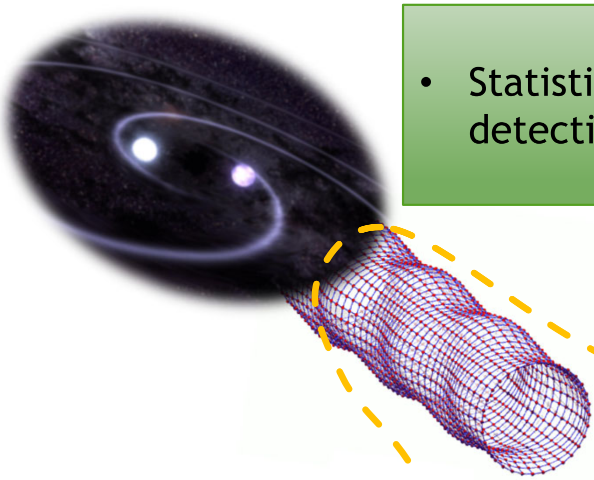
Signal detection and estimation

HUST Summer school, China

Soumya D. Mohanty

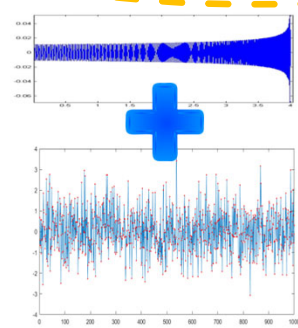
UTRGV

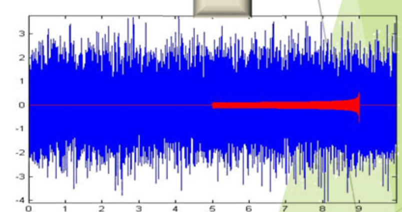


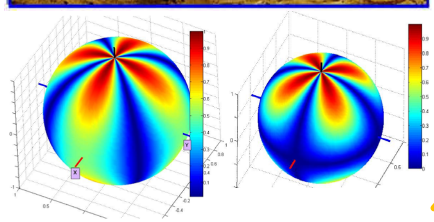
- Statistical theory of signal detection and estimation

What is the signal shape?

Signal present?

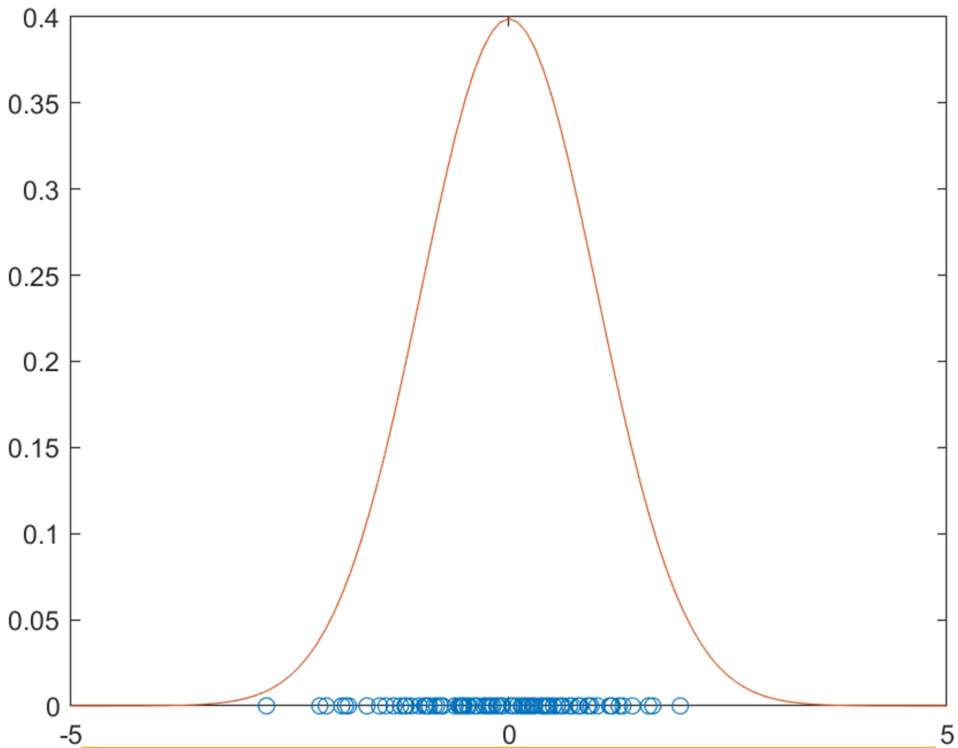




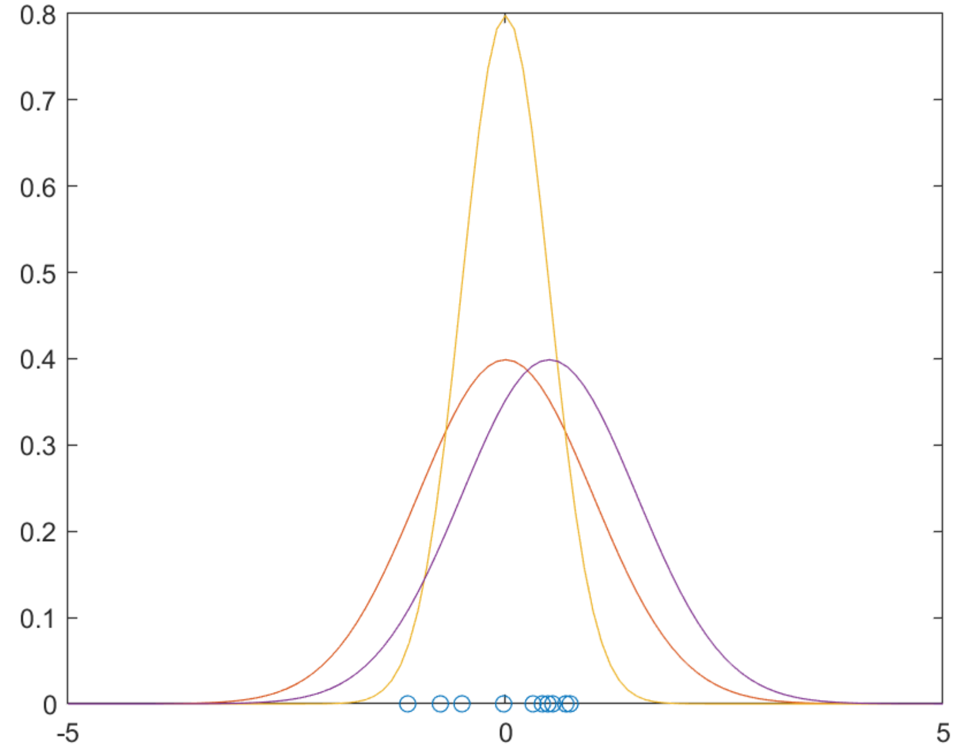


Estimation problem

Simple estimation problem



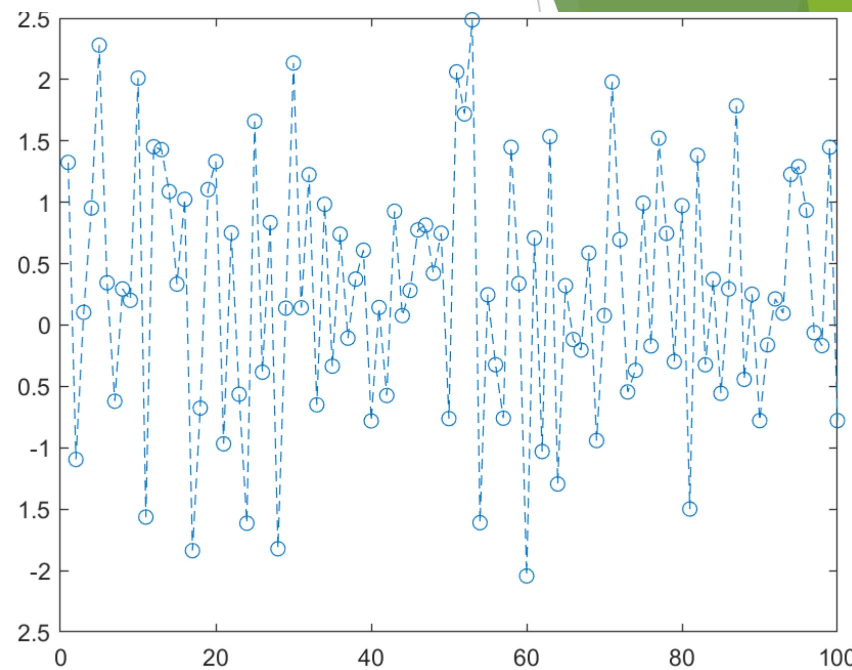
Given a pdf, we can draw trial values (data) from it



Estimation: Given trial values (data), what pdf were they drawn from?

General estimation problem

- **Given**: data $\bar{y} \in \mathbb{R}^N$
  - realization of a stochastic process $\bar{Y} = (Y_0, Y_1, \dots, Y_{N-1})$
- **Given**: set of possible joint pdf's describing the stochastic process : $p_{\bar{Y}}(\bar{y}\,;\Theta)$
  - Θ: a set of parameters
  - $\bar{y}$ is drawn form one of these pdf's with parameters Θ_{true}
  - The value of Θ_{true} is unknown
- Task: Estimate Θ_{true}



Example: Assume that the given data is drawn from WGN with unknown mean μ and unit variance

The joint pdf of the data is

$$p_{\bar{Y}}(\bar{y};\mu) = \prod_{i=0}^{99} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i-\mu)^2\right)$$

Task: Estimate μ_{true}

Exercise: Prove, starting with multivariate Normal pdf and $C_{ij} = \delta_{ij}$

Likelihood

- We would like to pick the pdf that is most "likely" to have produced the data: If we use this choice over a large number of data realizations, we should get the best estimation performance on the average

- One way to make this idea mathematically precise is the **likelihood function**
  - Set of parameters: Θ
  - Joint pdf of the data: $p_{\bar{Y}}(\bar{y};\Theta)$
  - Likelihood function: consider $p_{\bar{Y}}(\bar{y};\Theta)$ as a function of Θ for the given $\bar{y}$ (data)
  - Alternative notation: $\mathrm{L}(\Theta;\bar{y})$

- A high likelihood value means the corresponding pdf gives a higher probability of occurrence for the given data

Example: WGN with unknown mean μ and unit variance

- Set of parameter Θ is just μ
- $\bar{y} = (y_0, y_1, \dots, y_{N-1})$
- The joint pdf of the data is

$$p_{\bar{Y}}(\bar{y};\mu) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i-\mu)^2\right)$$

- Likelihood function

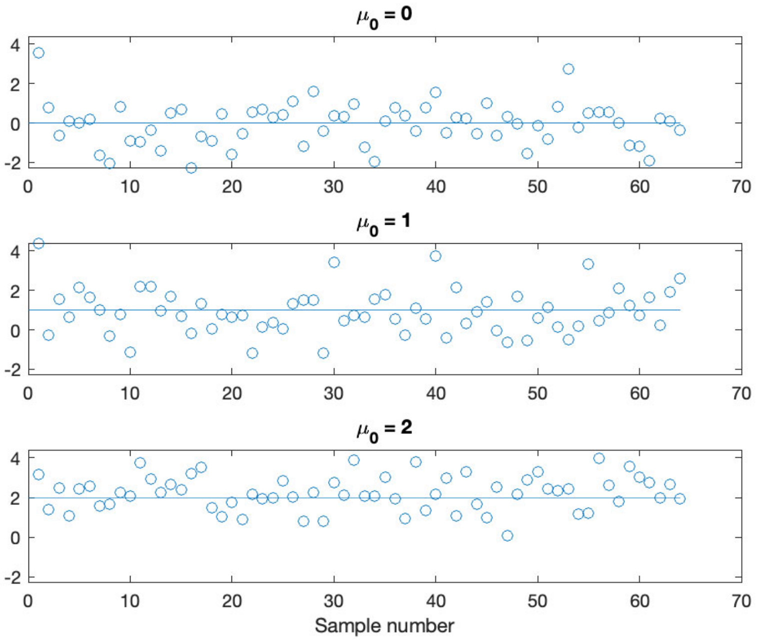
$$\mathrm{L}(\mu;\ \bar{y}) = p_{\bar{Y}}(\bar{y};\mu)$$

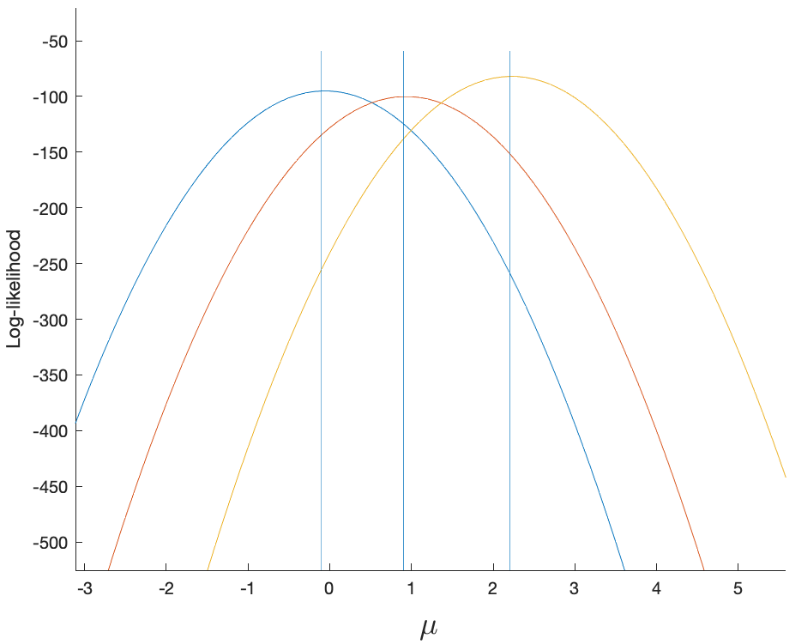
Likelihood function: WGN with unknown mean

Likelihood function $L(\mu; \bar{y})$:

$$\prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \mu)^2\right)$$

See DATASCIENCE_COURSE / DETEST / loglikewgndc.mlx





Maximum Likelihood Estimation

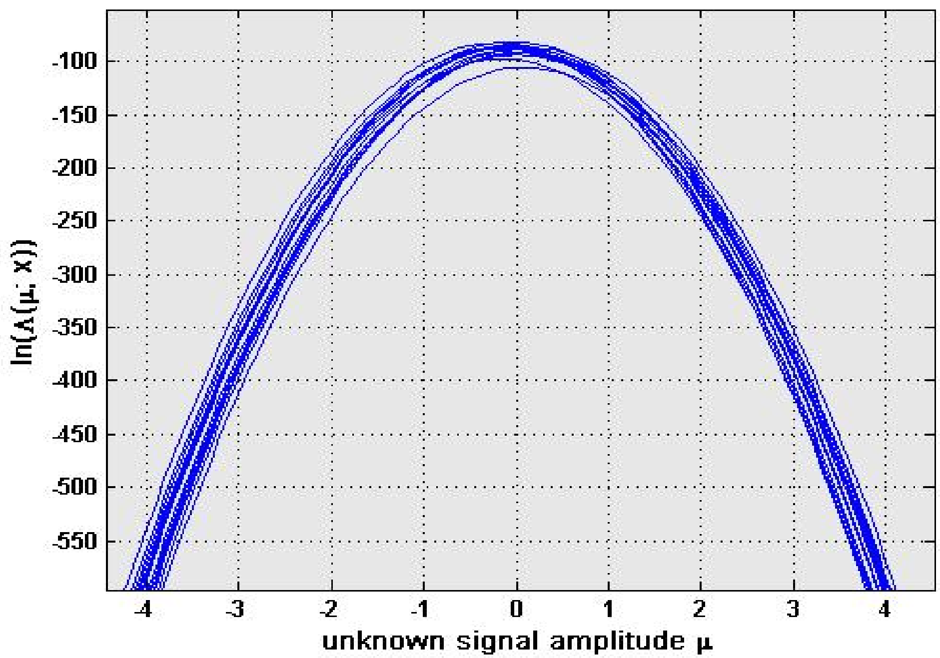
- Find Θ at which the Likelihood, $L(\Theta; \bar{y})$, has maximum value
  - This value of Θ is called the Maximum Likelihood Estimate (MLE)

$$\Theta_{MLE} = \arg\max_{\Theta} L(\Theta; \bar{y})$$

- Instead of maximizing the likelihood, we can use any monotonic function of the likelihood
  - e.g., $\ln(L(\Theta; \bar{y}))$ (log-likelihood)
- MLE is just one possible way to get an estimated value; other estimators are possible

Estimation error

- For any estimation method, the estimated value of the parameters will not match their true values due to the presence of noise
- An estimate must be provided with an estimated range of variation: estimation error



Example

- WGN with mean μ_0
- Different realizations, $\bar{y}_i$, of data
- Log-likelihood function, $\ln(L(\mu; \bar{y}_i))$, for each realization
- Scatter in the location of the maximum of the log-likelihood → Scatter in estimated parameter
- Estimation error could be given as the standard deviation of this scatter
- Note: For given observed data, the estimation error itself is an estimate based on hypothetical (or simulated) data realizations

Cramer-Rao Lower Bound

- CRLB: Lower bound on the variance of the estimate (for any estimator) $\widehat{\Theta}$

<u>Single parameter</u>

- Unbiased estimator: $E[\widehat{\Theta}] = \Theta_{true}$
- Variance of estimate:

$$\mathrm{var}(\widehat{\Theta}) = \Sigma_{\Theta}^2 = E[(\widehat{\Theta} - E[\widehat{\Theta}])^2] \geq \frac{1}{I(\Theta_{true})}$$

$$\text{Fisher information: } I(\Theta) = -E\left[\underbrace{\frac{\partial^2 \ln L(\Theta;\bar{Y})}{\partial \Theta^2}}_{Curvature}\right]$$

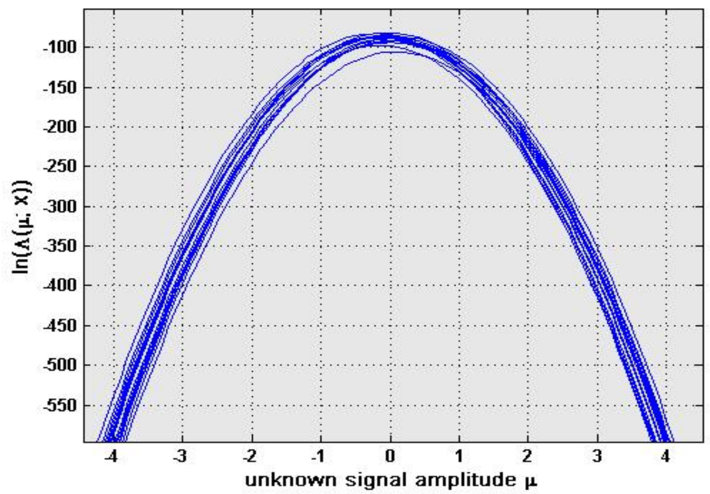
- **Exercise:** Compute CRLB for DC signal in WGN

<u>Multiple parameters</u>

- Unbiased estimator: $E[\widehat{\Theta}_i] = \Theta_{i,true}$
- Covariance matrix of estimates:

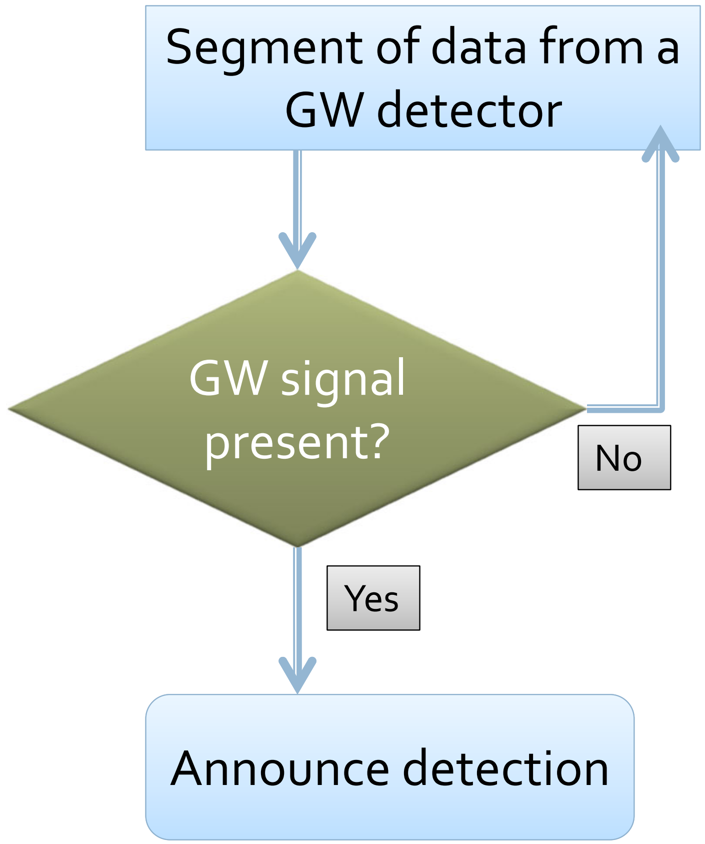
$$\text{Fisher information matrix } I_{ij}(\Theta) = -E\left[\frac{\partial^2 \ln L(\Theta;\bar{Y})}{\partial \Theta_i \partial \Theta_j}\right]$$

$$C_{ij} = E[(\widehat{\Theta}_i - E[\widehat{\Theta}_i])(\widehat{\Theta}_j - E[\widehat{\Theta}_j])] = \left(\mathrm{I}^{-1}(\Theta_{true})\right)_{ij}$$



DETECTION THEORY

Detection problem



Hypothesis testing

Given GW data we need to decide between two possibilities

Null hypothesis (H_0)

- $\bar{y} = \bar{n}$
- $p_{\bar{Y}}(\bar{y}|H_0)$: Joint pdf of the data is that of noise and **no signal**
- If decide H_0: Discard the data and get new data

Alternative hypothesis (H_1)

- $\bar{y} = \bar{s}(\Theta) + \bar{n}$ such that $\bar{s}(\Theta) \neq 0$
- $p_{\bar{Y}}(\bar{y}\,|H_1;\Theta)$: Joint pdf of the data is that of noise **plus signal**
- If decide H_1: Estimate Θ

False alarm and False dismissal probabilities

Two types of decision errors

False dismissal

deciding H_0 when H_1 is true

False alarm

deciding H_1 when H_0 is true

False alarm and False dismissal probabilities

Two types of decision errors

Noise in data $\Rightarrow$ Decision outcome is a **random variable**

False dismissal

deciding H_0 when H_1 is true

Probability: Q_1

False alarm

deciding H_1 when H_0 is true

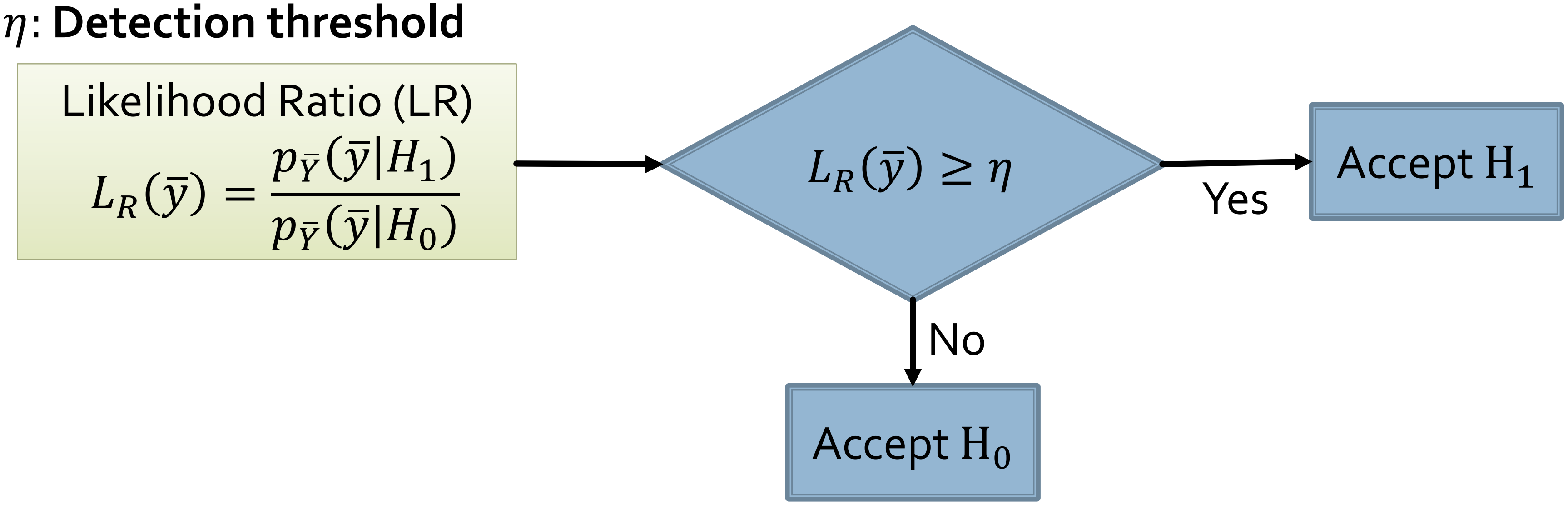
Probability: Q_0

Likelihood Ratio Test

- Any decision rule: based on the value of a detection statistic $\Gamma(\bar{y})$ to pick H_0 or H_1
- Is there a best $\Gamma(\bar{y})$? What criterion to use for "best" when comparing detection statistics?

<u>**Binary hypothesis case:** No free parameters Θ under H_0 or H_1</u>

- **Neyman-Pearson criterion**: Minimize Q_1 for fixed Q_0 $\rightarrow$ Optimal decision rule exists
- The decision rule is called the Likelihood Ratio (LR) test
- η: **Detection threshold**



Likelihood Ratio Test

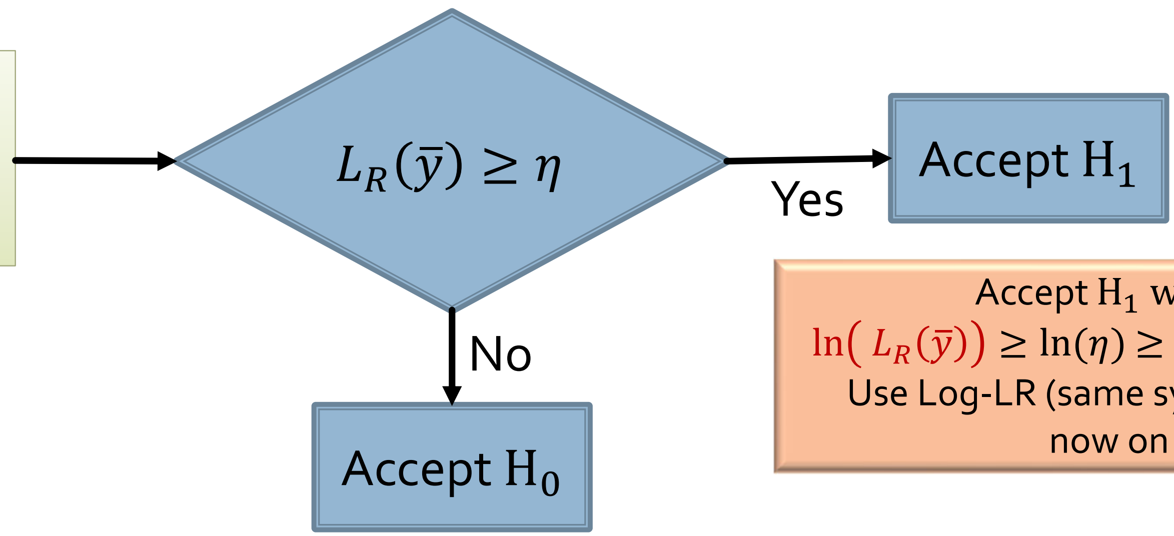
- Any decision rule: based on the value of a detection statistic $\Gamma(\bar{y})$ to pick H_0 or H_1
- Is there a best $\Gamma(\bar{y})$? What criterion to use for "best" when comparing detection statistics?

Binary hypothesis case: No free parameters Θ under H_0 or H_1

- **Neyman-Pearson criterion**: Minimize Q_1 for fixed Q_0 $\rightarrow$ Optimal decision rule exists
- The decision rule is called the Likelihood Ratio (LR) test
- η: **Detection threshold**

Likelihood Ratio (LR)

$$L_R(\bar{y}) = \frac{p_{\bar{Y}}(\bar{y}|H_1)}{p_{\bar{Y}}(\bar{y}|H_0)}$$



Accept H_1 when
$\ln(L_R(\bar{y})) \geq \ln(\eta) \geq$ new threshold
Use Log-LR (same symbol) from now on

Detection and false alarm probabilities

$$\overbrace{\text{False alarm Probability}}^{L_R(\bar{y})\geq\eta \text{ when } H_0 \text{ is true}}: \; Q_0 = \int_{\eta}^{\infty} p_{L_R}(x|H_0)dx$$

$$\overbrace{\text{False dismissal Probability}}^{L_R(\bar{y})\leq\eta \text{ when } H_1 \text{ is true}}: Q_1 = \int_{-\infty}^{\eta} p_{L_R}(x|H_1)dx$$

$$\overbrace{\text{Detection Probability}}^{\text{Pick } H_1 \text{ when } H_1 \text{ is true}}: \; Q_d = \int_{\eta}^{\infty} p_{L_R}(x|H_1)dx$$

$$= 1 - Q_1$$

Composite hypothesis test

(Also see Chapter 1 Sec 4 of textbook)

- The detection problem, in general, is **not** a binary hypothesis test: parameter values Θ are unknown

- H_0: $\bar{y}$ is only noise
- H_1: Not one but many alternative hypotheses corresponding to each possible value of $\Theta \rightarrow$ **Composite hypotheses test**

- Neyman-Pearson criterion for composite hypotheses: **No solution**, *in the general case*, for an optimal decision surface

GLRT: Generalized Likelihood Ratio Test

- $p_{\bar{Y}}(\bar{y}|H_0)$: pdf of the data under H_0
- $p_{\bar{Y}}(\bar{y}|H_1;\Theta)$: pdf of the data under H_1
  Note: there are many alternative hypotheses now corresponding to the different possible values of Θ

  Likelihood Ratio is a function of Θ:

$$L_R(\Theta;\bar{y}) = \frac{p(\bar{y}|H_1;\Theta)}{p(\bar{y}|H_0)}$$

Generalized Likelihood Ratio Test (GLRT): detection statistic

$$L_G(\bar{y}) = \max_{\Theta} L_R(\Theta;\bar{y})$$

- We can use the log-GLRT (same symbol): $L_G(\bar{y}) = \max_{\Theta} \ln L_R(\Theta;\bar{y})$

GLRT and MLE

- MLE and GLRT are intimately related:

GLRT	MLE
$L_G(\bar{y}) = \max_{\Theta} \frac{p(\bar{y}\|H_1;\Theta)}{p(\bar{y}\|H_0)} = \frac{\max_{\Theta} L(\Theta;\bar{y})}{p_{\bar{Y}}(\bar{y}\|H_0)}$ Value of the maximum of $L(\Theta;\bar{y})$ required	$\bar{\theta}_{MLE} = \arg\max_{\bar{\theta}} L(\Theta;\bar{y})$ Location of the maximum of $L(\Theta;\bar{y})$ required

- No separate MLE required when doing GLRT

GLRT for Gaussian noise

Gaussian noise: a common noise model

- Joint pdf of **any** subsequence of the noise is a zero-mean multivariate normal pdf

$$p_{\bar{X}}(\bar{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}\|\bar{x}\|^2)$$

- $\bar{x} = (x_0, x_1, \ldots, x_{N-1}) \in R^N$ (row vector)
- $\boldsymbol{C}$: Covariance matrix
- $|\boldsymbol{C}|$: Determinant of $\boldsymbol{C}$
- $\|\bar{x}\|^2 = \langle \bar{x}, \bar{x} \rangle$
- Inner product: $\langle \bar{x},\ \bar{y} \rangle = \ \bar{x}\boldsymbol{C}^{-1}\bar{y}^T$

Inner product for stationary noise

- White Gaussian Noise

$$\langle \bar{x}, \bar{y} \rangle \rightarrow \sum_{k=0}^{N-1} x_k y_k \Rightarrow \|\bar{x}\|^2 = \sum_{k=0}^{N-1} x_k^2$$

- Stationary Gaussian noise with Power Spectral Density (PSD) $S_n(f)$

$$\langle \bar{x}, \bar{y} \rangle \rightarrow \frac{\Delta}{N} \tilde{x} (\tilde{y}^\dagger \,./\, \bar{S}_n{}^T)$$

Where $\tilde{x} = F\bar{x}$ is the DFT

./ : Element-by-element division

GLRT for Gaussian noise in GW data analysis

- Gaussian stationary noise is the main model used in all GW data analysis algorithms
  - Extra steps are needed (e.g., line removal) to deal with the effects of non-Gaussian and non-stationary noise
- We will obtain the GLRT for Gaussian stationary noise by progressively maximizing the log-likelihood ratio over the parameters:
  - Amplitude
  - Time of arrival
  - (Oscillatory signal →) Initial phase (Exercise)
  - Amplitude, time of arrival, and initial phase
- These parameters are common to all oscillatory transient signals (e.g., binary inspirals)

GLRT for Gaussian noise: Starting point
(See Chapter 1.3 and 1.4 of textbook)

- Data :

$$\bar{y} = \bar{s}(\Theta) + \bar{n};$$

$\bar{n}$: realization of zero mean Gaussian noise
$\bar{s}(\Theta)$: GW signal (if present)
Θ: set of signal parameters

- $\Rightarrow \bar{y}$ is a realization of Gaussian noise with mean: $E[Y_i] = s_i(\Theta)$

- (Exercise) GLRT:

$$\mathrm{L}_G(\bar{y}) = \max_{\Theta} \ln L_R(\Theta; \bar{y}) = \max_{\Theta}\left[-\frac{1}{2}\|\bar{y} - \bar{s}(\Theta)\|^2 + \frac{1}{2}\|\bar{y}\|^2\right]$$

$$L_G(\bar{y}) = \max_{\Theta}\left(\langle \bar{y}, \bar{s}(\Theta)\rangle - \frac{1}{2}\|\bar{s}(\Theta)\|^2\right)$$

Amplitude

Amplitude normalization

- Convenient normalization of signals:

$$\bar{s}(\Theta) = \frac{\|\bar{s}(\Theta)\|}{\|\bar{s}(\Theta)\|}\bar{s}(\Theta\,) = \overbrace{\|\bar{s}(\Theta)\|}^{\substack{vector\\ length}} \; \overbrace{\frac{\bar{s}(\Theta)}{\|\bar{s}(\Theta)\|}}^{unit\ vector} = \|\bar{s}(\Theta)\|\bar{q}(\Theta) = A\,\bar{q}(\Theta)$$

$\mathrm{A} = \|\bar{s}(\Theta)\|$; $\|\bar{q}(\Theta)\| = 1 \Rightarrow \bar{q}(\Theta)$ is the unit norm signal (signal template)

- Any overall factor in $\bar{s}(\Theta)$ is now absorbed in A
- A is called the signal-to-noise ratio (SNR)
  - $\ln L_R(\Theta;\bar{y}) = \langle \bar{y}, \bar{s}(\Theta)\rangle - \frac{1}{2}\|\bar{s}(\Theta)\|^2$
  - (Exercise) Prove:

$$SNR = \frac{E[\ln L_R\,|H_1]}{var\,(\ln L_R\,|H_0)}$$

- SNR measures how far a signal will shift the value of $\ln L_R$ (on average) compared to its scatter due to noise alone: The greater the shift, the better will be the detectability of the signal

Amplitude normalization

- Convenient normalization of signals:

$$\bar{s}(\Theta) = \frac{\|\bar{s}(\Theta)\|}{\|\bar{s}(\Theta)\|}\bar{s}(\Theta) = \overbrace{\|\bar{s}(\Theta)\|}^{\substack{vector\\ length}} \overbrace{\frac{\bar{s}(\Theta)}{\|\bar{s}(\Theta)\|}}^{unit\ vector} = \|\bar{s}(\Theta)\|\bar{q}(\Theta) = A\,\bar{q}(\Theta)$$

$A = \|\bar{s}(\Theta)\|$; $\|\bar{q}(\Theta)\| = 1 \Rightarrow \bar{q}(\Theta)$ is the unit norm signal (signal template)

- The set of parameters now becomes

$$\Theta = \{A, \Theta'\}$$

where Θ' denotes all remaining parameters

- $\bar{q}(\Theta)$ now depends only on Θ':

$$\bar{q}(\Theta) \to \bar{q}(\Theta')$$

- **Then**

$$L_G(\bar{y}) = \max_{\Theta}\left(\langle\bar{y}, \bar{s}(\Theta)\rangle - \frac{1}{2}\|\bar{s}(\Theta)\|^2\right) \to L_G(\bar{y}) = \max_{A,\Theta'}\left(A\langle\bar{y}, \bar{q}(\Theta')\rangle - \frac{1}{2}A^2\right)$$

GLRT

(Also see Appendix C.1 of textbook)

$$L_G(\bar{y}) = \max_{A,\Theta'}\left(A\langle\bar{y},\bar{q}(\Theta')\rangle - \frac{1}{2}A^2\right) = \max_{\Theta'}\left(\max_{A}\left(A\langle\bar{y},\bar{q}(\Theta')\rangle - \frac{1}{2}A^2\right)\right)$$

- Solution of inner minimization:

$$A = \langle\bar{y},\bar{q}(\Theta')\rangle$$

- Hence

$$L_G = \max_{\Theta'}\ \langle\bar{y},\bar{q}(\Theta')\rangle^2$$

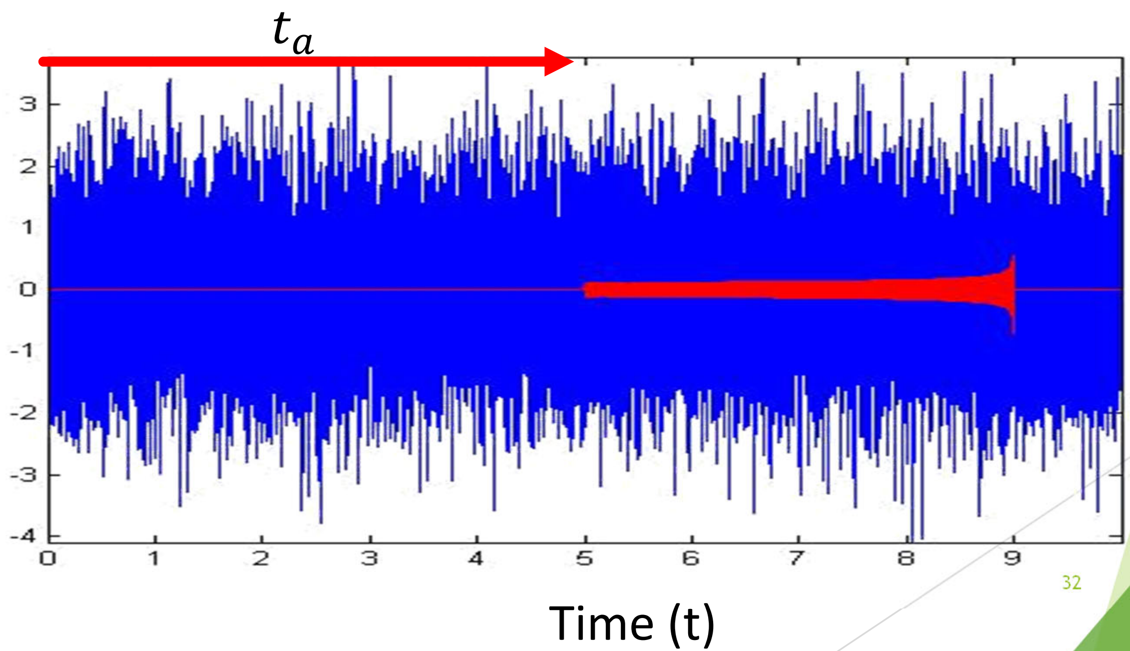
- (From now, $\Theta' \rightarrow \Theta$ the set of parameters **besides SNR**)

$$L_G = \max_{\Theta}\langle\bar{y},\bar{q}(\Theta\ )\rangle^2$$

Time of arrival

Time of arrival

- Signal start time ("time of arrival") parameter: t_a
- $q(t; t_a, \Theta') = q^{(0)}(t - t_a; \Theta')$
- $q^{(0)}(t; \Theta')$: template at $t_a = 0$



Matched filtering

- $L_G = \max_{\Theta}\langle \bar{y}, \bar{q}(\Theta)\rangle^2 = \max_{\Theta', t_a}\langle \bar{y}, \bar{q}(t_a, \Theta')\rangle^2 = \max_{\Theta'}\left(\max_{t_a}\langle \bar{y}, \bar{q}(t_a, \Theta')\rangle^2\right)$
- Obtaining $\langle \bar{y}, \bar{q}(t_a, \Theta')\rangle$ as a function of t_a is a **filtering** operation

$$\langle \bar{y}, \bar{q}(t_a, \Theta')\rangle = \bar{y}\mathbf{C}^{-1}\,\bar{q}^T(t_a, \Theta') = \bar{z}\bar{q}^T(t_a, \Theta')$$

Assume Uniform sampling,

$$= \sum_{k=0}^{N-1} z_k q_k(t_a, \Theta') = \frac{1}{\underbrace{\delta t}_{\substack{Sampling\\ interval}}}\delta t \sum_{k=0}^{N-1} z_k q^{(0)}(t_k - t_a, \Theta')$$

Digital filtering: Shift → Multiply → Sum

- Filtering done with filter that “matches” the signal $\rightarrow$ **Matched filtering**

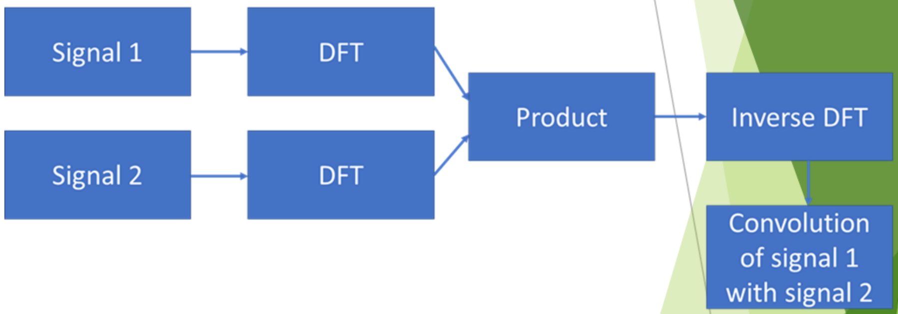
- Since $q^{(0)}(t;\Theta')$ is finite in length, the filtering operation is FIR filtering
- Convolution theorem $\Rightarrow \langle \bar{y}, \bar{q}\,(t_a, \Theta')\rangle$ can be implemented efficiently using FFT based correlation

1. Divided (sample by sample) FFT of data $\tilde{y}$ by PSD $\rightarrow \tilde{z}$
2. Multiply (sample by sample) $\tilde{z}$ and (complex conjugate) of FFT of template (having $t_a = 0$)
    1. Complex conjugate because no reflection operation on template → Correlation, not convolution
3. Take inverse FFT

$$F^{-1}\left[(\tilde{y}./\bar{S}_n)\ .*\ (\tilde{q}^{(0)})^*\right] \rightarrow \langle \bar{y}, \bar{q}_a(t_a, \Theta')\rangle$$

$$\text{for } t_a = k\Delta, k = 0,1, \dots, N-1$$

$$\mathrm{DFT}[\bar{x} * \bar{y}] = \tilde{x}.*\,\tilde{y} \Rightarrow \bar{x} * \bar{y} = \mathrm{DFT}^{-1}[\tilde{x}.*\,\tilde{y}]$$



Unknown Amplitude, initial phase, and time of arrival

$$\mathrm{L}_G = \max_{\Theta'} \left[\max_{t_a} (\langle \bar{y}, \bar{q}_0(t_a, \Theta') \rangle^2 + \langle \bar{y}, \bar{q}_1(t_a, \Theta') \rangle^2) \right] = \max_{\Theta'} \lambda(\Theta')$$

- $\bar{q}_p(t_a, \Theta'), p = 0, 1$: Quadrature templates

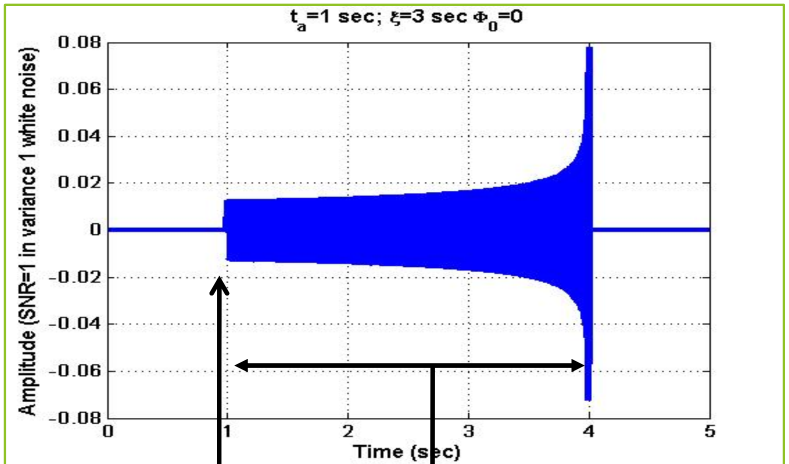
1. For a given Θ', Evaluate $\langle \bar{y}, \bar{q}_p(t_a, \Theta') \rangle$, $p = 0, 1$, using FFT based filtering $\rightarrow$ Two output time series
2. Square the samples of each time series
3. Add the two resulting time series
4. Find the maximum of this time series $\rightarrow$ get $\lambda(\Theta')$

Again, $\Theta' \rightarrow \Theta$: Excluding amplitude, initial phase, time of arrival

$$L_G = \max_{\Theta} \lambda(\Theta)$$

Example: Newtonian Binary Inspiral signal

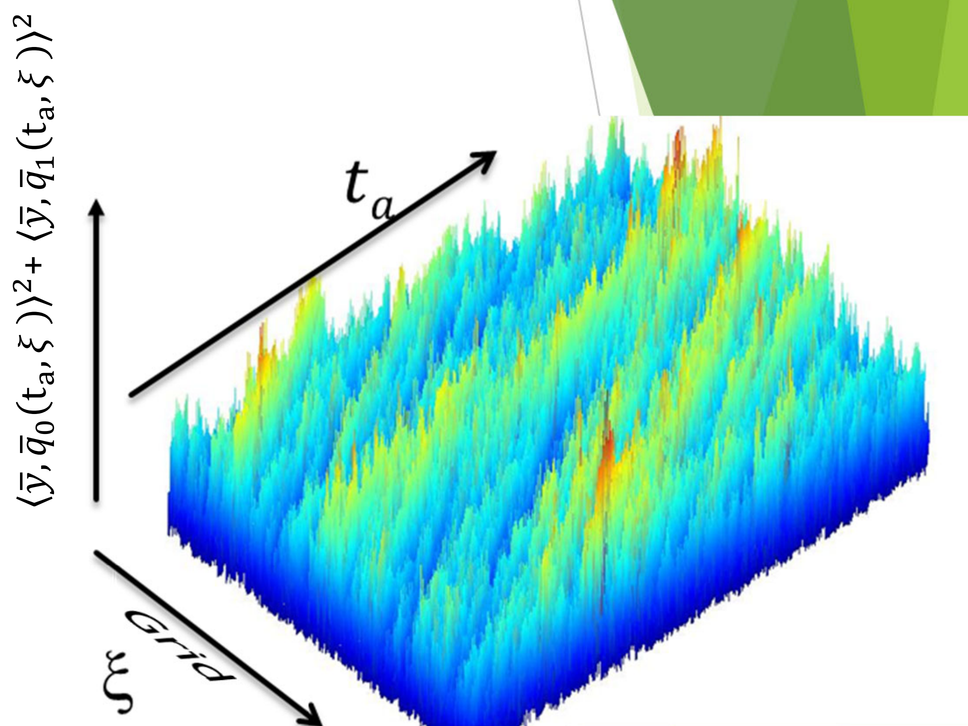
Newtonian inspiral template



t_a : time of arrival

Φ_0 : Initial phase

ξ : Chirp time



Summary: GLRT for Gaussian stationary noise

- We have obtained the General form of the GLRT for transient oscillatory signals (e.g., binary inspiral)
- Most GW data analysis algorithms are designed for the case of Gaussian stationary noise (but must be enhanced for real data)
- In GW data analysis,

Extrinsic parameters	Intrinsic parameters
• Parameters that can be maximized analytically or efficiently	• Parameters that must be maximized over numerically and/or are challenging to maximize

Summary: GLRT for Gaussian stationary noise

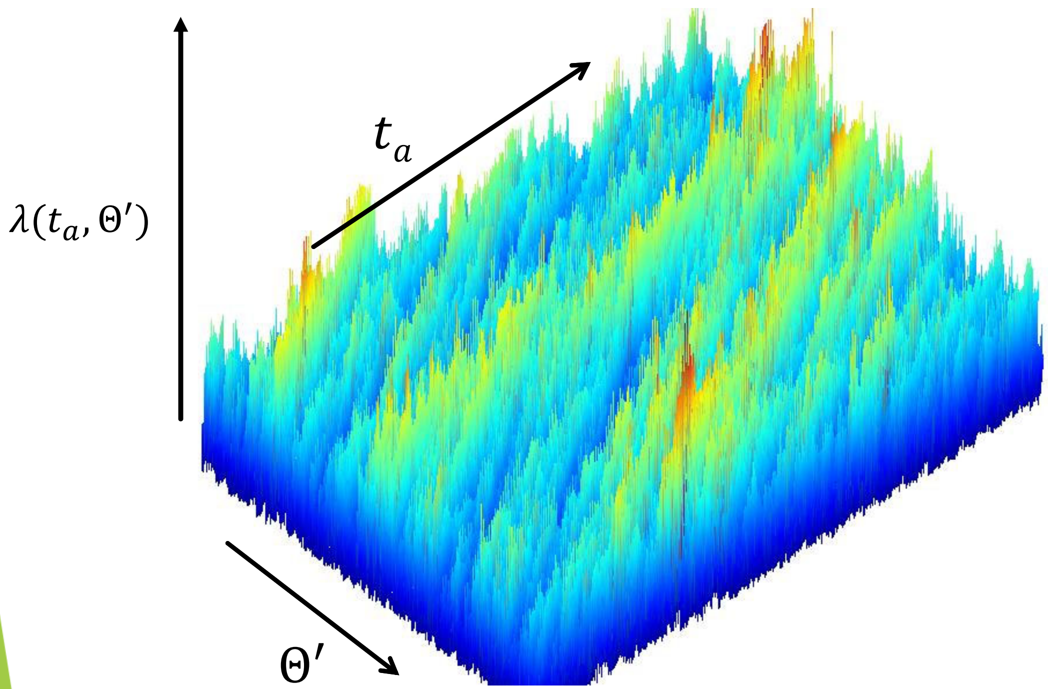
- General form of GLRT in GW data analysis:

$$L_G = \max_{\substack{intrinsic \\ parameters}} \max_{\substack{extrinsic \\ parameters}} \langle \bar{y}, \bar{q}(\Theta_{extrinsic}, \Theta_{intrinsic}) \rangle^2$$

$$= \max_{\substack{intrinsic \\ parameters}} \lambda(\Theta_{intrinsic})$$

- The maximization of $\lambda(\Theta_{intrinsic})$ must be done using numerical optimization methods

Binary inspiral search



The numerical optimization problem is

1. Intrinsically difficult
   - Large number of maxima
   - Becomes worse as the number of parameters increases
2. Grid-search: Computationally expensive
   - Binary inspiral network analysis for ground-based detectors grid based search: $\approx 10^8$ points in $\Theta_{intrinsic}$ space with $\approx 10^7$ floating point operations per point (1 hour segments) $\Rightarrow$ 0.3 Tflops to just keep up with the incoming data rate
   - Computational bottleneck ⇒ current searches follow a sub-optimal approach ⇒ Lower sensitivity ⇒ Reduced rate of detections

Homework

MLE for signal in general Gaussian noise

- Data model:

$$\underbrace{\bar{y}}_{\substack{Data\\ realization}} = \underbrace{\bar{s}(\Theta)}_{\substack{Signal\ with\\ parameters\\ \Theta}} + \underbrace{\bar{n}}_{\substack{Noise\\ realization}}$$

- Where $\bar{n}$ is a realization of zero mean Gaussian noise
  - Review: Homework on multivariate Normal pdf

$$p_{\bar{X}}(\bar{x}\,) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\|\bar{x}\|^2\right)$$

$$\|\bar{x}\|^2 = \bar{x}\mathbf{C}^{-1}\,\bar{x}^T$$

MLE for signal in Gaussian noise

- For a given Θ, the probability of getting the data realization is the probability of getting the noise realization: $\bar{y} - \bar{s}(\Theta)$
- Therefore, for a given Θ, the joint pdf of the data $\bar{y}$ is:

$$p_{\bar{Y}}(\bar{y};\Theta) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\|\bar{y} - \bar{s}(\Theta)\|^2\right)$$

- Log-likelihood function

$$L(\Theta;\bar{y}) = const. - \frac{1}{2}\|\bar{y} - \bar{s}(\Theta)\|^2$$

$$\max_{\theta} L(\Theta;\bar{y}) \text{ is equivalent to } \min_{\theta}\|\bar{y} - \bar{s}(\Theta)\|^2$$

Least-squares for iid noise

- For iid noise: $\mathbf{C} = \sigma^2\mathbb{I}$ where $\mathbb{I}$ is the identity matrix

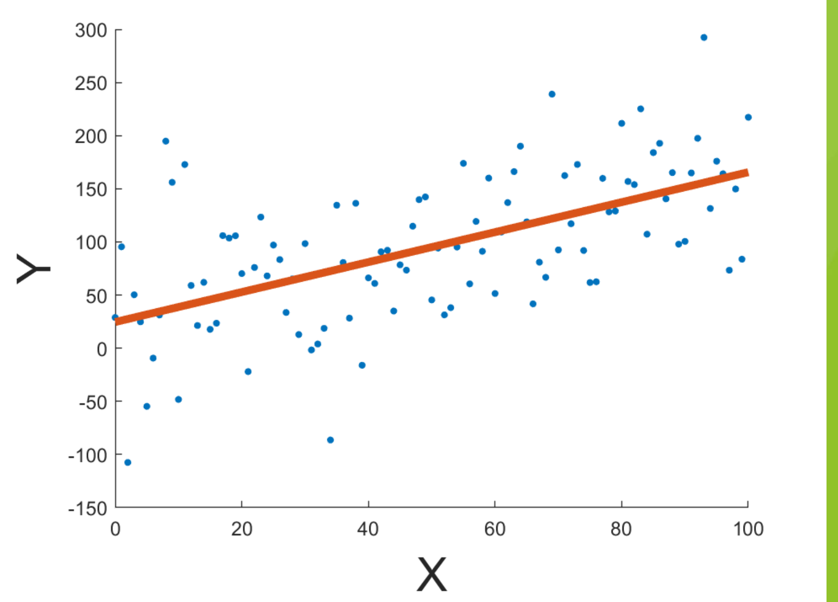
$$\min_{\theta}\|\bar{y} - \bar{s}(\Theta)\|^2 = \min_{\theta}\frac{1}{\sigma^2}\sum_{i=0}^{N-1}\left(y_i - s_i(\Theta)\right)^2$$

- This is nothing but the method of least-squares

- Example: Fitting a straight line to pairs of points $(y_i, x_i), i = 0,1,\ldots,N-1$

$$s_i = ax_i + b \rightarrow \Theta = \{a, b\}$$

$$\min_{a,b}\frac{1}{\sigma^2}\sum_{i=0}^{N-1}(y_i - (ax_i + b))^2$$



Additive dc signal in zero mean, stationary, white Gaussian noise

data: $\overline{x}$

Null hypothesis pdf:

$$p(\overline{x} \mid H_0) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=0}^{N-1} x_i^2 \right)$$

Alternative hypothesis pdf:

$$p(\overline{x} \mid H_1) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=0}^{N-1} (x_i - \mu)^2 \right)$$

Likelihood Ratio :

$$\Lambda(\mathbf{x}) = \frac{p(\overline{x} \mid H_1)}{p(\overline{x} \mid H_0)} = \exp\left(\frac{1}{\sigma^2} \left(\mu \sum_{i=0}^{N-1} x_i - \frac{1}{2} N \mu^2 \right) \right)$$

$$\Lambda(\mathbf{x}) \geq \eta \quad \Rightarrow \quad \frac{1}{N} \sum_{i=0}^{N-1} x_i \geq \frac{\ln \eta}{\mu} + \mu = \text{const.}(\Gamma)$$

Decision rule : Compute $\hat{\mu} = \frac{1}{N} \sum_{i=0}^{N-1} x_i$ and compare with a threshold

Use $Q_0 = \int_\Gamma^\infty dz \; p_{\hat{\mu}}(z \mid H_0)$ to fix Γ

GLRT: unknown amplitude & initial phase

Initial phase

- Any oscillatory signal will start from an initial phase that is unknown in general

- Sinusoidal signal
  - $s(t) = A\sin(2\pi f_0 t + \phi_0)$
  - Parameters: A, f_0, ϕ_0
- Linear chirp signal
  - $s(t) = A\sin(2\pi(f_0 t + f_1 t^2) + \phi_0)$
  - Parameters: A, f_0, f_1, ϕ_0
- Sine-Gaussian signal
  - $s(t) = A\exp\left(-\frac{(t-t_0)^2}{2\sigma^2}\right)\sin(2\pi f_0 t + \phi_0)$
  - Parameters: $A, t_0, \sigma, f_0, \phi_0$

Initial phase

- Any oscillatory signal will start from an initial phase that is unknown in general

$$\bar{s}(\Theta) = A\,\bar{q}(\Theta) \text{ where}$$

$$\bar{q}(\Theta) = \bar{q}(\phi_0, \Theta')$$

with

$$q_k(\phi_0, \Theta') = N(\Theta')a(t_k; \Theta')\sin(\phi(t_k; \Theta') + \phi_0)$$

- Sinusoidal signal
  - $s(t) = A\sin(2\pi f_0 t + \phi_0)$
  - Parameters: A, f_0, ϕ_0
- Linear chirp signal
  - $s(t) = A\sin(2\pi(f_0 t + f_1 t^2) + \phi_0)$
  - Parameters: A, f_0, f_1, ϕ_0
- Sine-Gaussian signal
  - $s(t) = A\exp\left(-\frac{(t-t_0)^2}{2\sigma^2}\right)\sin(2\pi f_0 t + \phi_0)$
  - Parameters: $A, t_0, \sigma, f_0, \phi_0$

Initial phase → Quadratures

$$q_k(\phi_0,\Theta') = N(t_k;\Theta')\sin(\phi(t_k;\Theta') + \phi_0)$$

$$\Rightarrow q_k(\phi_0,\Theta') = \underbrace{N(\Theta')a(t_k;\Theta')\sin(\phi(t_k;\Theta'))}_{q_{0,k}}\ \underbrace{\cos(\phi_0)}_{X} + \underbrace{N(\Theta')\,a(t_k;\Theta')\cos(\phi(t_k;\Theta'))}_{q_{1,k}}\,\underbrace{\sin(\phi_0)}_{Y}$$

- $\|\bar{q}(\phi_0,\Theta')\| = 1 \Rightarrow \langle\bar{q}_0,\bar{q}_0\rangle\cos^2\phi_0 + \langle\bar{q}_1,\bar{q}_1\rangle\sin^2\phi_0 + 2\langle\bar{q}_0,\bar{q}_1\rangle cos\phi_0\sin\phi_0$
- If $a(t_k;\Theta')$ varies slowly compared to the oscillation period, $\langle\bar{q}_0,\bar{q}_0\rangle \approx \langle\bar{q}_1,\bar{q}_1\rangle$ and $\langle\bar{q}_0,\bar{q}_1\rangle \approx 0 \Rightarrow \langle\bar{q}_0,\bar{q}_0\rangle \approx \langle\bar{q}_1,\bar{q}_1\rangle \approx 1$
- Therefore,

$$\bar{q}(\phi_0,\Theta') = X\,\bar{q}_0(\Theta') + Y\,\bar{q}_1(\Theta')$$

$$X^2 + Y^2 = 1$$

- $\bar{q}_{0,1}(\Theta')$: Quadrature components of template $\bar{q}(\phi_0,\Theta')$ or quadrature templates

Unknown amplitude and initial phase

- GLRT:

$$L_G = \max_{\Theta} \langle \bar{y}, \bar{q}(\Theta) \rangle^2$$
$$= \max_{\Theta'} \max_{\substack{X,Y \\ X^2+Y^2=1}} \left(X\langle \bar{y}, \bar{q}_0(\Theta') \rangle + Y\langle \bar{y}, \bar{q}_1(\Theta') \rangle \right)^2$$

- The quantity to be maximized is of the form

$$(X\langle \bar{y}, \bar{q}_0(\Theta') \rangle + Y\langle \bar{y}, \bar{q}_1(\Theta') \rangle)^2 = (n_1 A_1 + n_2 A_2)^2 = (\hat{n}.\bar{A})^2$$

Where $\hat{n} = (X, Y)$ is a unit vector since $X^2 + Y^2 = 1$

and $\bar{A} = (A_1, A_2)$

Unknown amplitude and initial phase

- Solution: $(\bar{A}.\hat{n})^2$ is maximized when the unit vector $\hat{n}$ points along vector $\bar{A} \Rightarrow$

$$X = \frac{\langle \bar{y}, \bar{q}_0 \rangle}{\sqrt{\langle \bar{y}, \bar{q}_0 \rangle^2 + \langle \bar{y}, \bar{q}_1 \rangle^2}}, \mathrm{Y} = \frac{\langle \bar{y}, \bar{q}_1 \rangle}{\sqrt{\langle \bar{y}, \bar{q}_0 \rangle^2 + \langle \bar{y}, \bar{q}_1 \rangle^2}}$$

$$\Rightarrow L_G = \max_{\Theta'} [\langle \bar{y}, \bar{q}_0(\Theta') \rangle^2 + \langle \bar{y}, \bar{q}_1(\Theta') \rangle^2]$$

- (From now, $\Theta' \rightarrow \Theta$ the set of parameters **besides amplitude and initial phase**)

$$\mathrm{L}_G = \max_{\Theta} [\langle \bar{y}, \bar{q}_0(\Theta) \rangle^2 + \langle \bar{y}, \bar{q}_1(\Theta) \rangle^2]$$

Monochromatic signal in WGN

- Special case: Monochromatic signal in WGN:
$$q(t;\Theta) \rightarrow q(t;\omega) = N(\omega)\sin(\omega t + \phi_0)$$
- N samples and Uniform sampling $\Rightarrow t_k = k\Delta, k = 0,1,\dots,N-1$
- WGN $(\sigma^2 = 1) \Rightarrow \langle \bar{x}, \bar{z} \rangle = \sum_{k=0}^{N-1} x_k z_k$ and $\|\bar{x}\|^2 = \langle \bar{x}, \bar{x} \rangle = \sum_{k=0}^{N-1} x_k^2$

$$\therefore \|\bar{q}(\omega)\|^2 = 1 \Rightarrow \langle \bar{q}(\omega), \bar{q}(\omega) \rangle = N^2(\omega) \sum_{k=0}^{N-1} \sin^2(\omega t_K)$$

$$\approx \frac{N^2(\omega)}{\Delta} \int_0^{T=(N-1)\Delta} dt\ \sin^2 \omega t \approx \frac{N^2(\omega)}{2\Delta} T = 1 \Rightarrow N(\omega) = \sqrt{\frac{2\Delta}{T}}$$

$\Rightarrow$ Normalization factor $N(\omega) = N_0$ is independent of ω

$$\Rightarrow q_{1,k}(\omega) - i q_{0,k}(\omega) = N_0(\cos \omega t_k - i \sin \omega t_k) = N_0 e^{-i\omega t_k} = N_0 e^{-i\omega k\Delta}$$

Monochromatic signal in WGN

- Now, $A^2 + B^2 = |A - i\,B|^2$

$$\Rightarrow \langle \bar{y}, \bar{q}_0(\omega)\rangle^2 + \langle \bar{y}, \bar{q}_1(\omega)\rangle^2 = |\langle \bar{y}, \bar{q}_1(\omega)\rangle - i\langle \bar{y}, \bar{q}_0(\omega)\rangle\,|^2$$

$$= \left|\underbrace{\langle \bar{y}, \bar{q}_1(\omega) - i\,\bar{q}_0(\omega)\rangle}_{\langle \bar{x},\bar{z}\rangle = \sum_{k=0}^{N-1} x_k z_k}\right|^2 = \left|\sum_{k=0}^{N-1} y_k \underbrace{\left(q_{1,k}(\omega) - i\,q_{0,k}(\omega)\right)}_{q_{1,k}(\omega) - i q_{0,k}(\omega) = N_0 e^{-i\omega k\Delta}}\right|^2$$

$$\mathrm{L}_G = \max_{\omega}\left|\sum_{k=0}^{N-1} y_k e^{-i\omega k\Delta}\right|^2 \text{(Ignoring overall constant factors)}$$

- Performing the search for the maximum over a regularly spaced grid in ω, given by $\omega_p = p \times \frac{2\pi}{N\Delta}$, means

$$\sum_{k=0}^{N-1} y_k e^{-i\omega k\Delta} = \sum_{k=0}^{N-1} y_k e^{-i2\pi pk/\mathrm{N}} \text{ which is the DFT}$$

- $\Rightarrow L_G$ for monochromatic sinusoid in WGN: Compute the magnitude of the DFT of the data and find the frequency with the largest peak