Beyond Current Binary Black Hole simulations: What to do next?

Some quick, rough, and possibly misinformed considerations to guide our thinking now that binary black hole simulations are merely extremely difficult.

Motivation: Doing the best science with data and simulations!

What do we need to do/know in order to use these waveforms and this data in order to get out the science that we claim we can do?

Goal: A users' manual for understanding how to estimate when waveforms are "Accurate enough"

1. A rough, pedagogical derivation of the match, from the viewpoint of filtering data.

2. How the match appears in a maximum likelihood formalism and guides us to a "minimal match" criteria for waveforms to be "good."

3. Why the match is *not* a useful tool for estimating how accurate waveforms must be for science analyses ("parameter estimation").

Background references

Manual for GRASP (Gravitational Radiation Analysis and Simulation Package):

http://www.lsc-group.phys.uwm.edu/~ballen/grasp-distribution/

Textbook "Applications of Classical Physics," Blandford & Thorne, Chapter 5, "Theory of Random Processes"

http://www.pma.caltech.edu/Courses/ph136/yr2006/text.html

L. S. Finn, PRD **46**, 5236 (1992)
L. S. Finn and D. E. Chernoff, PRD **47**, 2198 (1993)

Tutorial: A derivation of the match

Begin: Consider data stream *s* containing noise *n* and a signal *h*:

$$s(t) = h(t) + n(t)$$

An "observation" of this data means that we cross correlate the stream with some filter *Q*. *For now*, leave this filter completely arbitrary:

$$S = \int_{-\infty}^{\infty} s(t) Q(t)\, dt$$

$$= \int_{-\infty}^{\infty} \tilde{s}(f) \tilde{Q}^*(f)\, df$$

Tutorial: A derivation of the match

Further progress requires some information about statistics of noise.

Zero mean:

$$\langle n \rangle = 0$$

Variance set by the spectral density:

$$\langle \tilde{n}(f)\tilde{n}^*(f') \rangle = \frac{1}{2} S_h(|f|)\delta(f - f')$$

Averaging over ensemble of all possible realizations of detector noise:

$$\langle f(n) \rangle = \int f(n) P(n)\, \mathcal{D}n$$

Tutorial: A derivation of the match

Mean value of the filtered datastream is now pretty obvious:

$$\langle S \rangle = \int_{-\infty}^{\infty} \tilde{h}(f)\tilde{Q}^*(f)\, df$$

Difference between filter output and mean filter output defines noise:

$$N = S - \langle S \rangle$$

Noise is characterized by its variance:

$$\langle N^2 \rangle = \frac{1}{2}\int_{-\infty}^{\infty} S_h(|f|)|\tilde{Q}(f)|^2 df$$

Tutorial: A derivation of the match

Signal to noise ratio for this general filter *Q* is now defined as the ratio of mean signal output to rms noise:

$$\left(\frac{S}{N}\right)^2 = \frac{\langle S\rangle^2}{\langle N^2\rangle}$$

To facilitate further manipulations, very useful to define the following inner product:

$$(A, B) = \int_{-\infty}^{\infty} A(f) B^*(f) S_h(|f|) df$$

Tutorial: A derivation of the match

With this definition,

$$\langle S \rangle = (\tilde{h}/S_h, \tilde{Q})$$
$$\langle N^2 \rangle = \frac{1}{2}(\tilde{Q}, \tilde{Q})$$

which allows us to write the SNR as

$$\left(\frac{S}{N}\right)^2 = \frac{2(\tilde{h}/S_h, \tilde{Q})^2}{(\tilde{Q}, \tilde{Q})}$$

Key question: **Given that the data contains signal *h*, what is the filter *Q* that maximizes the SNR?**

Tutorial: A derivation of the match

Answer: Schwartz inequality!

$$(A, B)^2 \le (A, A)(B, B)$$

with equality for *A* ∝ *B*.

So, we choose our filter to *match* the signal, modulo some weighting with noise,

$$\tilde{Q}(f) = \tilde{h}(f)/S_h(f)$$

yielding the "optimal" SNR:

$$\left(\frac{S}{N}\right)^2 = 2\left(\frac{\tilde{h}}{S_h}, \frac{\tilde{h}}{S_h}\right) = 2\int_{-\infty}^{\infty} \frac{|\tilde{h}(f)|^2}{S_h(f)} df$$

Tutorial: A derivation of the match

Back up: Choose filter from family of "templates" T_i that we hope faithfully represents the signal. Make the following definitions:

$$\tilde{Q}(f) = \frac{2\tilde{T}_i(f)}{S_h} \qquad \left(\frac{\tilde{T}_i}{S_h}, \frac{\tilde{T}_i}{S_h}\right) = \frac{1}{2} \qquad \text{(so } \langle N^2 \rangle = 1\text{)}$$

Then,

$$\rho_i \equiv \left(\frac{S}{N}\right)_{\text{Measuring } h \text{ with } T_i} = 2\left(\frac{\tilde{h}}{S_h}, \frac{\tilde{T}_i}{S_h}\right)$$

$$= 2\int_{-\infty}^{\infty} \frac{\tilde{h}(f)\tilde{T}_i^*(f)}{S_h(f)} df \equiv \langle h | T_i \rangle$$

Finally: The match!

The "SNR" between *two templates* in our bank defines the "match" between those two:

$$\mu_{ij} = \langle T_i | T_j \rangle \leq 1$$

Match defines fraction of optimal SNR we *retain* if:

1. We search for a signal proportional to template *i* using template *j*,
2. The signals that nature provides are faithfully represented by this template bank:

$$\left(\frac{S}{N}\right)_{\mathrm{achieved}} = \langle h | T_j \rangle = \mu_{ij} \left(\frac{S}{N}\right)_{\mathrm{optimal}}$$

More fundamental viewpoint on measurement

Starting point is the *likelihood function*:
The fundamental probability distribution that describes how likely it is that model fits our data.

$$\Lambda[h(\vec{\theta})] = p(\vec{\theta})P[s|h(\vec{\theta})]/P(s|0)$$

Prior probability that *h* is described by "parameters" θ

More fundamental viewpoint on measurement

Starting point is the *likelihood function*:
The fundamental probability distribution that describes how likely it is that model fits our data.

$$\Lambda[h(\vec{\theta})] = p(\vec{\theta}) P[s|h(\vec{\theta})]/P(s|0)$$

Prior probability that *h* is described by "parameters" θ

CAUTION: Easiest way to introduce bias or otherwise shoot yourself in the foot is by assuming a bad prior! Best not to assume very much unless you have a data-motivated reason for doing so.

More fundamental viewpoint on measurement

Starting point is the *likelihood function*:
The fundamental probability distribution that describes how likely it is that model fits our data.

$$\Lambda[h(\vec{\theta})] = p(\vec{\theta}) P[s|h(\vec{\theta})]/P(s|0)$$

Probability of measuring data *s* assuming that waveform *h* with parameters θ is present.

Probability of measuring data *s* assuming that *only noise* is present.

Details on probability

Given the noise statistics, key probability distributions here take a simple form:

$$P(s|0) = \exp\left(-\frac{1}{2}\langle s|s\rangle\right)$$

This is the distribution for *s* to be pure noise! Note Gaussian form — a happy fantasyland.

$$P[s|h(\vec{\theta})] = P[s - h(\vec{\theta})|0]$$

Rigorous discussion begins with distributions, brings in statistics of noise, and then (via theory of random processes), derives statistic which leads to match as argument of these distributions: See Finn '92.

Detection

"Detection": you're stating that it is highly likely that ***a*** gravitational-wave is present in the data.

Key point: *You don't necessarily care about the details of that wave — just that it's there.*

Formally, marginalize over waveforms

$$\Lambda = \int \mathcal{D}h \, \Lambda[h]$$

$$= \int \mathcal{D}h \, p(h) P(s|h)/p(s|0)$$

and declare we've made "a detection" if Λ exceeds a threshold.

Detection

Operationally boils down to setting a threshold on the signal-to-noise ratio we achieve:

$$\Lambda_{\rm thresh} \to \left(\frac{S}{N}\right)_{\rm thresh}$$

$$\Lambda > \Lambda_{\rm thresh} \longrightarrow$$

$$\left(\frac{S}{N}\right) = \langle s|T_i\rangle \approx \langle h_{\rm true}|T_i\rangle = (1-\epsilon)\left(\frac{S}{N}\right)_{\rm opt}$$

$$> \left(\frac{S}{N}\right)_{\rm thresh}$$

As long as *some* template gives us SNR above threshold, we can have confidence in a detection.

Detection

Operationally boils down to setting a threshold on the signal-to-noise ratio we achieve:

$$\Lambda_{\rm thresh} \to \left(\frac{S}{N}\right)_{\rm thresh}$$

$$\Lambda > \Lambda_{\rm thresh} \longrightarrow$$

$$\left(\frac{S}{N}\right) = \langle s | T_i \rangle \approx \langle h_{\rm true} | T_i \rangle = (1 - \epsilon) \left(\frac{S}{N}\right)_{\rm opt}$$

$$> \left(\frac{S}{N}\right)_{\rm thresh}$$

Note: Signal range cut by at most ε; volume covered by measurement degraded by ~3ε.

Detection

Operationally boils down to setting a threshold on the signal-to-noise ratio we achieve:

$$\Lambda_{\rm thresh} \to \left(\frac{S}{N}\right)_{\rm thresh}$$

$$\Lambda > \Lambda_{\rm thresh} \longrightarrow$$

$$\left(\frac{S}{N}\right) = \langle s | T_i \rangle \approx \langle h_{\rm true} | T_i \rangle = (1 - \epsilon) \left(\frac{S}{N}\right)_{\rm opt}$$

$$> \left(\frac{S}{N}\right)_{\rm thresh}$$

Caveat: This is a discussion in Gaussian noise! Can't quite be so cavalier for the "real" case ...

Measurement

"Measurement": you've decided a signal is present. Now you want to characterize it as accurately as possible: Find the *h* that maximizes the likelihood.

$$\Lambda[h(\vec{\theta})] = p(\vec{\theta})P[s|h(\vec{\theta})]/P(s|0)$$
$$= p(\vec{\theta})\exp\left[-\frac{1}{2}\langle h(\vec{\theta})|h(\vec{\theta})\rangle + \langle s|h(\vec{\theta})\rangle\right]$$

For well characterized waveforms, this process means figuring out which parameters θ describe the waveforms in the data.

Measurement

"Measurement": you've decided a signal is present. Now you want to characterize it as accurately as possible: Find the *h* that maximizes the likelihood.

$$\Lambda[h(\vec{\theta})] = p(\vec{\theta}) P[s|h(\vec{\theta})]/P(s|0)$$
$$= p(\vec{\theta}) \exp\left[-\frac{1}{2}\langle h(\vec{\theta})|h(\vec{\theta})\rangle + \langle s|h(\vec{\theta})\rangle\right]$$

For well characterized waveforms, this process means figuring out which parameters θ describe the waveforms in the data.

Not all measurements are parameter estimation! Example: How do we use merger waveforms to "test general relativity"?

Parameter estimation

For measurements that *are* parameter estimation, find parameters that maximize the likelihood:

$$\left.\frac{\partial \Lambda(h)}{\partial \theta^i}\right|_{\theta^i = \hat{\theta}^i} = 0$$

The parameters $\hat{\theta}^i$ that maximize Λ are presumably close to the true value $\tilde{\theta}^i$ characterizing the signal in our data. How close? Examine distribution of δθ:

$$\delta\theta^i \equiv \hat{\theta}^i - \tilde{\theta}^i$$

Parameter estimation

We would like this error to be determined by noise. If it is, then

$$\delta\theta^i = \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h}{\partial\theta^j} \middle| n \right\rangle$$

where

$$\Gamma_{ij} = \left\langle \frac{\partial h}{\partial\theta^i} \middle| \frac{\partial h}{\partial\theta^j} \right\rangle$$

The distribution of this variable is given by

$$\langle\delta\theta^i\rangle = 0$$

$$\langle\delta\theta^i\delta\theta^j\rangle = \left(\Gamma^{-1}\right)^{ij}$$

Parameter estimation

The result $$\delta\theta^i = \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h}{\partial \theta^j} \middle| n \right\rangle$$

assumes we have a template that *exactly* matches the waveform in our data. A better measure of this error is

$$\delta\theta^i = \left(\Gamma^{-1}\right)^{ij} \left(\left\langle \frac{\partial h_{\rm T}}{\partial \theta^j} \middle| n \right\rangle + \left\langle \frac{\partial h_{\rm T}}{\partial \theta^j} \middle| h_{\rm N} - h_{\rm T} \right\rangle \right)$$

$$= \delta\theta^i_{\rm stat} + \delta\theta^i_{\rm sys}$$

where $h_{\rm N}$ is nature's waveform, $h_{\rm T}$ is our template.

Accuracy goal: Systematic errors smaller than statistical errors.

$$\left\langle \left(\delta\theta^i_{\rm stat}\right)^2 \right\rangle^{1/2} > \delta\theta^i_{\rm sys}$$

$$\left[\left(\Gamma^{-1}\right)^{ii}\right]^{1/2} > \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h}{\partial \theta^j} \middle| \Delta h \right\rangle$$

Examine how different terms in this relation scale with SNR and phase error:

$$\left(\Gamma^{-1}\right)^{ij} \propto (1/{\rm SNR})^2$$

$$\left\langle \frac{\partial h}{\partial \theta^j} \middle| \Delta h \right\rangle \propto ({\rm SNR})^2 \Delta\Phi$$

Accuracy goal: Systematic errors smaller than statistical errors.

So the parameter accuracy requirement

$$\left\langle \left(\delta\theta^i_{\rm stat}\right)^2 \right\rangle^{1/2} > \delta\theta^i_{\rm sys}$$

Turns into a waveform requirement

$$\Delta\Phi < 1/\mathrm{SNR}$$

Bear in mind this is a *very* crude rule of thumb! To get it, neglected parameter correlations, only considered SNR scaling of waveform and matrices.

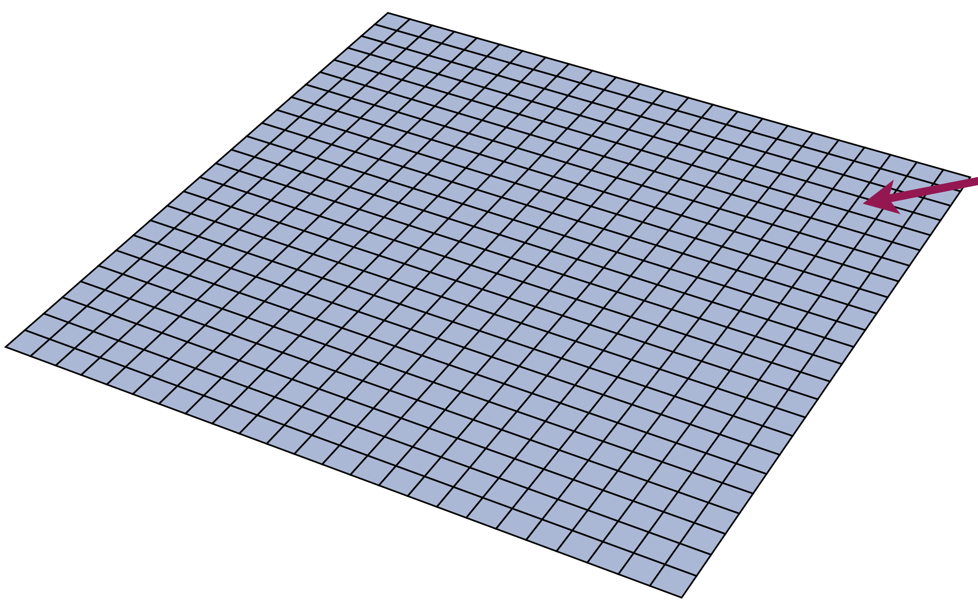
Probably too optimistic:

Real requirements more stringent.

What happens if this accuracy is not achieved/achievable?

That is, what if the waveforms we use to do our parameter estimate do not faithfully represent the signals that nature puts in our detectors?

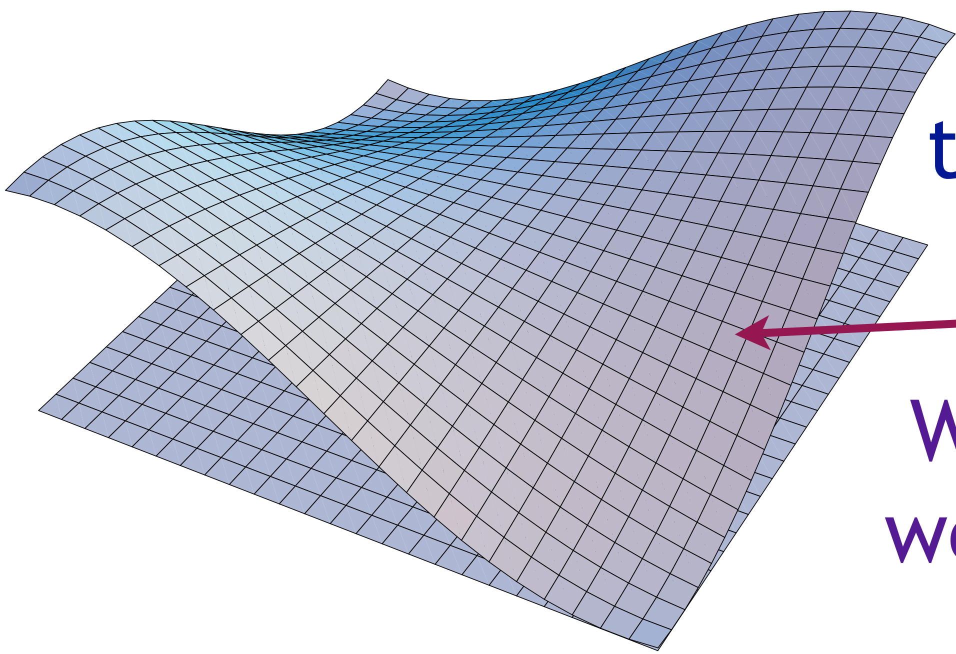
Akin to using templates that live in this manifold



What happens if this accuracy is not achieved/achievable?

That is, what if the waveforms we use to do our parameter estimate do not faithfully represent the signals that nature puts in our detectors?

Akin to using templates that live in this manifold



to measure waveforms that live in this manifold.

We can make the detection, we'll estimate parameters ... but that estimate will be systematically biased.

Bias

The bias due to displacement of these manifolds:

$$\delta\theta^i_{\rm bias} = \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h_{\rm T}}{\partial\theta^j} \middle| h_{\rm N} - h_{\rm T} \right\rangle$$

(Details, formalism, and numerical results, evaluated for LISA: Cutler and Vallisneri, arXiv:0707.2982. Numerical results obtained by using pN order *N* as a model for extracting signal of pN order *N* + 1.)

A key issue is that all of the parameter errors are *correlated* — sometimes highly. Bias in mass or spin can strongly impact our inferred values of other parameters (e.g., sky position, distance to event).

A (bad) example covariance matrix

Table 2. Full errors and correlations

	$\ln D_L$	$\cos\bar\theta_L$	$\cos\bar\theta_N$	$\bar\phi_L$	$\bar\phi_N$	t_c	Φ_c	$\ln\mathcal{M}$	$\ln\eta$	β	σ
$\ln D_L$	0.233	−0.984	0.878	0.509	0.213	0.0801	0.246	0.227	−0.186	0.205	−0.106
$\cos\bar\theta_L$	·	0.467	−0.861	−0.350	−0.071	−0.040	−0.098	−0.178	0.138	−0.156	0.0622
$\cos\bar\theta_N$	·	·	0.0006	0.465	0.203	0.0709	0.231	0.201	−0.166	0.181	−0.095
$\bar\phi_L$	·	·	·	0.687	0.782	0.232	0.827	0.337	−0.317	0.328	−0.259
$\bar\phi_N$	·	·	·	·	0.0017	0.193	0.691	0.252	−0.244	0.250	−0.210
t_c	·	·	·	·	·	63.1	0.705	0.923	−0.955	0.942	−0.993
Φ_c	·	·	·	·	·	·	4.00	0.742	−0.747	0.748	−0.726
$\ln\mathcal{M}$	·	·	·	·	·	·	·	0.0010	−0.995	0.998	−0.956
$\ln\eta$	·	·	·	·	·	·	·	·	0.303	−0.999	0.981
β	·	·	·	·	·	·	·	·	·	1.11	−0.971
σ	·	·	·	·	·	·	·	·	·	·	0.722

No spin precession; LISA noise curve used for match; m1 = 3 x 10^6 Msun, m2 = 10^6 Msun; restricted 2PN.

Diagonal entries: rms $\delta\theta$

Off-diagonal entries: Correlations.

A (bad) example covariance matrix

Table 2. Full errors and correlations

	$\ln D_L$	$\cos\bar\theta_L$	$\cos\bar\theta_N$	$\bar\phi_L$	$\bar\phi_N$	t_c	Φ_c	$\ln\mathcal{M}$	$\ln\eta$	β	σ
$\ln D_L$	0.233	−0.984	0.878	0.509	0.213	0.0801	0.246	0.227	−0.186	0.205	−0.106
$\cos\bar\theta_L$	·	0.467	−0.861	−0.350	−0.071	−0.040	−0.098	−0.178	0.138	−0.156	0.0622
$\cos\bar\theta_N$	·	·	0.0006	0.465	0.203	0.0709	0.231	0.201	−0.166	0.181	−0.095
$\bar\phi_L$	·	·	·	0.687	0.782	0.232	0.827	0.337	−0.317	0.328	−0.259
$\bar\phi_N$	·	·	·	·	0.0017	0.193	0.691	0.252	−0.244	0.250	−0.210
t_c	·	·	·	·	·	63.1	0.705	0.923	−0.955	0.942	−0.993
Φ_c	·	·	·	·	·	·	4.00	0.742	−0.747	0.748	−0.726
$\ln\mathcal{M}$	·	·	·	·	·	·	·	0.0010	−0.995	0.998	−0.956
$\ln\eta$	·	·	·	·	·	·	·	·	0.303	−0.999	0.981
β	·	·	·	·	·	·	·	·	·	1.11	−0.971
σ	·	·	·	·	·	·	·	·	·	·	0.722

Message: Correlations are large enough that a bias in one parameter is going to skew other parameters as well!

Match is not enough!

Key finding of Cutler and Vallisneri:

There can be significant bias introduced by a "bad" waveform model even if the match is extremely good.

Example they considered:

- Model the true waveform as 3.5PN
- Model the template as 3PN

Compare statistical and systematic errors. Also compute the match to see whether a bad result would have been "obvious."

Example

Chirp mass error:

$\delta\theta_{stat} = 10^{-5}$ $\delta\theta_{sys} = 3\mathrm{x}10^{-4}$

Bias by a factor of 30.

Reduced mass error:

$\delta\theta_{stat} = 2\mathrm{x}10^{-3}$ $\delta\theta_{sys} = 0.14$

Bias by a factor of 70.

Sky position:

$\delta\theta_{stat} \approx 1$ degree $\delta\theta_{sys} \approx 1.5$ degree

Inferred position is outside the “true” error box.

Example

Chirp mass error:

$\delta\theta_{stat} = 10^{-5}$ $\delta\theta_{sys} = 3\mathrm{x}10^{-4}$

Bias by a factor of 30.

Reduced mass error:

$\delta\theta_{stat} = 2\mathrm{x}10^{-3}$ $\delta\theta_{sys} = 0.14$

Bias by a factor of 70.

Sky position:

$\delta\theta_{stat} \approx 1$ degree $\delta\theta_{sys} \approx 1.5$ degree

Inferred position is outside the "true" error box.

Match = 0.9999

Why??

This counterintuitive result can be understood by comparing match to the parameter bias:

$$\mathrm{match} = \frac{\langle h_T | h_N \rangle}{\sqrt{\langle h_T | h_T \rangle \langle h_N | h_N \rangle}}$$

$$\delta\theta^i_{\rm bias} = \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h_{\rm T}}{\partial \theta^j} \middle| h_{\rm N} - h_{\rm T} \right\rangle$$

The difference is largely due to derivatives: If waveform is very sensitive to a parameter, the large derivative can compensate for the fact that $h_{\rm N}$ and $h_{\rm T}$ are nearly equal.

Why??

This counterintuitive result can be understood by comparing match to the parameter bias:

$$\text{match} = \frac{\langle h_T | h_N \rangle}{\sqrt{\langle h_T | h_T \rangle \langle h_N | h_N \rangle}}$$

$$\delta\theta^i_{\text{bias}} = \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h_{\mathrm{T}}}{\partial \theta^j} \middle| h_{\mathrm{N}} - h_{\mathrm{T}} \right\rangle$$

Correlations in the covariance matrix Γ_{ij} also play a huge role! May have severe consequences when this analysis is done for LIGO measurements.

Punchline:

Use the match to assess whether waveforms are useful for *detection* purposes:

$$\langle T_i | T_j \rangle = 1 - \epsilon$$

Tells us about SNR loss — just right to assess how useful templates are for a GW search.

Punchline:

Need something much more sophisticated and much more complicated to assess whether waveforms are useful for parameter estimation!

Goal is

$$\left\langle \left(\delta\theta^i_{\rm stat}\right)^2 \right\rangle^{1/2} > \delta\theta^i_{\rm sys}$$

$$\left[\left(\Gamma^{-1}\right)^{ii}\right]^{1/2} > \left(\Gamma^{-1}\right)^{ij} \left\langle \frac{\partial h}{\partial \theta^j} \middle| \Delta h \right\rangle$$

There's a **lot** more work to do ...