

Chapter 6

LISA Data Disturbances

As mentioned in §2.4 the LISA environment is not free from complications. Solar flares, cosmic rays, and rotations of the communications antenna disturb the proof mass. In particular, data during the disturbance may be lost. This is called a “data gap.” In addition, at the end of the disturbance period, the proof mass may have a new position, a new velocity, and a new acceleration relative to the distant proof mass defining the other end of the measured arm of the interferometer. These new parameters will be referred to as the “disturbance parameters.”

In this chapter we demonstrate a fairly simple method for locating a gravitational wave signal in the LISA data stream in the presence of data disturbances, and reducing the effects these disturbances have on the scientific results.

The aim of this chapter is to determine some of the effects data disturbances have on signal extraction from the LISA data stream. For a simple case, we wish to find information on the frequency of data disturbances which would inhibit the extraction of science from the data. Since data disturbances inhibit science, operational planning will require some sort of baseline information on the effects data disturbances impose.

Ground-based gravitational wave detectors usually will not be affected by brief disturbances which can be fit out from the data. However, both ground-based and space-based detectors will be affected by data gaps, even if no other disturbances are present. Preliminary analyses of these have been performed by Sylvestre and Armstrong [89] and by Cornish [11], and our results agree when we set the disturbance parameters equal to zero.

In §6.1 we describe the construction of our simulated LISA data stream; in §6.3 we describe our search algorithm and the methods used for locating the signals and disturbances; in §6.4 we apply our search algorithm to our simulated LISA data stream; in §6.5 we analyze the effect of multiple disturbances on the science results.

6.1 Simulated LISA Data Stream

Our simulated LISA data stream consists of noise, a monochromatic signal, and a disturbance. This section describes each of the three components in detail.

6.1.1 Noise Spectrum Realization

We use a suggestion for the LISA sensitivity curve described in [5] which extends down to $3\ \mu\text{Hz}$. We further extend this curve to $0.3\ \mu\text{Hz}$. Our noise realization also includes an approximation for the binary confusion noise present in our galaxy [28, 7, 30]. Above $3\ \text{mHz}$ the sensitivity is determined mainly by white noise in measuring the distances between the proof masses, and is dominated by shot noise in the LISA lasers. The spectrum from $3\ \text{mHz}$ down to $0.1\ \text{mHz}$ is dominated by the binary confusion noise. This part of the spectrum rises steeply at first and then roughly as $1/f$ from 1.6 to $0.1\ \text{mHz}$. Below $0.1\ \text{mHz}$ the spectrum is dominated by acceleration noise. Thermal and proof mass charge fluctuations cause the noise level to rise with a power law exponent of $-5/2$. For frequencies below $10\ \mu\text{Hz}$ the noise due to these fluctuations is expected to become worse, and another power of \sqrt{f} is added. It is assumed here that the spectrum continues to rise as $1/f^3$ down to $0.3\ \mu\text{Hz}$, the lowest frequency observable in this study.

We create a realization of the LISA sensitivity curve in frequency space by generating gaussian deviates to multiply each amplitude. Inverse Fourier transforming the spectral amplitude gives us a time series. Our time series contains 2^{15} points sampled at 100 second intervals, making the observation time just under 38 days. Throughout this chapter t_i is the time stamp at measurement i , and y_i is the simulated LISA data stream with disturbance and signal at t_i .

We produced 35 different realizations of the noise in the LISA data stream to gain confidence in our search algorithm. One of these realizations was selected at random for use in performing our analysis. We did most of our analysis on only one realization for simplicity in presenting our results. If not otherwise mentioned, all results refer to this realization.

6.1.2 Signal and Disturbance

For this analysis we have chosen to look at single monochromatic sinusoids with constant amplitude. The actual LISA signals will be considerably more complex. For instance, there may be a large number of separable signals below $100\ \mu\text{Hz}$, and many may change their frequencies considerably in one month. However, information for the single signal case will give some indication of what is likely to happen in more complicated cases.

We analyze two single signal cases. The first signal has a frequency of $3 \mu\text{Hz}$, which is near the “limit” of the proposed LISA sensitivity [5]. This signal contains nearly 10 complete cycles within our 38 day data set. In view of the low frequency, this signal should give us strong indications of limitations on the lengths and frequency of data disturbances.

The second case has a signal at $100 \mu\text{Hz}$. As the LISA constellation orbits the Sun, gravitational wave signals will be modulated in amplitude and frequency. However, the wavelength corresponding to $100 \mu\text{Hz}$ is 20 AU, so this signal will have minimal spread in frequency space. Therefore we are justified in approximating a signal at this frequency as monochromatic.

Our signals are characterized by an amplitude, a frequency, and a phase. We represent the signal symbolically as a function $h(t_i; A, f, \phi)$. We define a disturbance as an interruption in the data stream where no data is present, also known as a data gap, and where at the end of the interruption the proof mass may have a new position, velocity, and acceleration. The interruption has a duration of length L . The disturbance is characterized by a starting time, an ending time, and three parameters $[p, u, a]$ which are the additional position, velocity, and acceleration of the proof mass. We represent the disturbance symbolically as a function $g(t_i; p, u, a)$.

Figures 6.1 and 6.2 show one realization of the LISA data stream with a monochromatic signal at $3 \mu\text{Hz}$ and a disturbance located at 11.6 days after the beginning of the data set.

6.2 A Simple Guess

Looking at the frequency spectrum of the raw data, Figure 6.2, we should be able to guess the amplitude and frequency of the signal. The input signal parameters are $(A, f, \phi)_{\text{input}} = (2 \times 10^{-17}, 3 \mu\text{Hz}, 0^\circ)$. The frequency bin with the maximum power is $3.052 \mu\text{Hz}$. The width of each bin is set by the integration time and is $0.3052 \mu\text{Hz}$. Therefore a naïve determination of the frequency yields a value with a 1.7% error and a 10% uncertainty. The value of the power in that frequency bin gives an amplitude for the sinusoid of 1.75×10^{-17} with an estimated uncertainty of about 0.5×10^{-17} , which is 29%. The ratio of the two complex components of the FFT at that peak determines the phase of the signal to be 143.7° .

By comparison, our algorithm, which removes the disturbance, attains less than 0.3% error in the determination of the frequency, the amplitude better than 2%, and the phase to within 5° (see §6.4.1).

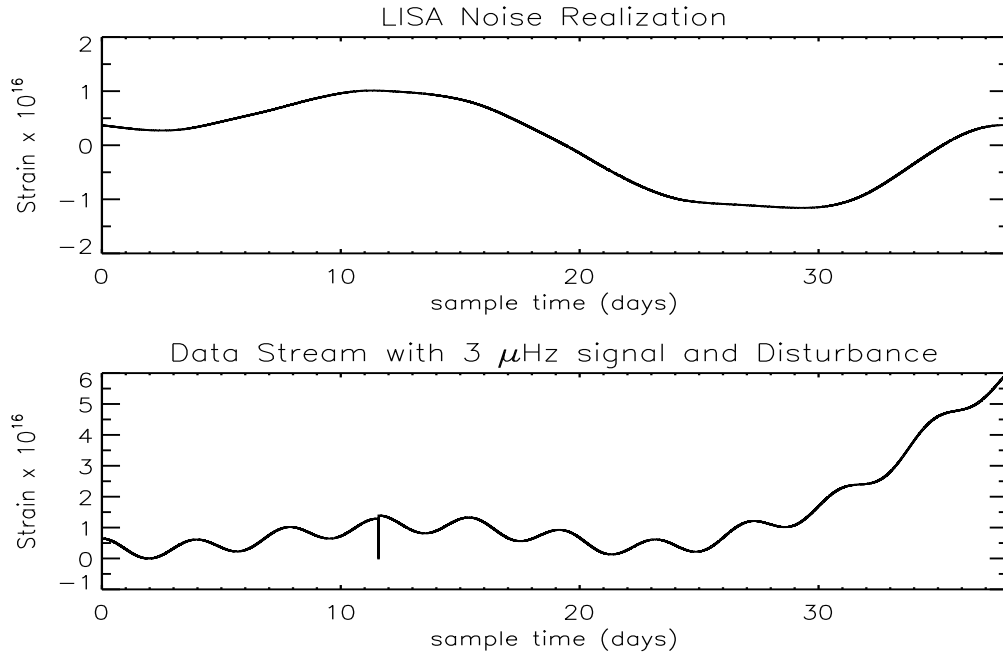


Figure 6.1: **Top panel:** One realization of the LISA noise spectrum in the time domain. **Bottom panel:** Added to the noise is a monochromatic sinusoid with $(A, f, \phi) = (2 \times 10^{-17}, 3 \mu\text{Hz}, 0^\circ)$, and a disturbance with $[p, u, a] = [10^{-17}, 10^{-23}, 10^{-28}]$ located at 11.6 days with an interruption length of $L = 17$ minutes. During this disturbance there is a data gap where no data is recorded and the strain values are zero.

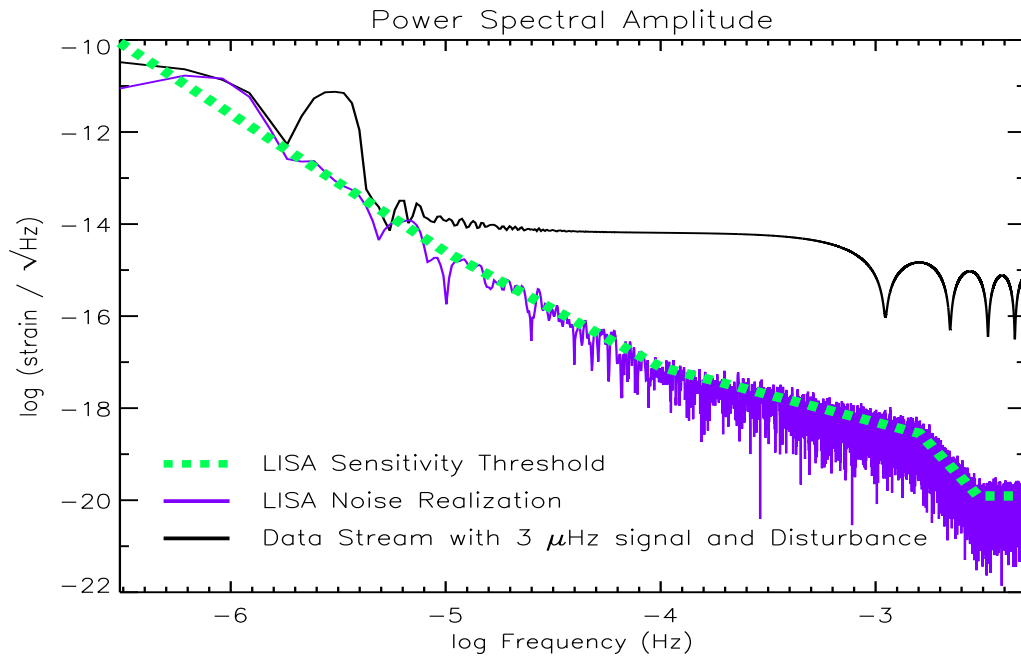


Figure 6.2: Power spectral amplitudes of the data shown in Figure 6.1. Our realization of the LISA noise includes the confusion noise background due to close white dwarf binaries in the galaxy. The $3 \mu\text{Hz}$ signal is clearly visible in this spectrum. This signal has a SNR of 5 as explained in §6.3.3. The sinc pattern at high frequencies as well as the flat response is due to the sharp edges of the data gap.

Doing these same naïve guesses of the parameters when no disturbance is present yields an amplitude of 1.92×10^{-17} and a phase of 1.34° , much closer to the input values. The presence of the disturbance has dramatically changed the derived value of the amplitude and phase of the signal.

The naïve determinations of the sinusoid parameters are lacking in precision and accuracy, and downstream scientific results, such as the masses, spins, and distances to supermassive black holes, will be hindered. These results imply that the disturbance should be fit for and removed from the data stream during data analysis.

6.3 Search Algorithm

We use an iterative approach for finding the signal and disturbance parameters. Our algorithm utilizes both the downhill simplex method of Nelder and Mead [61, 71] and the directional downhill descent method of Marquardt [71]. Given an initial collection of possible solutions to the search problem (generated randomly) we use the Nelder and Mead method to quickly find a local minimum of the problem.

As explained in [45] the Nelder and Mead method is very useful for producing a rapid initial decrease in estimator values in low parameter dimensions. But in higher dimensions the Nelder and Mead method fails to find minimized estimator values. However, the Nelder and Mead method can be forced to continue minimizing the estimator with prodding. Some of our prodding is randomly restarting our simplex centered about the previous solution. Other prodding involves modifying the estimator.

Once the Nelder and Mead method ceases to refine the solution we turn to the directional downhill descent method of Marquardt. This search method uses derivative information to minimize the estimator. Finishing with the Marquardt method guarantees a minimized estimator, typically with a value much less than that provided by the Nelder and Mead method.

There are many search algorithms in the literature. It is not our aim to characterize search algorithms, or to promote any one algorithm over another. We are not necessarily suggesting a methodology for LISA data analysis. Instead, we wish to characterize the effect data disturbances have on the extraction of science from the LISA data stream. We chose the Nelder and Mead method because of its simplicity and speed. Since we can constrain our parameter space fairly tightly initially (based on examination of the data in the time domain and the frequency domain) the Nelder and Mead method is suitable for finding the global minimum of our problem.

6.3.1 Evaluation Criteria

The search algorithm is dependent on an evaluation criterion which determines how likely a solution is to be the answer. Since we have correlated noise, we cannot determine an uncertainty associated with each measurement point in the time domain. But we do have some knowledge of the noise: we know its frequency spectrum. Therefore we can construct a χ^2 in the frequency domain with a diagonal correlation matrix:

$$\chi^2 = \sum_{i=0}^{N-1} \left[\frac{r(f_i)}{\sigma(f_i)} \right]^2, \quad (6.1)$$

where $r(f_i)$ is the Fourier transform of $y_i - h(t_i) - g(t_i)$, and $\sigma(f_i)$ is the LISA sensitivity threshold.

There are two modifications we make to our estimator during our search. The first is to window the data before taking the Fourier transform. The discrete Fourier transform of a sinusoid is a delta function in frequency space with large sidelobes dropping off as $1/f$. Using a modified Hanning window we suppress the sidelobes of the function, but we also spread the peak of the function out to cover more frequency bins. This modification helps the Nelder and Mead method to locate the frequency and amplitude of the sinusoid in our data stream. The windowing function is applied to all sharp edges in the data set, at the beginning, at the end, and at the two edges of the interruption. This windowing has been done for the spectrum shown in Figure 6.2. Without this windowing the spectral amplitude is dominated by the sinc function response of the data gap.

After the Nelder and Mead method has found modest parameter values we no longer window the data. At this point the disturbance has been mostly removed, and we fit across the data gap a quadratic function which matches the data before and after the gap. There is now no sinc function response from the data gap and strong windowing is not necessary. Without windowing the frequency of the signal is more readily attainable.

The second modification we make to our estimator is to weight the residuals. Our search algorithm can minimize the χ^2 by lowering the frequency of the sine wave, which has the effect of reducing the transform lobes at high frequencies. Since there are many more points at high frequencies than there are at low frequencies the value of χ^2 will be reduced, but there will be extra residuals at low frequencies, since the sine wave frequency is not being fit appropriately. To force our algorithm to fit the sine wave frequency we weight the residuals at low frequencies more than the residuals at high frequencies. We justify this weighting by the fact that we know we are looking for a low frequency sine wave, so we preferentially want to reduce the χ^2 contribution from the low frequency residuals.

In χ^2 -tests a reduced χ^2 of 1 indicates a good fit. We have 32768-point data sets, with 6 parameters to be fit, which yields 32762 degrees of freedom. Any value of χ^2 around this value is

considered a good fit. But how good is good? We would like some supercriterion to determine how well the search algorithm is actually performing. The goal, of course, is to find out how well the algorithm does in removing the disturbance so that the disturbance does not affect the science results. Our supercriterion compares the algorithm's fit for the signal, in the presence of both noise and a disturbance, to the algorithm's fit for the signal in the presence of noise only:

$$\Delta^2 = \sum_{i=0}^{N-1} [h(t; A, f, \phi) - h(t; A, f, \phi)_{\text{nodis}}]^2. \quad (6.2)$$

In this way we can judge how well different search criteria and search algorithms remove the disturbance from different data streams.

6.3.2 Disturbance Identification and Removal

We know the location and duration of each disturbance with some certainty. Our search algorithm is fed the locations and durations of each disturbance and is then asked to remove those disturbances as best as possible. The bulk of our analysis had only one disturbance present in the data set. The disturbance function is a quadratic. It easily can be shown that the accuracy to which you can fit a quadratic is dependent on the number of points available. Therefore the location of the disturbance within our 38-day data set should affect our ability to accurately identify the disturbance parameters.

We examined three disturbance locations in detail. For the rest of this article we refer to the three disturbances as $g1$, $g2$, and $g3$. $g1$ has a disturbance start time of 84,000 seconds (0.97 days) after the beginning of the time series. $g2$ has a disturbance start time of 500,000 seconds (5.8 days) after the beginning of the time series. $g3$ has a disturbance start time of 1,000,000 seconds (11.6 days) after the beginning of the time series. Each of these disturbances had the following parameters $[p, u, a]_{\text{input}} = [10^{-17}, 10^{-23}, 10^{-28}]$, which are referred to as the input disturbance parameters. These values correspond to a position change of about 26 nm, a velocity change of about 26 fm/s, and an acceleration change of about 0.26 attometer/s²[†] of the proof mass. The length of the data interruption is ten data points, i.e., 1000 seconds, or about 17 minutes.

Table 6.1 shows the results of our search algorithm when no signal is present. The determination of the position is very good, and the derived velocity and acceleration make a smooth residual curve throughout the interruption region. Note that the algorithm has found a lower χ^2 than obtained by using the input parameters. Figure 6.3 contains our results for all 35 noise realizations. We can use the standard deviation of the derived parameters throughout the 35 noise realizations as an estimate

[†]1 attometer $\equiv 10^{-18}$ m

Table 6.1: Search results for the three disturbance locations examined with no signal present. In all three cases, the search algorithm has minimized χ^2 below that which the input parameters give. $p^* = (p - 10^{-17}) \times 10^{20}$, u is in multiples of 10^{-23} , a is in multiples of 10^{-28} .

	p^*	u	a	χ^2
input	0.000	1.000	1.000	31 364
$g1$	-0.142	0.660	1.011	29 985
$g2$	-0.045	0.935	1.002	30 971
$g3$	-0.135	0.727	1.012	30 232

of our error in any one derived parameter. The errors are roughly [0.01%, 18%, 0.67%]. Therefore we can conclude that our algorithm finds confident values for the position and acceleration, but the velocity may not be accurate. Note that in Figure 6.3 the acceleration is systematically higher than the input value, and that the velocity (by correlation is therefore) lower than the input value. This slight shift in the acceleration and velocity has allowed our algorithm to find a lower χ^2 value. The fact that we have found “incorrect” values for the disturbance parameters should not deter us since the uncertainties should not depend on the assumed values, and since our main goal is to see if and how this affects the science results.

6.3.3 Signal Identification and Removal

For our realization of the LISA data stream, a root-mean-square (rms) strain amplitude of 2×10^{-17} has a signal-to-noise ratio (SNR) of 5 for a $3 \mu\text{Hz}$ signal. We define the SNR using the frequency spectrum of the data set. The SNR is basically the amplitude of the sine wave divided by the rms noise level at the frequency of the sine wave:

$$\rho = 2 \frac{A \sqrt{T_{\text{obs}}}}{\sigma(f = 3 \mu\text{Hz})} \quad (6.3)$$

where $T_{\text{obs}} = 38$ days is the observation period of the data set. This is the same formula for the signal-to-noise ratio, ρ , found in the literature [67, 89].

We injected a signal with this amplitude, frequency, and a phase of zero degrees into our data stream, so that $(A, f, \phi)_{\text{input}} = (2 \times 10^{-17}, 3 \mu\text{Hz}, 0^\circ)$. Without a disturbance in the data stream, our search algorithm found these signal parameters $(A, f, \phi)_{\text{nodis}} = (2.045 \times 10^{-17}, 2.996 191 \mu\text{Hz}, 4.990^\circ)$ with $\chi_{\text{nodis}}^2 = 21 164$. These will be the parameters used in our calculation of our supercriterion Δ^2 . Using the input signal parameters we found $\chi_{\text{input}}^2 = 31 219$. Taking the sum of the squares of the difference between the solution and the input signal gives a value of 0.437×10^{-31} . The correlated noise in the data stream “pulls” the solution away from the input signal parameters. In effect, our search algorithm has been able to lower χ^2 by fitting some of the noise.

Figure 6.4 displays the results from our search algorithm for all 35 noise realizations in the case of a signal only, with no disturbance. This data shows how well our algorithm can find the 3 μHz signal in the noisy data. We located the amplitude to within 1%, the frequency is consistently low by about 2%, and the phase is correspondingly high by about 6 degrees. The systematic error in our derived frequency is a result of the correlated noise rising as $1/f^3$ at low frequencies. Our algorithm is able to achieve a lower χ^2 by fitting a slightly lower frequency sine wave, removing some of the noise in the process. The sine wave frequency is anti-correlated with the phase, so the phase is correspondingly high when the frequency is low. This systematic error should not deter us in this investigation because we wish to find the limitations that disturbances add to our identification of the science signal, not necessarily how well we can fit the sinusoid individually.

6.4 Effects on Science

We now examine the full LISA data stream with a monochromatic signal and a disturbance.

6.4.1 Lower Frequency Signal

A strong low frequency signal can alter the determined disturbance parameters. This is why we utilize random restarts of our simplex to prevent becoming caught in local minima. In this way we can better determine all six parameters of interest. Table 6.2 contains our search results for the signal and disturbance parameters for the three disturbance locations examined. Our search algorithm was again able to minimize χ^2 below that which the input parameters yield.

Comparing the disturbance values in this table with those in Table 6.1 we see that the derived positions and accelerations are very close, but the velocities are different. Comparing the results for all 35 noise realizations we see that in Figure 6.3 (which contains a disturbance but no signal) the velocity is systematically low. Whereas in Figure 6.5 (which contains both a disturbance and a signal) the

Table 6.2: Search algorithm results for the three disturbances examined, and with a low frequency signal. A is in multiples of 10^{-17} , $f^* = (f - 3 \mu\text{Hz})/\text{nHz}$, $p^* = (p - 10^{-17}) \times 10^{20}$, u is in multiples of 10^{-23} , a is in multiples of 10^{-28} , and Δ^2 is in multiples of 10^{-31} .

	A	f^*	ϕ	p^*	u	a	χ^2	Δ^2
input	2.000	0.000	0.000°	0.000	1.000	1.000	31 369	0.437
nodis	2.045	-3.809	4.990°				21 164	0.000
$g1$	1.995	-10.754	1.875°	0.190	1.264	0.996	21 449	2.443
$g2$	1.991	-14.507	-0.378°	0.012	1.037	1.009	22 088	6.196
$g3$	1.998	-10.561	-1.322°	-0.170	0.667	1.027	22 914	4.628

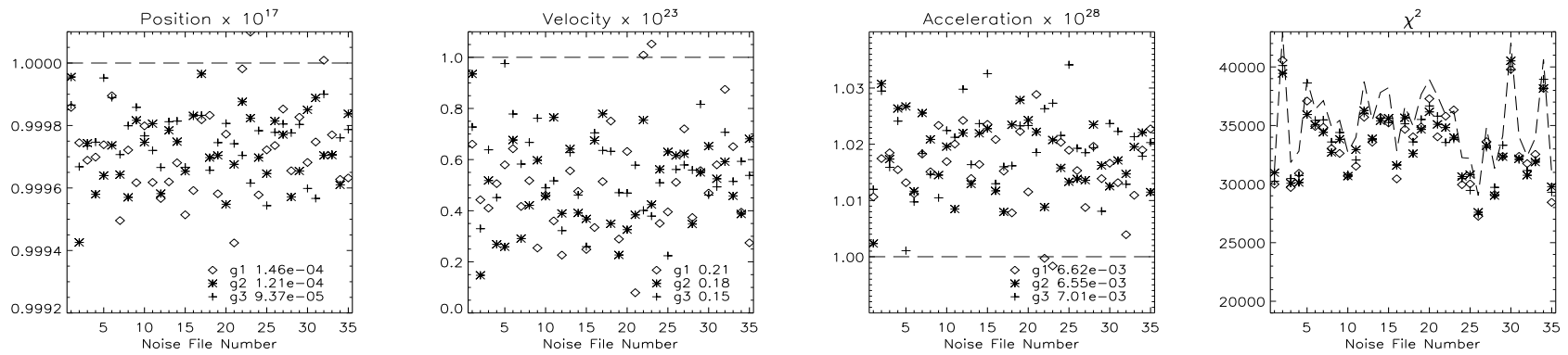


Figure 6.3: Search algorithm results for the three disturbance locations examined, in the 35 realizations of the LISA data stream. The dashed lines represent the input parameters. In the case of χ^2 , the dashed line is the value of χ^2 in the original LISA data stream realization. From the collection of 35 values for each parameter we compute the standard deviation. We can take this number as an indicator of the error in our derived quantity. This number is printed at the bottom of each plot, for each disturbance location.

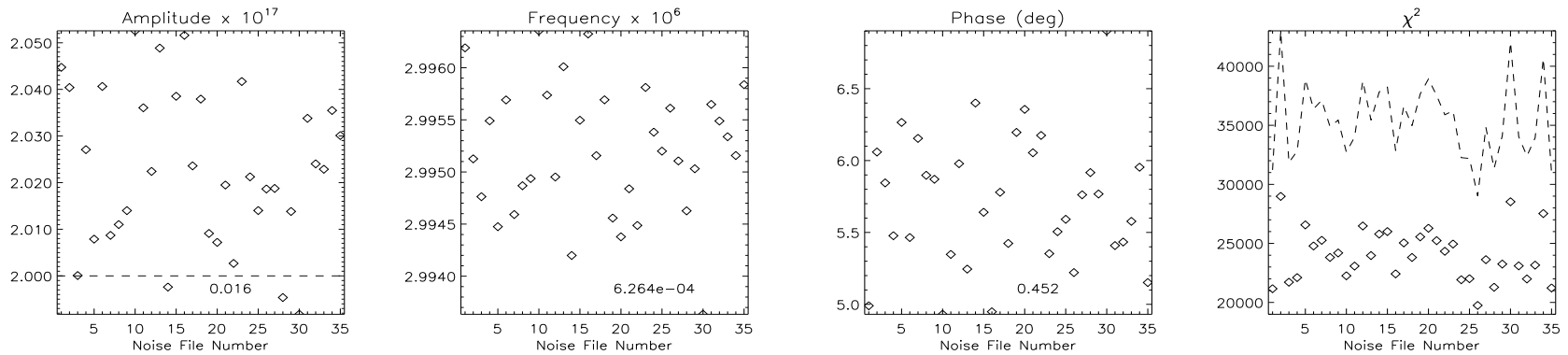


Figure 6.4: Search algorithm results for the 35 realizations of the LISA data stream in the presence of a SNR=5 signal only. The dashed lines represent the input parameters. In the case of χ^2 , the dashed line is the value of χ^2 in the original LISA data stream realization. The standard deviation of each collection is printed at the bottom of each plot.

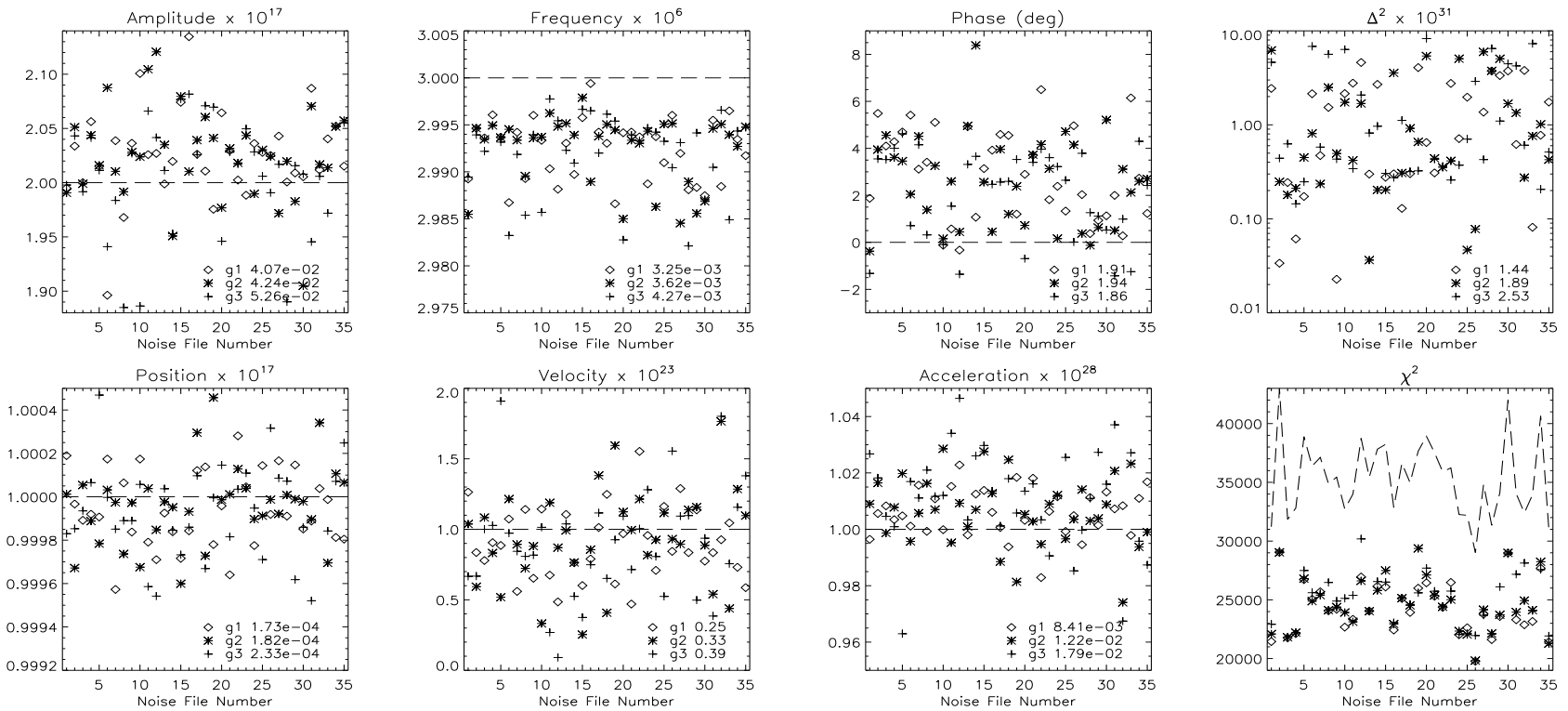


Figure 6.5: Search algorithm results for the 35 realizations of the LISA data stream with a SNR=5 3 μ Hz signal and a disturbance with an interruption length of 17 minutes. The dashed lines represent the input parameters. In the case of χ^2 , the dashed line is the value of χ^2 in the original LISA data stream realization. As in the other figures, the standard deviations are printed at the bottom of each plot.

systematic low has been removed. The Fourier transform of a linear slope is a $1/f$ power spectrum. A monochromatic sine wave will have $1/f$ lobes. Therefore it is not unreasonable for there to be a correlation between the frequency of the sinusoid (which determines the amplitude of the lobes) and the velocity. This correlation has caused the derived velocities to be different when there is a sine wave present.

Figure 6.5 contains results for all 35 noise realizations. From these results we again compute standard deviations and use these as estimates of the errors in each of our derived parameters. The errors in the signal parameters are roughly (2.3%, 0.12%, 0.53%). The errors in the disturbance parameters are roughly [0.02%, 32%, 1.3%]. With a disturbance present our algorithm has performed worse at obtaining the true values of the signal parameters. In addition, the spread in derived disturbance parameters also has increased. However, the derived errors from the 35 realizations imply that the input parameters are just within our solution. This does not mean that our algorithm did a good job at locating the signal, rather that the disturbance did not significantly push the solution away from the input.

At this point the reader may be interested in the results of searching for the science signal while ignoring the data disturbance. Matched filtering of a sinusoid alone without removing the disturbance unfortunately is not effective. After a sufficiently long amount of time the strain data will not be centered about zero and matched filtering will never produce a large correlation result.

Our search algorithm, which creates a χ^2 in the frequency domain, also has a hard time fitting the signal parameters when the disturbance is present and not removed. Table 6.3 contains our search results for the signal when the disturbance is not removed. As can be seen, the search algorithm is confused by the presence of extra power at lower frequencies deriving from the velocity and acceleration terms in the disturbance. This increase in power has caused the search algorithm to choose a much larger value for the amplitude of the sinusoid. A comparison of these derived sinusoid parameters with the ones mentioned for the naïve guess described in §6.2 show that our algorithm, when not fitting

Table 6.3: Search algorithm results for the three disturbances examined, and with a low frequency signal, without removing the disturbance. A is in multiples of 10^{-17} , f is in μHz , and Δ^2 is in multiples of 10^{-31} .

	A	f	ϕ	χ^2	Δ^2
input	2.000	3.000	0.000°	31 369	0.437
nodis	2.045	2.996	4.990°	21 164	0.000
g_1	37.468	3.495	-174.757°	4.15×10^{15}	45 568.500
g_2	28.368	3.192	-176.303°	5.28×10^{14}	25 642.700
g_3	19.348	3.187	-174.834°	1.24×10^{15}	11 819.300

for the disturbance, is equivalent to guessing the sinusoid parameters using the frequency domain information alone.

6.4.2 Higher Frequency Signal

Disturbances should have a stronger effect on low frequency sources than high frequency sources. At higher frequencies, data disturbances should be less noticeable since there are many more cycles present in the data set. Fitting for the disturbance parameters with only a high frequency signal present should be much like fitting for the parameters with no signal present at all.

We choose to look at a 100 μHz signal for our higher frequency case. A signal-to-noise ratio of 5 implies an rms strain amplitude of 1.6×10^{-21} for our realization. Our input parameters are now $(A, f, \phi)_{\text{input}} = (1.6 \times 10^{-21}, 100 \mu\text{Hz}, 0^\circ)$. With no disturbance, our search algorithm found the following parameters: $(1.682110 \times 10^{-21}, 99.98995 \mu\text{Hz}, 10.60217^\circ)$ with $\chi^2 = 30692$. Table 6.4 contains results for this high frequency signal in the presence of a disturbance with an interruption of 17 minutes in length. Comparing with Table 6.1 we see that the derived disturbance parameters are almost identical. The frequency of the signal has been found to within 10 nHz, or 0.015%, of the input frequency. The disturbance has had little effect on the ability of our algorithm to determine the high frequency signal parameters.

Table 6.4: Search algorithm results for the three disturbances examined, and with a high frequency signal. A is in multiples of 10^{-21} , $f^* = (f - 100 \mu\text{Hz})/\text{nHz}$, $p^* = (p - 10^{-17}) \times 10^{20}$, u is in multiples of 10^{-23} , a is in multiples of 10^{-28} , and Δ^2 is in multiples of 10^{-39} .

	A	f^*	ϕ	p^*	u	a	χ^2	Δ^2
input	1.600	0.000	0.000°	0.000	1.000	1.000	31364	1.302
nodis	1.682	-10.050	10.602°				30692	0.000
$g1$	1.687	-9.950	10.413°	-0.148	0.652	1.011	29925	0.366
$g2$	1.689	-10.080	10.778°	-0.047	0.930	1.003	30916	0.382
$g3$	1.684	-10.040	10.586°	-0.121	0.750	1.011	30185	0.394

6.5 Disturbance Limitations on Operations

As we have seen in the previous section, isolated disturbances do not significantly affect our ability to identify and remove sources from the data stream, given enough data before and after the disturbance location. The following question can now be posed: ‘‘How does the frequency of disturbances affect our search algorithm?’’ For this analysis we examined only the lower frequency signal case since the identification of higher frequency signals is not perturbed much in the presence of a disturbance.

6.5.1 Frequency

Each disturbance contains three parameters which need to be fitted. As mentioned earlier, the accuracy to which you can determine the parameters depends on the number of sample points available. Within a 38-day period, the inclusion of multiple disturbances effectively reduces the amount of information available to determine each disturbance. We can simulate the effect of having multiple disturbances by placing one disturbance near the end of the data stream. This assumes that all previous disturbances have been fit and removed accurately. Table 6.5 shows our results for one disturbance, with $L = 1000$ seconds, at various times from the end of the data stream. With a small number of points after the data interruption it is hard to determine the acceleration parameter. This implies that the assumption that we removed all other disturbances exactly is wrong — we probably would not have been able to determine the acceleration well. Therefore we must fit for all of the disturbances at once since the acceleration parameter will strongly affect the data at large times.

As a preliminary investigation we examined several cases with multiple disturbances. We simulated data streams with the maximal number of disturbances possible all equally spaced by T , and of

Table 6.5: Search results for a low frequency signal with one disturbance. The disturbance start time is T before the end of the data stream, and the length L is 1000 seconds. The last three data sets are $g3$, $g2$, and $g1$ from Table 6.2. A is in multiples of 10^{-17} , $f^* = (f - 3 \mu\text{Hz})/\text{nHz}$, $p^* = (p - 10^{-17}) \times 10^{20}$, u is in multiples of 10^{-23} , a is in multiples of 10^{-28} , and Δ^2 is in multiples of 10^{-31} .

T	A	f^*	ϕ	p^*	u	a	χ^2	Δ^2
input	2.000	0.000	0.000°	0.000	1.000	1.000	33 575	0.437
nodis	2.045	-3.809	4.990°				21164	0.000
1.2 hrs	2.000	-0.049	0.007°	-0.048	1.013	17.369	32 933	0.441
2.4 hrs	1.999	-0.223	0.029°	-0.096	0.849	19.351	30 669	0.454
6.0 hrs	1.998	-0.666	0.091°	-0.241	0.431	10.602	28 215	0.492
12 hrs	1.996	-1.237	0.168°	-0.311	0.284	5.710	25 756	0.549
1.0 day	1.989	-3.552	0.493°	-0.145	0.625	3.548	21 438	0.866
1.2 days	1.991	-2.790	0.388°	-0.404	0.392	2.737	22 334	0.745
2.4 days	1.985	-6.858	1.562°	0.434	1.941	1.191	21 653	1.271
3.5 days	1.987	-4.003	0.568°	-0.371	0.194	1.457	24 205	0.942
5.0 days	1.988	-5.911	2.318°	-0.131	0.612	1.194	21 772	0.747
7.0 days	2.001	-3.662	1.557°	-0.323	0.166	1.171	24 628	0.519
9.5 days	1.986	-6.369	2.492°	-0.271	0.547	1.086	21 947	0.796
14 days	1.929	-8.449	1.817°	0.461	1.699	0.963	25 784	1.905
19 days	2.021	-6.491	2.646°	0.302	1.535	0.976	22 295	0.695
23 days	1.978	-8.641	3.191°	0.173	1.395	0.987	22 349	1.132
26 days	1.998	-10.561	-1.322°	-0.170	0.667	1.027	22 914	4.628
32 days	1.991	-14.507	-0.378°	0.012	1.037	1.009	22 088	6.196
37 days	1.995	-10.754	1.875°	0.190	1.264	0.996	21 449	2.443

length $L = 1000$ seconds. Our results of this analysis are in Table 6.6. Note that our search algorithm was not designed to handle more than 20 disturbances, therefore instead of having 31 equally spaced disturbances in the $T = 1.2$ days data set and 37 disturbances in the $T = 1$ case, we only have 20.

It appears that our algorithm has trouble with disturbance frequencies of more than one every 3.5 days. The algorithm not only has trouble fitting for the frequency and phase, but the velocities and accelerations for most of the the disturbances are much worse — this is evident by the large values of χ^2 . Some refinement of the search algorithm for multiple gaps may help to fit the velocities and accelerations better for a high frequency of disturbances.

Table 6.6: Search algorithm results in the presence of multiple disturbances. # is the number of disturbances present in the data stream, separated in time by T , and all of length $L = 1000$ seconds. Notice the degradation in the derived signal parameters and the value of χ^2 for disturbance frequencies more than one every 3.5 days. A is in multiples of 10^{-17} , $f^* = (f - 3 \mu\text{Hz})/\text{nHz}$, and Δ^2 is in multiples of 10^{-31} .

T	#	A	f^*	ϕ	χ^2	Δ^2
input	0	2.000	0.000	0.000°	31 218	0.437
nodis	0	2.045	-3.809	4.990°	21 164	0.000
1.0 day	20	2.447	-403.809	12.292°	4 991 500	287.479
1.2 days	20	2.093	-192.644	12.261°	3 853 900	311.622
2.4 days	15	1.991	-2.163	-1.474°	819 324	1.329
3.5 days	10	2.046	-3.735	5.180°	29 346	0.002
5.0 days	7	1.993	-6.640	3.392°	24 459	0.561
7.0 days	5	2.044	-5.313	4.941°	23 877	0.048
9.5 days	3	2.076	3.460	3.027°	23 818	0.277
14 days	2	2.018	-3.579	0.781°	21 993	0.708
19 days	1	2.108	-1.311	2.381°	21 393	0.222

6.6 Chapter Summary

We have presented an algorithm which identifies monochromatic signals in the presence of data disturbances in a LISA-like data stream. The results here are encouraging on the ability of removing the disturbances without affecting the science results. However, we found complications with data sets containing disturbances which are more frequent than about once every 3.5 days. There are a number of simplifications in our data stream, for instance the use of monochromatic signals and instantaneous drop out and recovery on the edges of interruptions. As mentioned in §6.3.2, longer data sets will allow all parameters to be fit more accurately. Therefore the results from our short 38 day long data set analysis should be taken only as a guide for longer data sets.

