Recommended text:
D.S. Sivia & J. Skilling, "Data Analysis: A Bayesian Tutorial"

Bayesian Inference

Problem: Have some data $\underset{\sim}{s}[t] = \{s^0, s^1, \ldots\}$

- Data is noisy.
- $\boxed{\text{Assume}}$ $\underset{\sim}{s}[t] = \underset{\sim}{n}[t] + \underset{\sim}{h}[t]$
  - $\underset{\sim}{n}[t]$: noise
  - $\underset{\sim}{h}[t] = \{h(t; \underset{\sim}{\lambda})\}$ signal, given by model $h(t; \underset{\sim}{\lambda})$.

The signal depends on a set of parameters $\underset{\sim}{\lambda}$.

Ex: a GW from a binary, $\underset{\sim}{\lambda} = \{m_1, m_2, D_L, \ldots\}$

We want to know:

What is the probability the signal has params $\underset{\sim}{\lambda}'$, given our data & signal model:

$$p(\underbrace{\underset{\sim}{\lambda}'}_{\text{params.}} \mid \underbrace{\underset{\sim}{s}, h}_{\text{givens}}) = ?$$

Bayes Theorem:

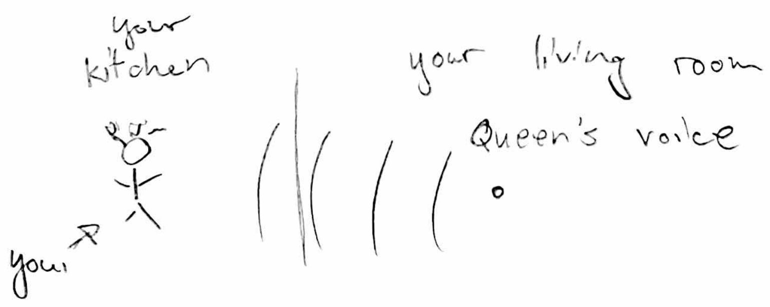
$$\underbrace{p(\underline{\lambda}'|\underline{s},h)}_{\text{"posterior"}} = \frac{\overbrace{p(\underline{s}|\underline{\lambda}',h)}^{\text{"likelihood"}}\,\overbrace{p(\underline{\lambda}'|h)}^{\text{"prior"}}}{\underbrace{p(\underline{s}|h)}_{\text{"evidence"}}}$$

- likelihood: prob. of observing the data assuming the signal has params $\underline{\lambda}'$

- prior: our prior belief that a signal can have parameters $\underline{\lambda}'$

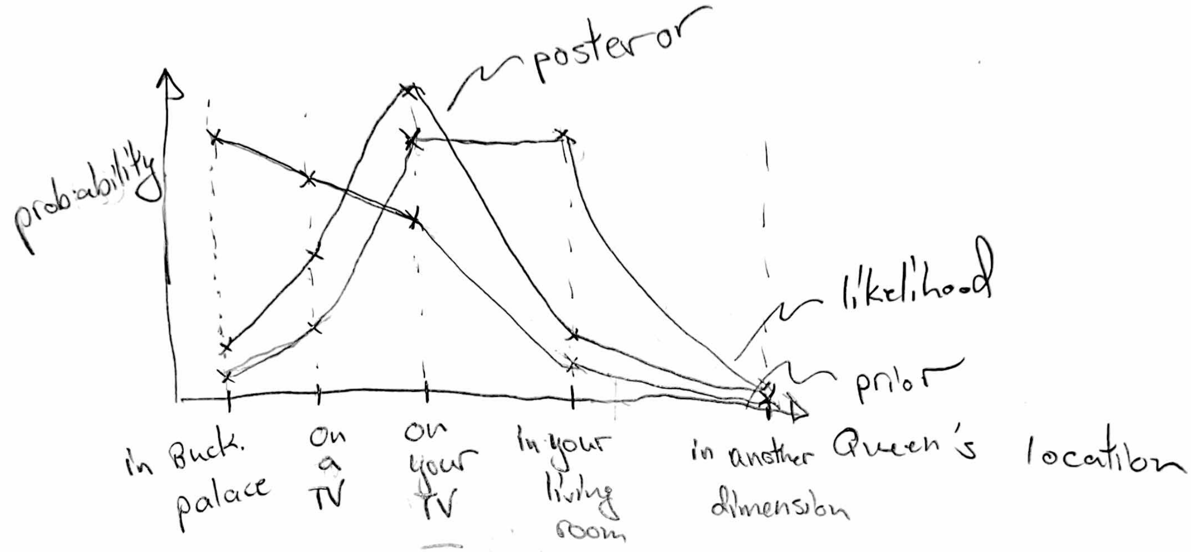
- evidence: normalization constant, used for comparing models.

Example:



Where is the Queen of England?

Given: you hear the Queen's voice in your living room


posterior
probability
likelihood
prior
in Buck. palace
On a TV
On your TV
In your living room
in another dimension
Queen's location

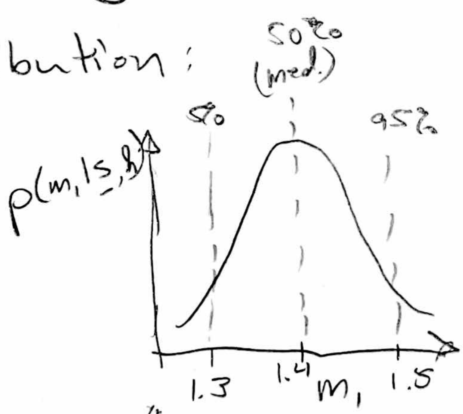
Marginalization

Full posterior prob. is multidimensional. Ex: for BBH, $\underline{\lambda}$ has 15 params $\Rightarrow$ 15 dimensional param. space.

Integrating posterior over a subset of params yields marginal:

$$\boxed{\begin{aligned} p(\{\lambda^0, \ldots, \lambda^k\} | \underline{s}, h) &= \int p(\underline{\lambda} | \underline{s}, h)\, d\lambda^{k+1} \ldots d\lambda^n \\ &= \frac{1}{p(\underline{s}|h)} \int p(\underline{s} | \underline{\lambda}, h)\, p(\underline{\lambda} | h)\, d\lambda^{k+1} \ldots d\lambda^n \end{aligned}}$$

Integrating over all but 1 param yields a 1D distribution:



Quote credible intervals using this

$\Rightarrow m_1 = 1.4^{+0.1}_{-0.1}\ M_\odot$

Evidence

Since post. is a pdf, marginalizing over all params:

$$\int p(\underline{\lambda} \mid \underline{s}, h)\, d\underline{\lambda} = 1$$

$$\Rightarrow \boxed{p(\underline{s} \mid h) = \int p(\underline{s} \mid \underline{\lambda}, h)\, p(\underline{\lambda} \mid h)\, d\underline{\lambda}}$$

Evidence is used for model selection. Say have 2 different signal models h_A, h_B.

Bayes Factor:

$$\boxed{\mathcal{B}(A, B \mid \underline{s}) = \frac{p(\underline{s} \mid h_A)}{p(\underline{s} \mid h_B)}}$$

$\log \mathcal{B} > 0 \rightarrow$ model A favored

$\log \mathcal{B} < 0 \rightarrow$ model B favored

$\log \mathcal{B} = 0 \rightarrow$ can't tell

Note: if only interested in posterior for a single model, sufficient to evaluate

$$p(\underline{\lambda} \mid \underline{s}, h) \propto p(\underline{s} \mid \underline{\lambda}, h)\, p(\underline{\lambda} \mid h)$$

In order to evaluate posterior, we need a signal model h <u>and</u> a noise model n.

- h comes from GR (... & other physics)
- What is the noise model?

Derivation of Likelihood

L.S. Finn, PRD 46, 5236-5249 (1992)

Wainstein & Zubakov, Extraction of signals from noise, 1962

Have network of K GW detectors (K = 3, HLV).
Sample their output at rate $1/\Delta t$ for time T,
get $N = \lfloor T/\Delta t \rfloor$ samples

$$\underline{s}_{Net} = \{ s_H^0, s_H^1, \dots, s_H^N, s_L^0, s_L^1, \dots, s_L^N, s_V^0, s_V^1, \dots s_V^N \}$$

Consider the hypothesis that the data contain no signal, $\underline{s} = \underline{n}$. What is the prob. of observing the given realization? $p(\underline{s}_{Net} | n) = ?$

$\boxed{\text{Assume}}$ noise is a stochastic Gaussian process

$$p(\underline{s}_{Net} | n) = \frac{\exp\left[-\frac{1}{2} \underline{s}_{Net}^T \underline{\underline{\Sigma}}_{net}^{-1} \underline{s}_{Net}\right]}{\left[(2\pi)^{NK} \det \underline{\underline{\Sigma}}_{net}\right]^{1/2}}$$

$\underline{\underline{\Sigma}}_{net}$ is the covariance matrix of the noise

Assuming 0 mean, is:

$$\underline{\underline{\Sigma}}_{net}[j,k] = \langle \underline{s}[j] \, \underline{s}[k] \rangle,$$ $\langle \cdots \rangle$ is an ensemble average

$$\underline{\Sigma}_{net} = \begin{pmatrix} \boxed{HH} & \boxed{HL} & \boxed{HV} \\ \boxed{\ldots} & \boxed{LL} & \boxed{LV} \\ \boxed{\ldots} & \boxed{\ldots} & \boxed{VV} \end{pmatrix}$$

Assume noise is uncorrelated between detectors

$\Rightarrow$ $\underline{\Sigma}_{net}$ block diagonal

$$p(s_{net} \mid n) = \frac{\exp\left[-\frac{1}{2} \sum_{d=1}^{K} \underline{s}_d^T \underline{\Sigma}_d^{-1} \underline{s}_d\right]}{\left[(2\pi)^{NK} \prod_{d=1}^{K} \det \Sigma_d\right]^{1/2}}$$

$\rightarrow$ **Need $\underline{\Sigma}^{-1}$** (will drop detector subscripts from here)

$$\Sigma[j,k] = \langle \underline{s}[j]\, \underline{s}[k] \rangle$$

$$= \langle \underline{s}[j]\, \underline{s}[\Delta_{kj} + j] \rangle, \qquad \Delta_{kj} \equiv k - j$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{l=0}^{n-1} \underline{s}^{(l)}[j]\, \underline{s}^{(l)}[\Delta_{kj} + j]$$

$\Rightarrow$ In general, cov. depends on both time considered $t_j = j \Delta t$ & displacement $\tau_{kj} = \Delta_{kj} \Delta t$

$\boxed{\text{Assume}}$ noise is Wide sense stationary (WSS)

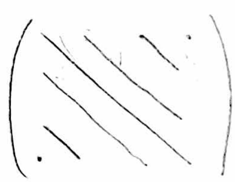
WSS: both mean & variance is constant

Then we can add any const. to both indices & get the same result:

$$\underline{\Sigma}[j+m, k+m] = \underline{\Sigma}[j, k]$$
$$= \langle \underline{s}[0], \underline{s}[\Delta_{kj}] \rangle$$
$$= \langle \underline{s}[-\Delta_{kj}] \, \underline{s}[0] \rangle$$

$\Rightarrow \underline{\Sigma}$ is: - symmetric

- diagonals are equal ("Toeplitz")
- elements are even functions of Δ_{kj}

$\underline{\Sigma} =$ 

=

Elements are ensemble averages.

$\boxed{\text{Assume}}$ data is ergodic $\Rightarrow$ get new realizations via time

Can replace ensemble average with time average:

$$\underline{\Sigma}[j,k] = \lim_{n\to\infty} \frac{1}{n} \sum_{\ell=0}^{n-1} \underline{s}^{(\ell)}[0]\, \underline{s}^{(\ell)}[\Delta_{kj}]$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{\ell=0}^{n-1} \underline{s}[\ell]\, \underline{s}[\Delta_{kj}+\ell]$$

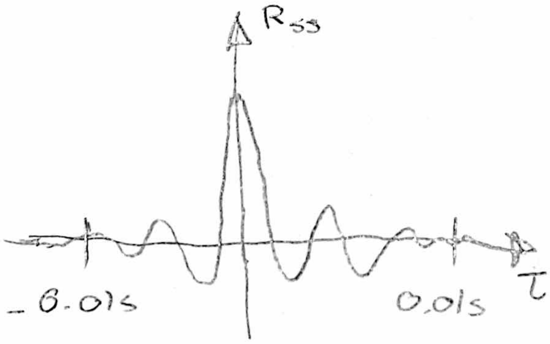
even func.

$$= \lim_{n\to\infty} \frac{1}{2n} \sum_{\ell=-n}^{n-1} \underline{s}[\ell]\, \underline{s}[\Delta_{kj}+\ell]$$

$$= \frac{1}{2} R_{ss}\left((k-j)\,\Delta t\right)$$

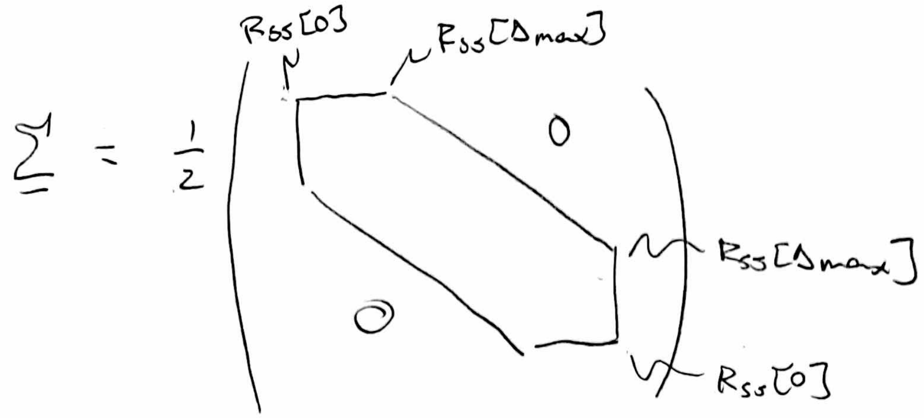
$R_{ss}(\tau)$ is the autocorrelation function of the data.

For GW detectors R_{ss} looks like:

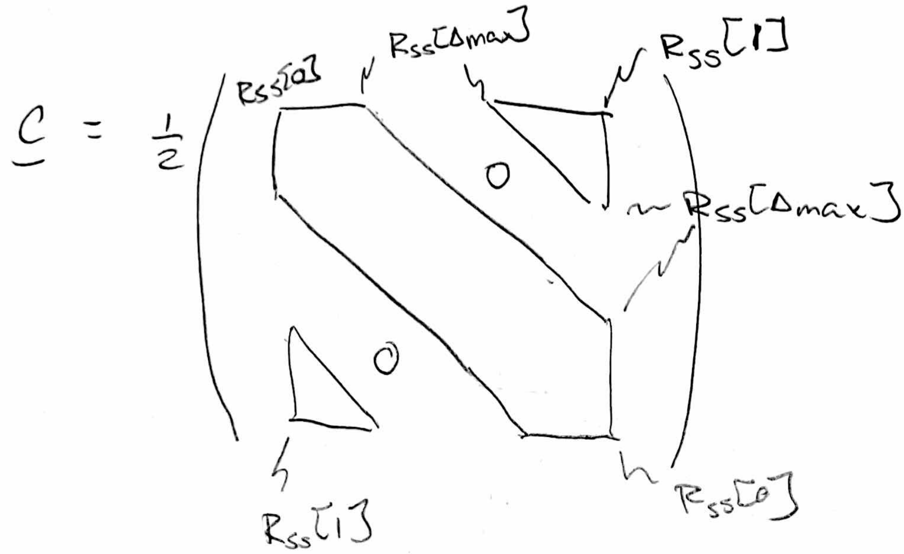


If $R_{ss}(\tau) \to 0$ in some finite time T_{max} then all diagonals with $|\Delta_{kj}| > \lfloor T_{max}/\Delta t \rfloor = \Delta_{max}$ $= 0$.

In this case, cov. matrix looks like:



This is similar to:



$\underline{C}$ is a <u>circulant</u> matrix; i.e., has form

$$\begin{pmatrix} a & b & c & d \\ b & a & b & c \\ c & b & a & b \\ d & c & b & a \end{pmatrix} \Rightarrow \underline{C}[j+N] = \underline{C}[j]$$

* All circulant matrices have the same eigenvectors:

$$\underline{u}_p[k] = \frac{1}{\sqrt{N}} e^{-2\pi i k p/N}$$

See R.M. Gray, "Toeplitz & Circulant Matrices: A Review"

Not true of Toeplitz matrices (like $\underline{\Sigma}'$), but b/c $\underline{\Sigma}'$ is a banded Toeplitz, it is asyntotically equivalent to $\underline{C}$, i.e.:

$$\lim_{N\to\infty} | \underline{\Sigma} - \underline{C} | = 0$$

This means: we can find eigenvalues of $\underline{\Sigma}'$ using

$$\underline{\Sigma}' \underline{u}_p \approx \lambda_p \underline{u}_p$$

as long as $\quad \Delta_{max} << N/2$

Solving, get:

$$\lambda_p = \frac{1}{2} \mathrm{Re}\left\{ \sum_{l=-N/2}^{N/2} R_{ss}[l] e^{-2\pi i p l/N} \right\}$$

R_{ss} even

$$= \frac{1}{2} \mathrm{Re}\left\{ \sum_{l=0}^{N-1} \cdots \right\}$$

$$= \frac{1}{2} \mathrm{Re}\left\{ \tilde{R}_{ss}[p] / \Delta t \right\}$$

The Re{ } arises b/c $\underline{\Sigma}$ is symm. & even.

where $\tilde{R}_{ss}$ is the discrete Fourier transform of R_{ss}:

$$\tilde{\underline{x}}[p] = \mathcal{F}\{x\}[p] = \Delta t \sum_{k=0}^{N-1} \underline{x}[k] e^{-2\pi i p k/N}$$

$$\underline{x}[k] = \mathcal{F}^{-1}\{\tilde{x}\}[k] = \Delta f \sum_{p=0}^{N-1} \tilde{\underline{x}}[p] e^{2\pi i p k/N}$$

$$\Delta f = \frac{1}{N\Delta t} = \frac{1}{T}$$

Wiener - Khinchin theorem: the F.T. of the autocorrelation function of a WSS stochastic process is equal to the power spectral density of the process $\mathcal{S}_n$:

$$\boxed{\mathcal{S}_n[p] = \Delta t \sum_{k=0}^{N-1} R_{ss}[k]\, e^{-2\pi i k p/N}}$$

$$\Rightarrow \boxed{\lambda_p = \frac{\mathcal{S}_n[p]}{2\Delta t}}$$

Construct matrix of eigenvec. $\underline{U}[k,p] = \frac{1}{\sqrt{N}} e^{-2\pi i k p/N}$

$\to \underline{U}$ is unitary $(U^\dagger U = \mathbb{1})$

& eigenval. $\underline{\Lambda}[\ell, p] = \lambda_p \delta_{\ell p}$.

to get:

$$\underline{\Sigma}^{-1}[j,k] = \frac{2\Delta t}{N} \sum_{p=0}^{N-1} \frac{e^{-2\pi i j p/N} e^{2\pi i k p/N}}{\mathcal{S}_n[p]}$$

b/c $\mathcal{S}_n$ symm around $N/2$

$$= 2\Delta f (\Delta t)^2 \sum_{p=0}^{N/2-1} \left(e^{-2\pi i (j-k)p/N} + e^{2\pi i (j-k)p/N}\right) \mathcal{S}_n^{-1}[p]$$

Note that b/c $\underline{s}$ is real,

$$(\Delta t)^2 \sum_{j,k=0}^{N-1} \underline{s}[j]\, \underline{s}[k] \left(e^{-2\pi i (j-k)p/N} + e^{2\pi i (j-k)p/N}\right) = 2|\tilde{\underline{s}}|^2[p]$$

We therefore have:

$$\underline{s}^T \underline{\underline{\Sigma}}^{-1} \underline{s} = 4 \Delta f \sum_{p=0}^{N/2-1} \frac{|\tilde{s}|^2[p]}{S_n[p]}$$

Note that if we define the inner product

$$\boxed{\langle \underline{a}, \underline{b} \rangle \equiv 4 \,\mathrm{Re}\left\{ \Delta f \sum_{p=0}^{N/2-1} \frac{\tilde{a}^*[p]\, \tilde{b}[p]}{S_n[p]} \right\}}$$

Then:

$$\underline{s}^T \underline{\underline{\Sigma}}^{-1} \underline{s} = \langle \underline{s}, \underline{s} \rangle$$

and:

$$\boxed{p(\underline{s}_{net} | n) \propto \exp\left[-\frac{1}{2} \sum_{d=1}^{K} \langle \underline{s}_d, \underline{s}_d \rangle \right]}$$

What about the signal hypothesis? $p(\underline{s} | \underline{\lambda}, h) = ?$

$$\underline{s}_{net} = \underline{n}_{net} + \underline{h}_{net}$$

$$\Rightarrow \underline{n}_{net} = \underline{s}_{net} - \underline{h}_{net}$$

$$\Rightarrow \boxed{p(\underline{s}_{net} | \underline{\lambda}, h) \propto \exp\left[-\frac{1}{2} \sum_{d} \langle \underline{s}_d - \underline{h}_d(\underline{\lambda}), \underline{s}_d - \underline{h}_d(\underline{\lambda}) \rangle \right]}$$

Signal-to-noise ratio (SNR)

Can't do full Bayesian inference on all times

- too computationally expensive
- glitches make significance estimation difficult

Instead a bank of templates is used to identify times when a signal exists. Calculate SNR for each template.

Likelihood ratio:

$$\mathcal{L}(\underline{\lambda}) \equiv \frac{p(\underline{s} \mid \underline{\lambda}, h)}{p(\underline{s} \mid n)}$$

$$= \frac{\exp\left[\sum_d \langle h_d(\underline{\lambda}), s_d\rangle - \frac{1}{2}\langle h_d(\underline{\lambda}), h_d(\underline{\lambda})\rangle \cancel{- \frac{1}{2}\langle s_d, s_d\rangle}\right]}{\cancel{\exp\left[-\frac{1}{2}\sum_d \langle s_d, s_d\rangle\right]}}$$

$$\Rightarrow \log \mathcal{L}(\underline{\lambda}) = \sum_d \langle h_d(\underline{\lambda}), s_d\rangle - \frac{1}{2}\langle h_d(\underline{\lambda}), h_d(\underline{\lambda})\rangle$$

Signals will have some unknown amplitude A & phase ϕ. Don't really care about these if we just want to know when a signal exists.

Amplitude: GW amplitude is inversely proportional to luminosity distance D. Generate a template $\underline{h}$ at some fiducial distance D_0 (usuall chosen to be 1 Mpc). Then for a signal $\underline{h}'$ at any distance,

$$\underline{h}' = \frac{D_0}{D}\underline{h} = A\underline{h} \quad \text{and:}$$

$$\log \mathcal{L} = A\langle \underline{h}, \underline{s}\rangle - \frac{1}{2}A^2\langle \underline{h}, \underline{h}\rangle$$

Note: we only consider a single detector here.

Maximizing over A:

$$\frac{d\log\mathcal{L}}{dA} = 0 \Rightarrow A_{max} = \frac{\langle \underline{h}, \underline{s}\rangle}{\langle \underline{h}, \underline{h}\rangle}$$

$$\Rightarrow \max_A \log\mathcal{L} = \frac{1}{2}\frac{\langle \underline{h}, \underline{s}\rangle^2}{\langle \underline{h}, \underline{h}\rangle} \equiv \frac{\rho^2}{2}, \quad \text{where } \boxed{\rho \equiv \frac{|\langle \underline{h}, \underline{s}\rangle|}{\sqrt{\langle \underline{h}, \underline{h}\rangle}}}$$

is the SNR.

Phase: Let $\underline{h} = \underline{h}_0\cos\phi + \underline{h}_{\pi/2}\sin\phi$, where

$$\underline{h}_{(0,\pi/2)}[t] \equiv (\mathrm{Re}, \mathrm{Im})\int_{-\infty}^{\infty}\hat{h}(f)e^{2\pi i f t}\,df$$

is the waveform with $\phi = 0$ & $\phi = \pi/2$, respectively. Then

$$\max_\phi \rho = \frac{\langle \underline{h}_0, \underline{s}\rangle\cos\phi + \langle \underline{h}_{\pi/2}, \underline{s}\rangle\sin\phi}{\sqrt{\langle \underline{h}_0, \underline{h}_0\rangle}}$$

Maximizing over ϕ gives:

$$\phi_{max} = \arctan\left(\frac{\langle \underline{h}_{\pi/2}, \underline{s} \rangle}{\langle \underline{h}_0, \underline{s} \rangle} \right)$$

Substituting back yields

$$\boxed{\max_{\phi} \rho = \left[\frac{\langle \underline{h}_0, \underline{s} \rangle^2 + \langle \underline{h}_{\pi/2}, \underline{s} \rangle^2}{\langle \underline{h}_0, \underline{h}_0 \rangle} \right]^{1/2}}$$

Note that if redefine the inner product as:

$$\langle\langle h, s \rangle\rangle = 4 \int \frac{\tilde{h}^*(f)\, \tilde{s}(f)}{S_n(f)} df \qquad \text{(i.e., let it be complex)}$$

Then:

$$\max_{\phi} \rho = \frac{|\langle\langle \underline{h}, \underline{s} \rangle\rangle|}{\sqrt{\langle\langle \underline{h}, \underline{h} \rangle\rangle}} \quad \& \quad \phi_{max} = \tan^{-1}\left(\frac{\mathrm{Im} \langle\langle \underline{h}, \underline{s} \rangle\rangle}{\mathrm{Re} \langle\langle \underline{h}, \underline{s} \rangle\rangle} \right)$$

=

Optimal SNR:

Since noise has zero mean, on average the SNR (assuming the template & signal are the same) will be:

$$\bar{\rho} = \frac{|\langle \underline{h}, \underline{h} + \bar{\underline{n}} \rangle|}{\sqrt{\langle \underline{h}, \underline{h} \rangle}} = \frac{|\langle \underline{h}, \underline{h} \rangle|}{\sqrt{\langle \underline{h}, \underline{h} \rangle}} .$$

$\Rightarrow$ $\boxed{\rho_{opt} = \sqrt{\langle \underline{h}, \underline{h} \rangle}}$ is the "optimal" SNR

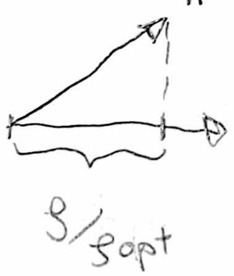
Overlap: Say the template $\underline{h}$ & signal $\underline{h}'$ are not the same – there's some mismatch. Then, on average the SNR will be:

$$\bar{\rho} = \frac{\langle \underline{h}, \underline{h}' \rangle}{\sqrt{\langle \underline{h}, \underline{h} \rangle}}$$

Relative to the optimal SNR for h', this is:

$$\boxed{\frac{\rho}{\rho_{opt}} = \frac{\langle \underline{h}, \underline{h}' \rangle}{\sqrt{\langle \underline{h}, \underline{h} \rangle \langle \underline{h}', \underline{h}' \rangle}} \equiv \text{"overlap"}}$$

$$\hat{h} \equiv \frac{h}{\sqrt{\langle h, h \rangle}}$$



$$\hat{h}' = \frac{h'}{\sqrt{\langle h', h' \rangle}}$$

Whitening: Can write the inner product as:

$$\langle a, b \rangle = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)}{\sqrt{S_n(f)}} \frac{\tilde{b}(f)}{\sqrt{S_n(f)}} \, df$$

$$a_w(t) = \int \frac{\tilde{a}^*(f)}{\sqrt{S_n(f)}} e^{2\pi i f t} \, df$$ is the "whitened" time series

$$\langle a, b \rangle = \mathcal{F}^{-1}\{\tilde{a}_w \cdot \tilde{b}_w\}(0) = a_w * b_w = \int a_w(\tau) b_w(\tau - 0) d\tau$$

convolution theorem

Time: Can do a trick to get SNR at all times τ.

$$\mathcal{F}\{h(t+\tau)\} = \int h(t+\tau)\, e^{-2\pi i f t}\, dt \qquad t' = t+\tau$$

$$dt' = dt$$

$$= \left(\int h(t') e^{-2\pi i f t'} dt'\right) e^{-2\pi i f \tau}$$

$$= \hat{h}(f)\, e^{-2\pi i f \tau}$$

Then, the SNR at some time offset τ is:

$$\rho(\tau) = \frac{|\langle h(t-\tau), s\rangle|}{\sqrt{\langle h,h\rangle}} = \frac{4}{\sqrt{\langle h,h\rangle}} \left| \int_{-\infty}^{\infty} \frac{\tilde{h}^*(f)\,\tilde{s}(f)}{S_n(f)}\, e^{2\pi i f \tau}\, df \right|$$

$$= \frac{4}{\sqrt{\langle h,h\rangle}} \left| \mathcal{F}^{-1}\{\tilde{z}\}(\tau) \right|, \qquad \tilde{z}(f) = \frac{\hat{h}^*(f)\,\tilde{s}(f)}{S_n(f)}$$

i.e can just take the inverse Fourier transform of $\tilde{z}$ to get SNR at all times.

$\Rightarrow$ $N \log N$ operations instead of N^2