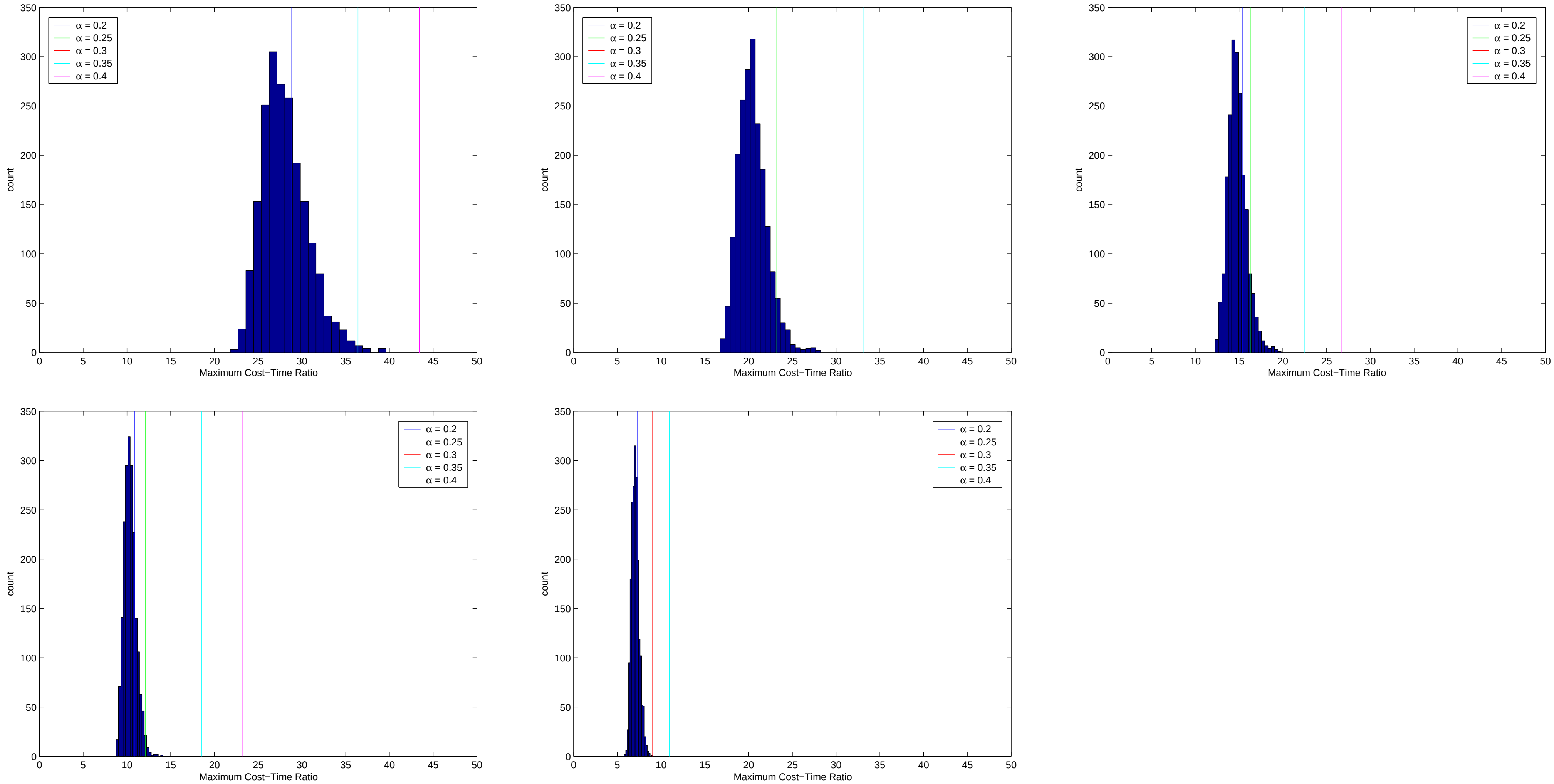
Practical Issues

- In practice, need to estimate cut-off
- Monte-Carlo simulations
- Key observation: under $\boldsymbol{H_0}$, MCTTR achieved for paths of length 1
- Refinement of detection strategy: multiple comparisons

Multiple Comparisons




count
100
50
0
0
5
10
15
20
25
30
35
40
Maximum Cost−Time Ratio


α = 0.2
α = 0.25
α = 0.3
α = 0.35
α = 0.4
count
Maximum Cost−Time Ratio

Colored Noise

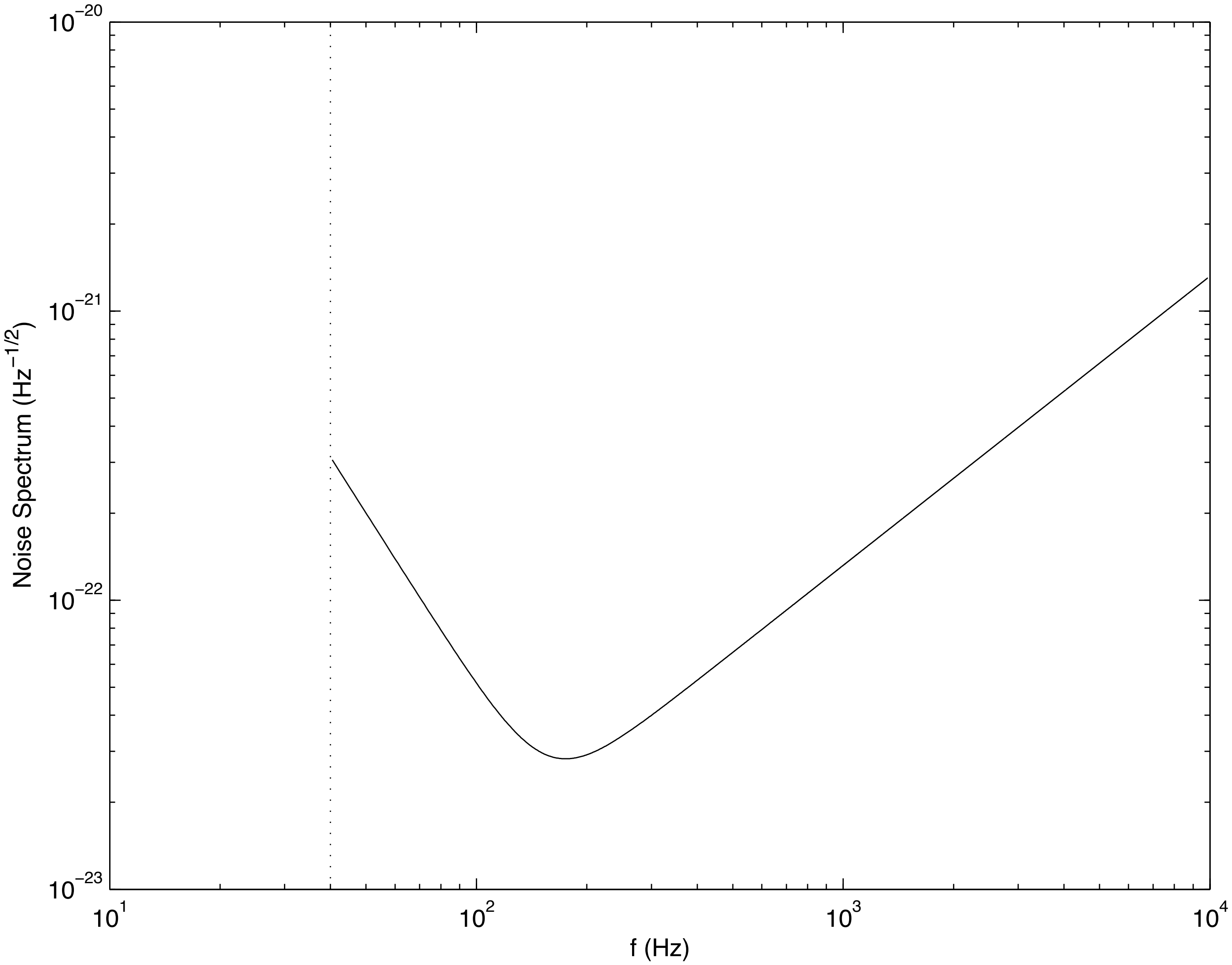


Figure 2: LIGO noise power spectrum

Likelihood Interpretation

- Model

$$Y = S + z, \qquad z \sim N(0, I)$$

- Likelihood: chirplet $f \in \mathcal{F}$ ($\|f\| = 1$)

$$L(y, \lambda f) \propto e^{-\frac{1}{2}\|y - \lambda f\|^2}$$

- Maximum likelihood

$$\max_{\lambda, f} L(y, \lambda f) = \min_{\lambda, f} \|y - \lambda f\|^2 = \min_{f} \|y - \langle y, f \rangle f\|^2$$

- Pythagoras

$$\|y\|^2 = \|y - \langle y, f \rangle f\|^2 + |\langle y, f \rangle|^2$$

- Equivalence

$$\max_{f} |\langle y, f \rangle|^2 \qquad \Leftrightarrow \qquad \max L(y, \lambda f)$$

- Chirplet path

$$C(V^*) = \max_{V} \sum_{v \in V} |\langle y, f_v \rangle|^2 \qquad \Leftrightarrow \qquad \min_{V} \|y - \sum_{v} \lambda_v f_v\|^2.$$

Detection in Colored Noise, I

- Model

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{z}, \qquad \boldsymbol{z} \sim \boldsymbol{N(0, \Sigma)}$$

- Maximum likelihood

$$\max_{\boldsymbol{\lambda, f}} \boldsymbol{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\lambda f})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\lambda f})}$$

or equivalently

$$\max_{\boldsymbol{f}} \frac{|\boldsymbol{y}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{f}|^2}{\boldsymbol{f}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{f}}$$

Detection in Colored Noise, II

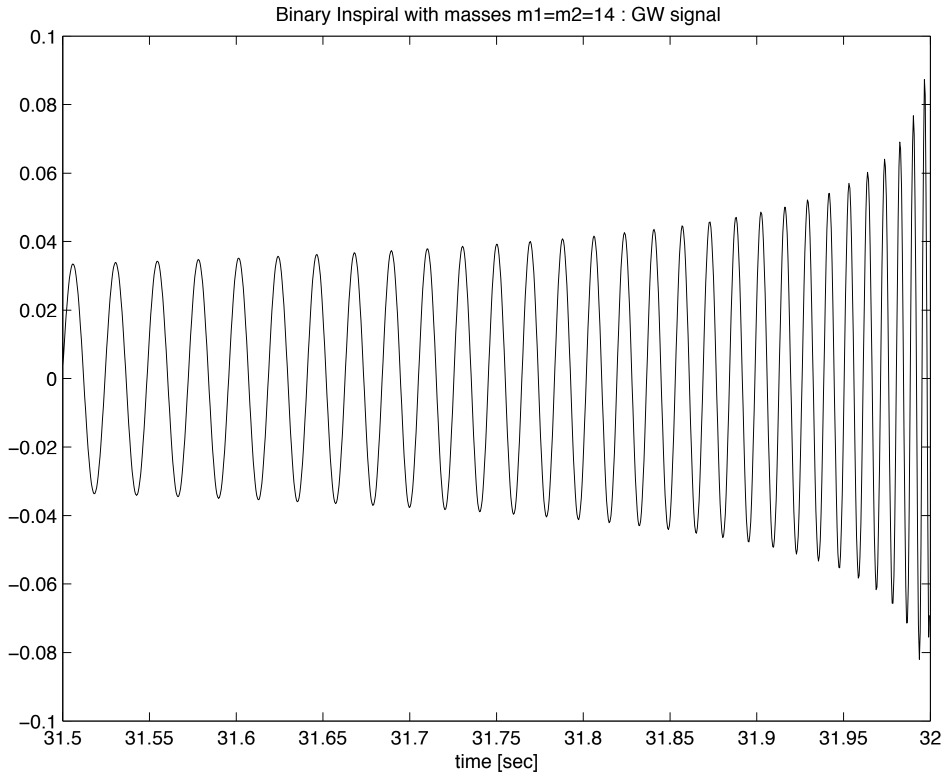
Chirplet test statistic

$$T^* = \sup_V \frac{\sum_{v \in V} C(v)}{\sum_{v \in V} 1}, \qquad C(v) = \frac{|y^T \Sigma^{-1} f_v|^2}{f_v^T \Sigma^{-1} f_v}$$

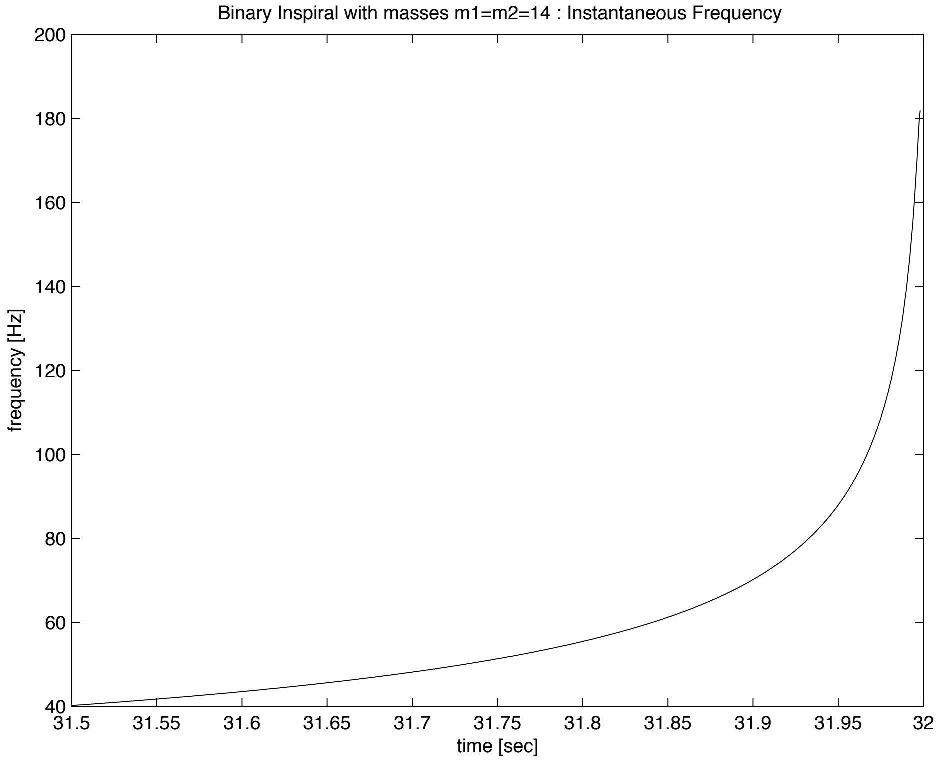
Same structure

- Interpretation as a min cost-to-time ratio
- Rapidly computable
- Similar decision theoretic results

Time-varying amplitude



Binary inspiral



Inspiral's instantaneous frequency

Time-varying amplitude

- Chirplets

$$f_v(t) = e^{i\phi_v(t)}/\sqrt{|I|}, \qquad \|f_v\| = 1.$$

- Allow for smoothly varying amplitude

$$\min \|y - P(t)\, e^{i\phi_v(t)}\|^2$$

with $P(t)$, polynomial of degree 2, say.

- Equivalence

$$C(v) = \sum_{i=0}^{2} C_i(v), \qquad C_i(v) = |\langle y, P_i(t) e^{i\phi_v(t)} \rangle|^2$$

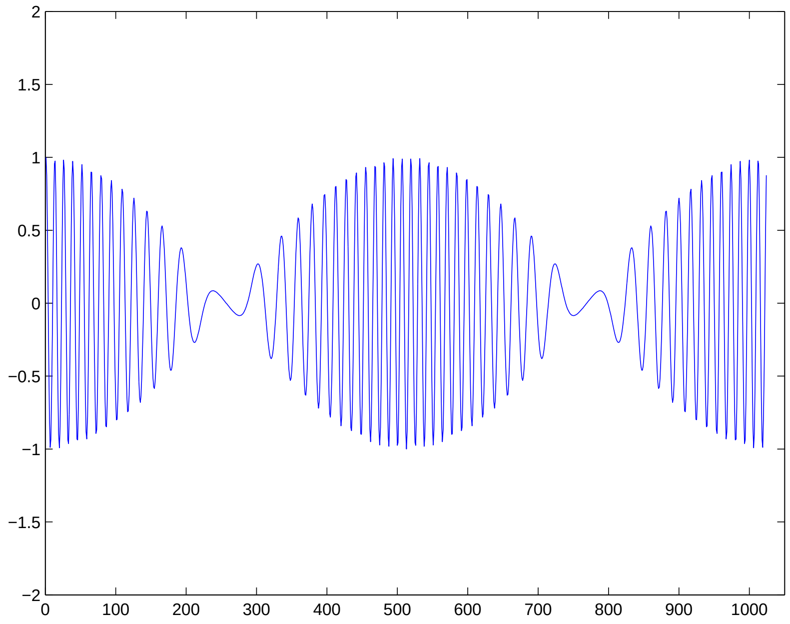
with P_0, P_1, P_2 three orthonormal polynomials of at most 2.

Chirplet test statistic

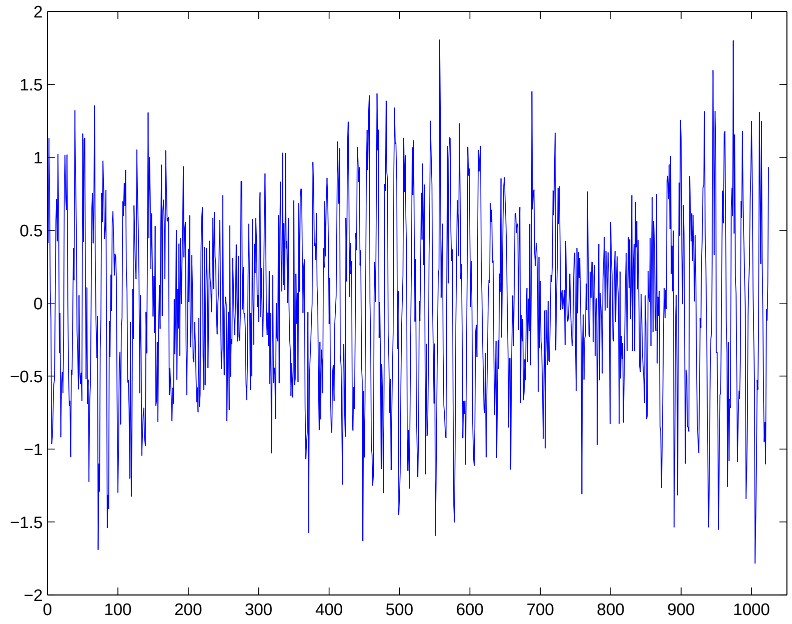
$$T^* = \sup_V \frac{\sum_{v \in V} C(v)}{\sum_{v \in V} 1}$$

Same structure

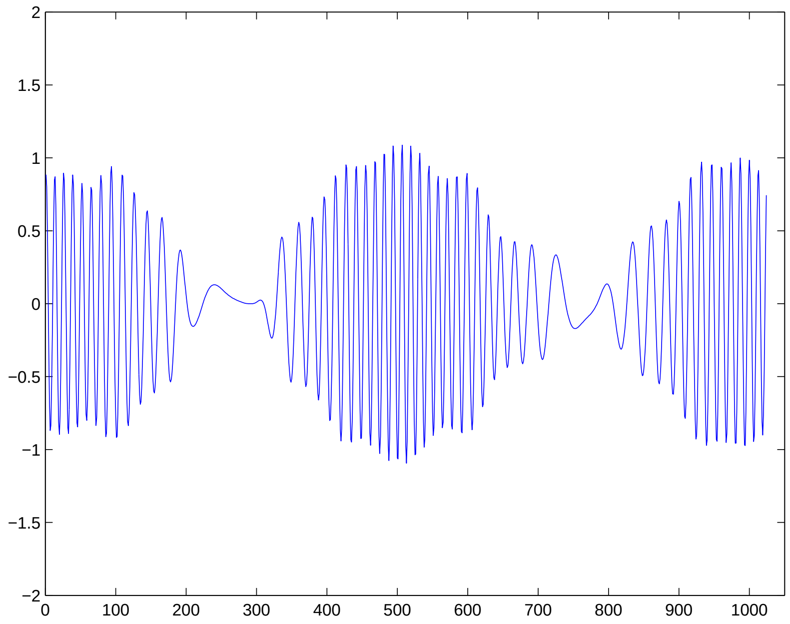
- Interpretation as a min cost-to-time ratio
- Rapidly computable
- Probably Similar decision theoretic results



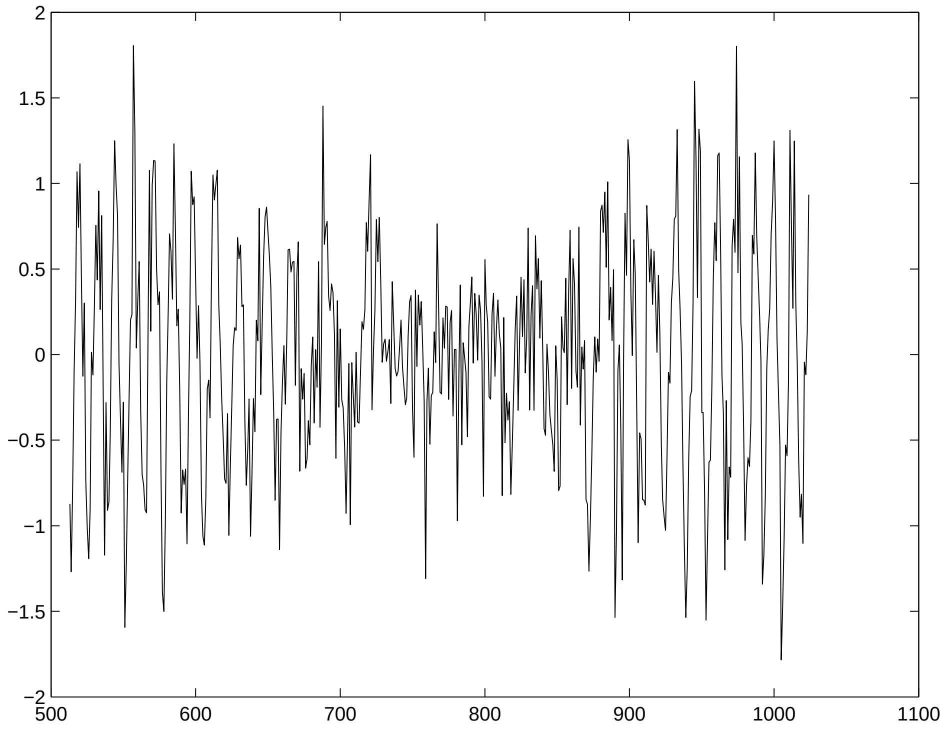
Chirp



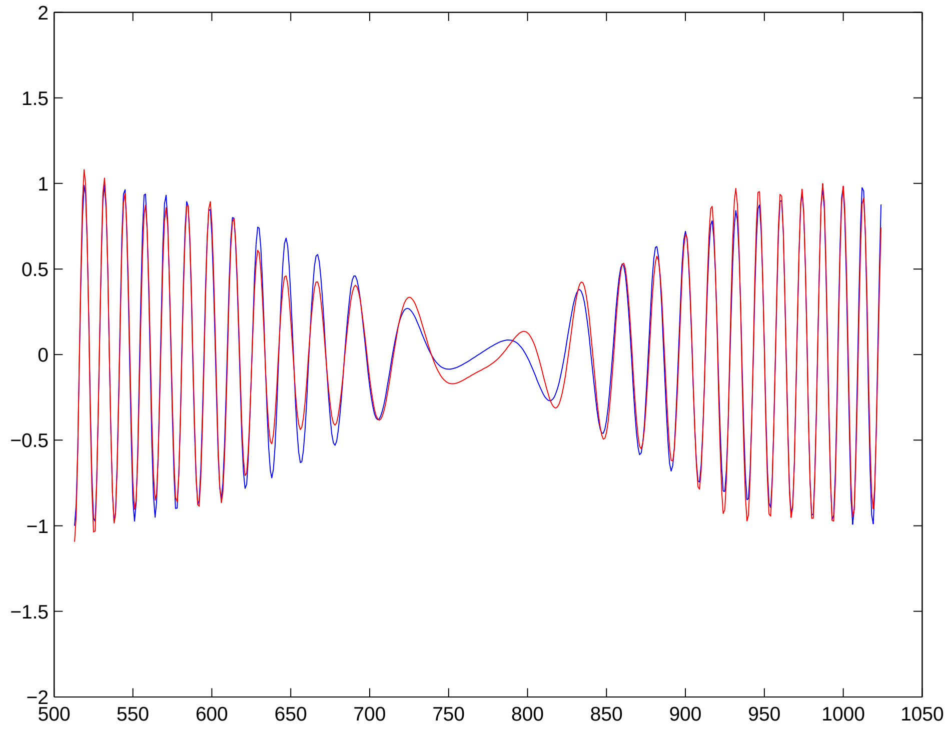
Noisy chirp



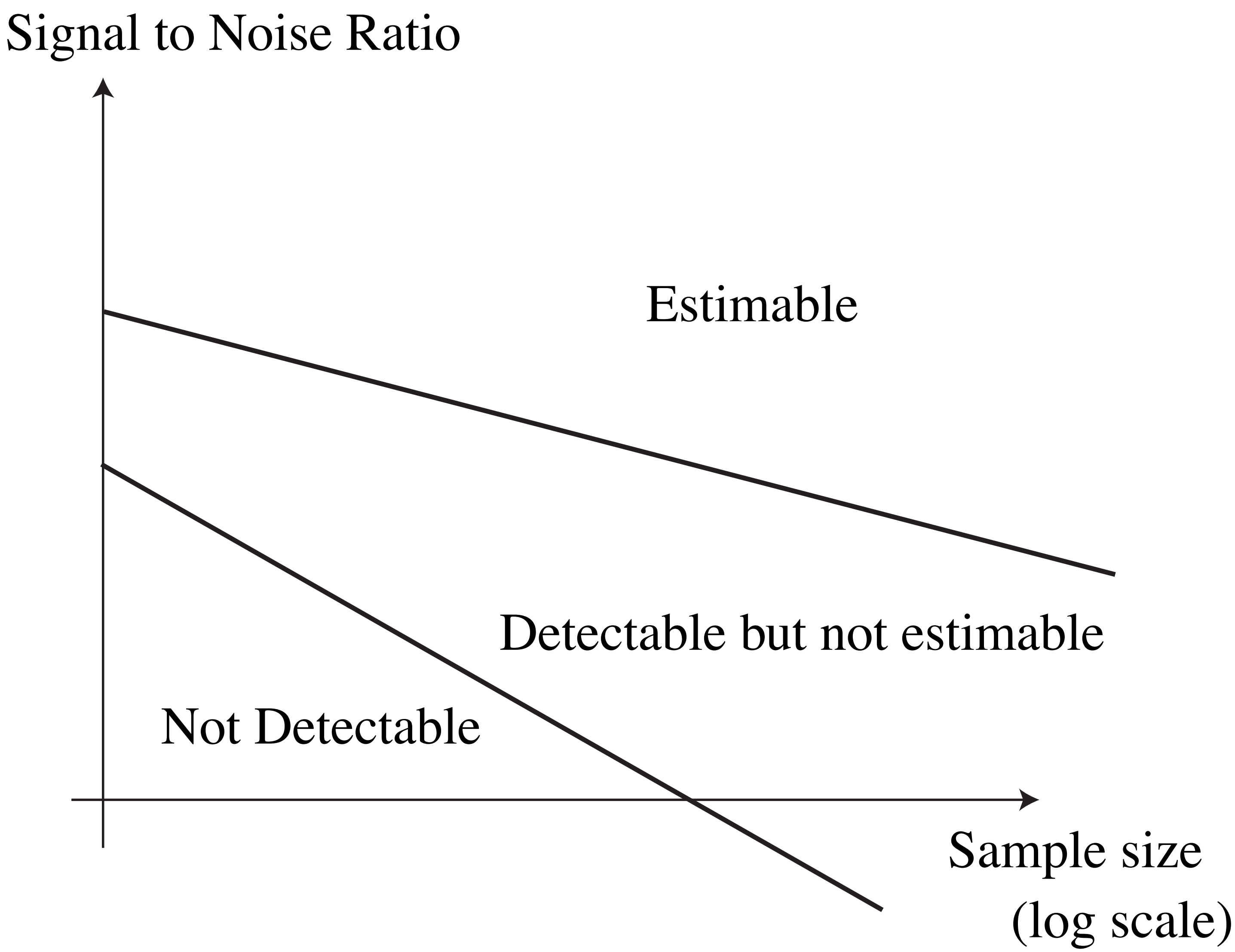
‘De-noised’ chirp



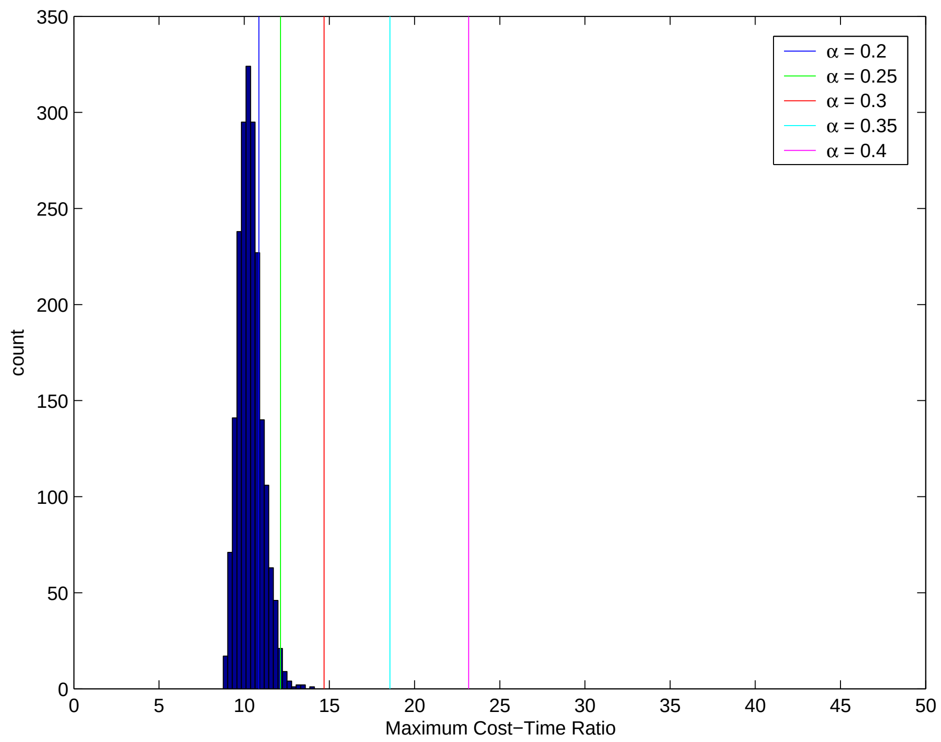
Noisy chirp



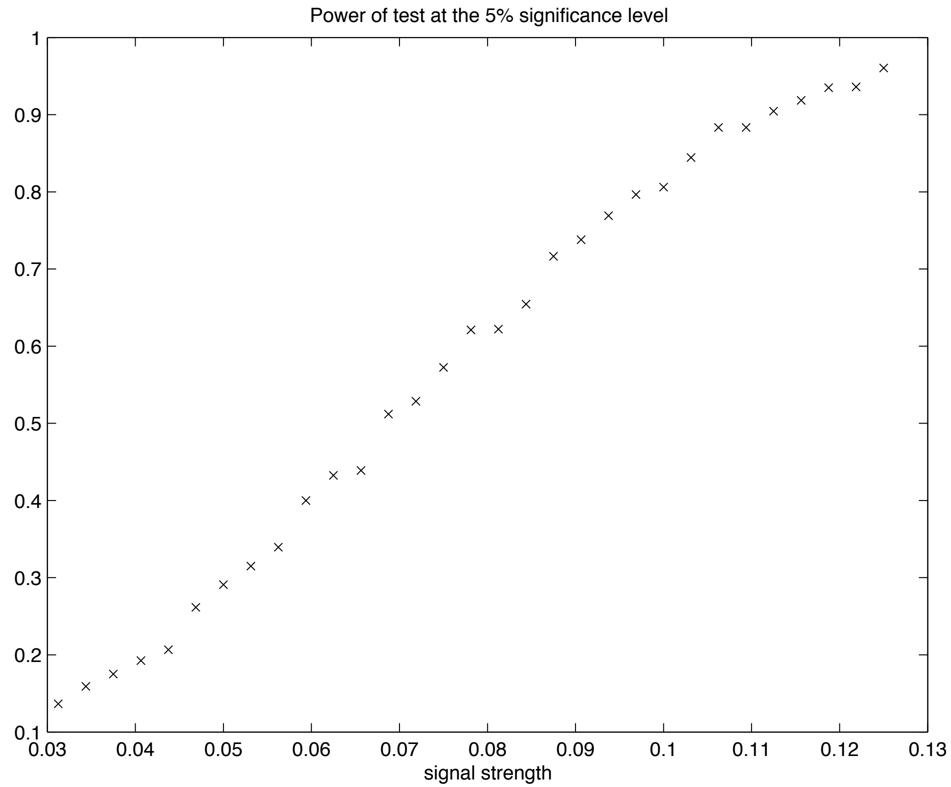
Original chirp and reconstruction


Signal to Noise Ratio
Estimable
Detectable but not estimable
Not Detectable
Sample size
(log scale)

Comparison Chirplet Detection vs. Optimal Detection



Chirplet detection



Power curve of simple LRT (Z score)

Summary

- Structured algorithms
- Methodology allows to detect signals whenever their strength makes them detectable by any method, no matter how intractable.
- Computational infrastructure: ChirpLab 1.0, ChirpLab 2.0
- Promising early numerical experiments
- Many extensions
  - Chirps with (unknown) finite duration and location
  - Several chirps—interfering or not.