Optimal Experimental Design: Introduction and Application to System Identification

Larry Price

LIGO-G1400084

Why bother?

same total excitation power in each case		White noise	Stepped Sine	Optimal
Pendulum With Bounce	time (normalized)	1	2.2	1
	max error	70%	70%	1%
OLG of PDH control servo	time (normalized)	1	2.8	1
	max error	16%	2%	1%

Overview

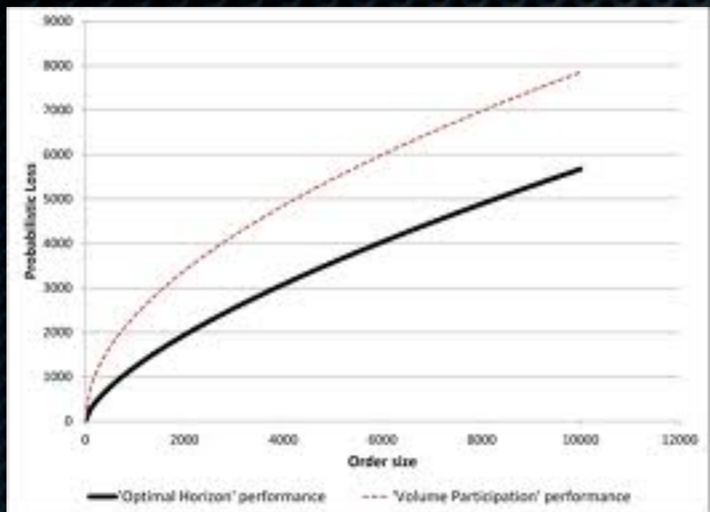
- Introduction to optimal design
- Basics of system identification
- Some examples
- Future work

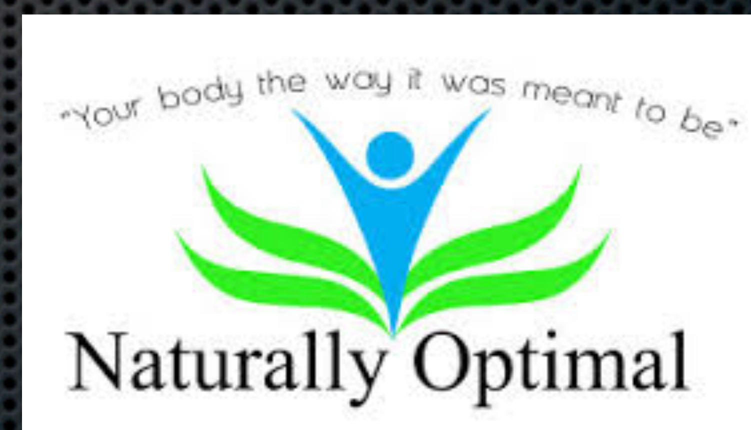
Optimal design in a nutshell

- First works date back to the early 19th century
  - Francis Bacon (*New Atlantis*, 1624):

    "Then after divers meetings and consults of our whole number, to consider of the former labors and collections, we have three that take care out of them to direct new experiments, of a higher light, more penetrating into nature than the former. These we call lamps."
- Basic goal: Distribute finite resources "optimally"

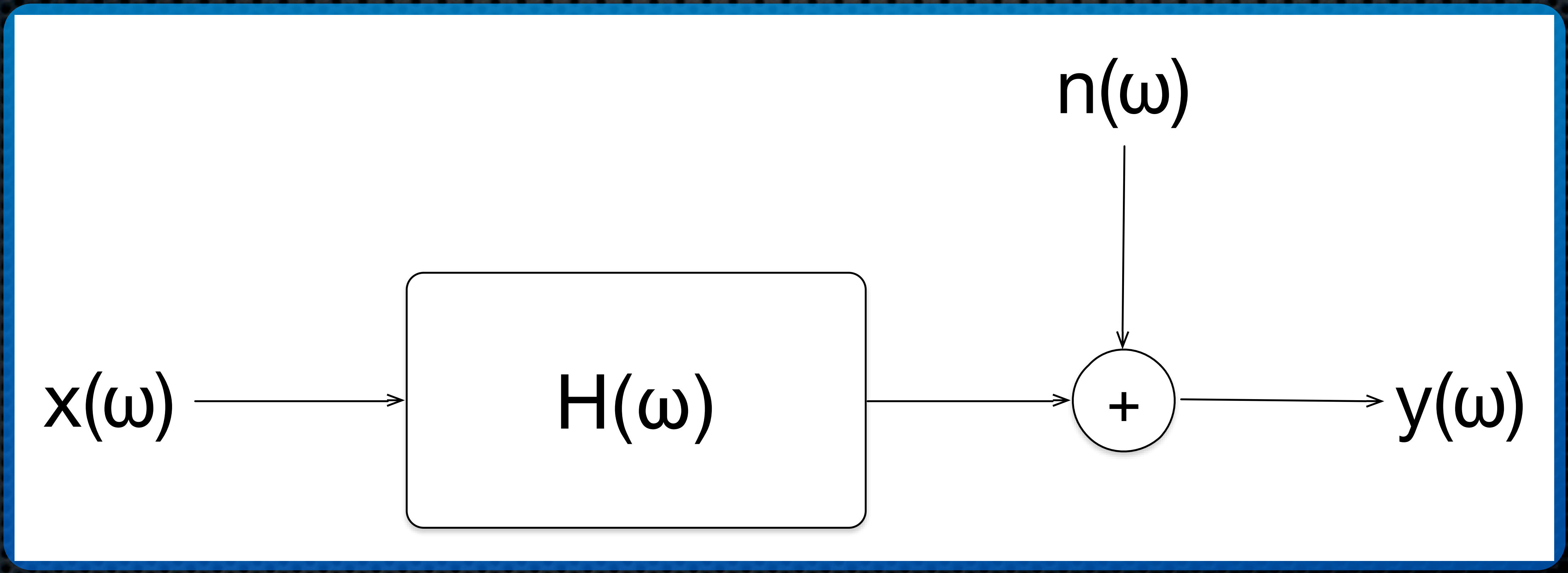
What is optimal?





System identification (sys id)

- Choose an x and find H



Sys id

- In the frequency domain

$$y(\omega) = x(\omega)H(\vec{\theta};\omega) + n(w)$$

- Estimate of H given by

$$H(\vec{\theta};\omega) = \frac{S_{xy}(\omega)}{S_x(\omega)}$$

- Next: parameter estimation*

Two goals of sys id

- Determine the features of your system beyond the TF: resonances, nonlinearity, noise (Non-parametric)
- Identify the parameters of your system, e.g., for control (Parametric)
    - Our use case: check the location of previously determined parameters
    - Excitation design is fundamentally linked to the method of parameter estimation!
    - Want to minimize errors on parameter estimates

Covariance matrix

- The covariance matrix generalizes the notion of variance when you have multiple parameters
- Diagonal elements are variances of the variables
- Off-diagonal elements are *covariances* between variables

$$\Sigma = \begin{bmatrix} \sigma^2_{\theta_1} & \sigma_{\theta_1\theta_2} & \cdots & \sigma_{\theta_1\theta_N} \\ \sigma_{\theta_1\theta_2} & \sigma^2_{\theta_2} & \cdots & \sigma_{\theta_2\theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_1\theta_N} & \sigma_{\theta_2\theta_N} & \cdots & \sigma^2_{\theta_N} \end{bmatrix}$$

Covariance matrix cont.'d

- In practice, the covariance matrix is difficult to compute before the experiment is performed
- Instead, we use a result from Cramer & Rao that says a lower bound on the covariance matrix is given by the inverse of the *Fisher information matrix.*
    - Maximize Fisher = Minimize Covariance
    - NB: In practice you tend to achieve results higher than the Cramer-Rao bound indicates
    - Our FOM is the *maximum* of the CR bounds of the system parameters

Fisher information

- Phrased in terms of the *likelihood function* (probability of some outcome given a set of parameter values)
- Explicitly

$$\mathcal{I}_{jk} = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\log\mathcal{L}(y;\vec{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log\mathcal{L}(y;\vec{\theta})\right)\right]$$

What to optimize?

- Fisher is a matrix but we want a scalar.
  Some options:
  - The trace
  - The determinant
  - The minimum eigenvalue
  - Something more complicated....
- In practice, optimization of one criteria results in a good design by other criteria

Optimal sys id

- What's the "best" excitation signal we can use given limited input power?

$$\int_0^\infty d\omega \, S_x(\omega) = 1$$

Fisher for generic TF

- First assume the noise is Gaussian, and write the likelihood

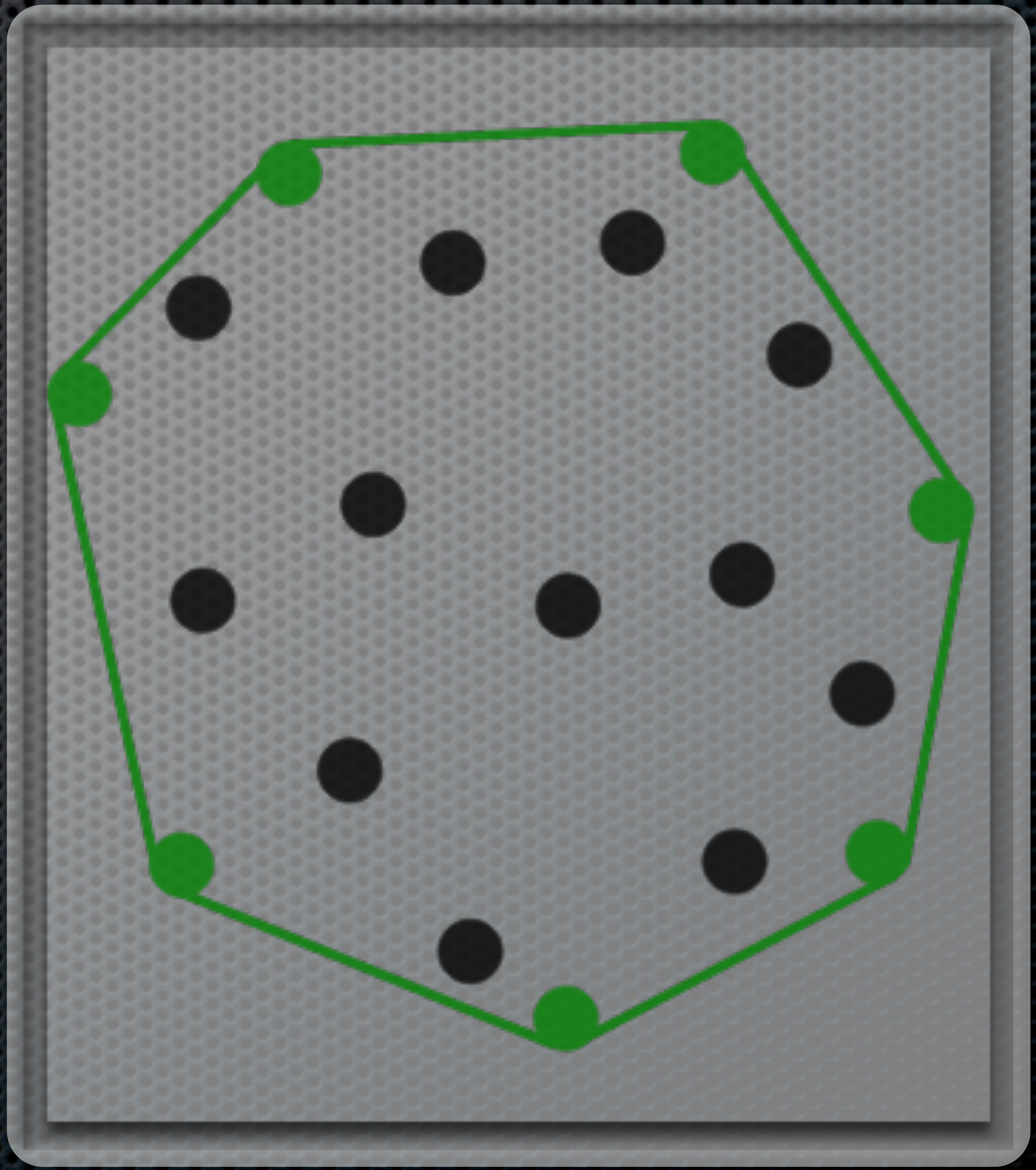
$$\mathcal{L}(y;\vec{\theta}) \propto \exp\left[-\frac{1}{2}\int_0^\infty \frac{d\omega}{S_n(\omega)}\left|y(\omega) - H(\vec{\theta};\omega)x(\omega)\right|^2\right]$$

- Fisher is then given by

$$\mathcal{I}_{jk} = \frac{1}{2}\int_0^\infty d\omega\,\frac{S_x(\omega)}{S_n(\omega)}\Re\left[\left(\frac{\partial H(\vec{\theta};\omega)}{\partial\theta_j}\right)^*\left(\frac{\partial H(\vec{\theta};\omega)}{\partial\theta_k}\right)\right]$$

Useful fact

- The set of all input power-constrained Fisher matrices forms the *convex hull* of single-frequency Fisher matrices
- This means the (non-unique!) optimal excitation is a linear combination of sine waves



Example:

- Transfer function: $$H(\theta;\omega) = \frac{1}{i\omega\theta + 1}$$

Example:

- Transfer function: $$H(\theta;\omega) = \frac{1}{i\omega\theta + 1}$$

- Fisher matrix: $$\mathcal{I} = \frac{1}{2S_n(\omega)} \frac{\omega^2}{(1+\theta^2\omega^2)^2}$$

Example:

- Transfer function: $$H(\theta;\omega) = \frac{1}{i\omega\theta + 1}$$
- Fisher matrix: $$\mathcal{I} = \frac{1}{2S_n(\omega)} \frac{\omega^2}{(1+\theta^2\omega^2)^2}$$
- Optimal design: $$\omega = \frac{1}{\theta}$$

Same example zpk-style

- Transfer function: $$H(\vec{\theta};\omega) = \frac{1/\theta}{i\omega + 1/\theta}$$

Same example zpk-style

- Transfer function: $$H(\vec{\theta};\omega) = \frac{1/\theta}{i\omega + 1/\theta}$$

- Fisher matrix:
$$\mathcal{I} = \frac{1}{2S_n(\omega)(\omega^2 + 1/\theta^2)^2}\begin{bmatrix}\omega^2 + 1/\theta^2 & -1/\theta^2 \\ -1/\theta^2 & 1/\theta^2\end{bmatrix}$$

Same example zpk-style

- Transfer function: $$H(\vec{\theta};\omega) = \frac{1/\theta}{i\omega + 1/\theta}$$

- Fisher matrix:

$$\mathcal{I} = \frac{1}{2S_n(\omega)(\omega^2 + 1/\theta^2)^2} \begin{bmatrix} \omega^2 + 1/\theta^2 & -1/\theta^2 \\ -1/\theta^2 & 1/\theta^2 \end{bmatrix}$$

$$\det(\mathcal{I}) = \frac{\omega^2/\theta^2}{4S_n(\omega)^2(\omega^2 + 1/\theta^2)^4}$$

Same example zpk-style

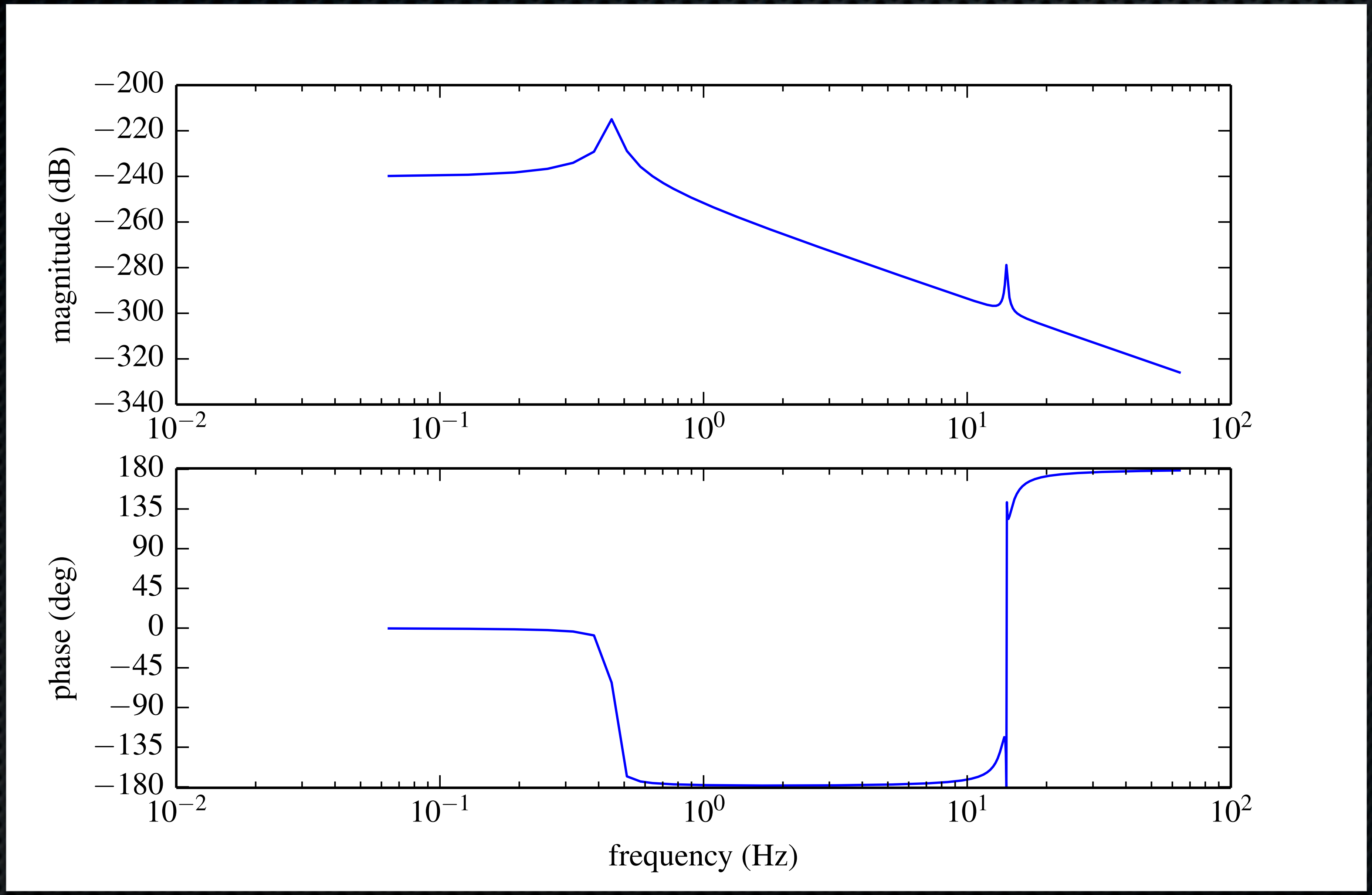
- Transfer function: $$H(\vec{\theta};\omega) = \frac{1/\theta}{i\omega + 1/\theta}$$
- Fisher matrix:
$$\mathcal{I} = \frac{1}{2S_n(\omega)(\omega^2 + 1/\theta^2)^2} \begin{bmatrix} \omega^2 + 1/\theta^2 & -1/\theta^2 \\ -1/\theta^2 & 1/\theta^2 \end{bmatrix}$$
$$\det(\mathcal{I}) = \frac{\omega^2/\theta^2}{4S_n(\omega)^2(\omega^2 + 1/\theta^2)^4}$$
- Optimal design: $$\omega = \frac{1}{\sqrt{3}\theta}$$

More generally

- Use an automated procedure:
  - Select frequencies of interest*
  - Iterate to determine the amplitudes

Simulations

- Compare white noise, swept (stepped!) sine, and optimal excitations
- Chose a fixed time and ramped up the amplitude until errors were ~1%
- Swept sine was chosen to take roughly the same measurement time (minus a 5 second dwell between each frequency) as the others
- Gaussian noise



Pendulum with bounce mode

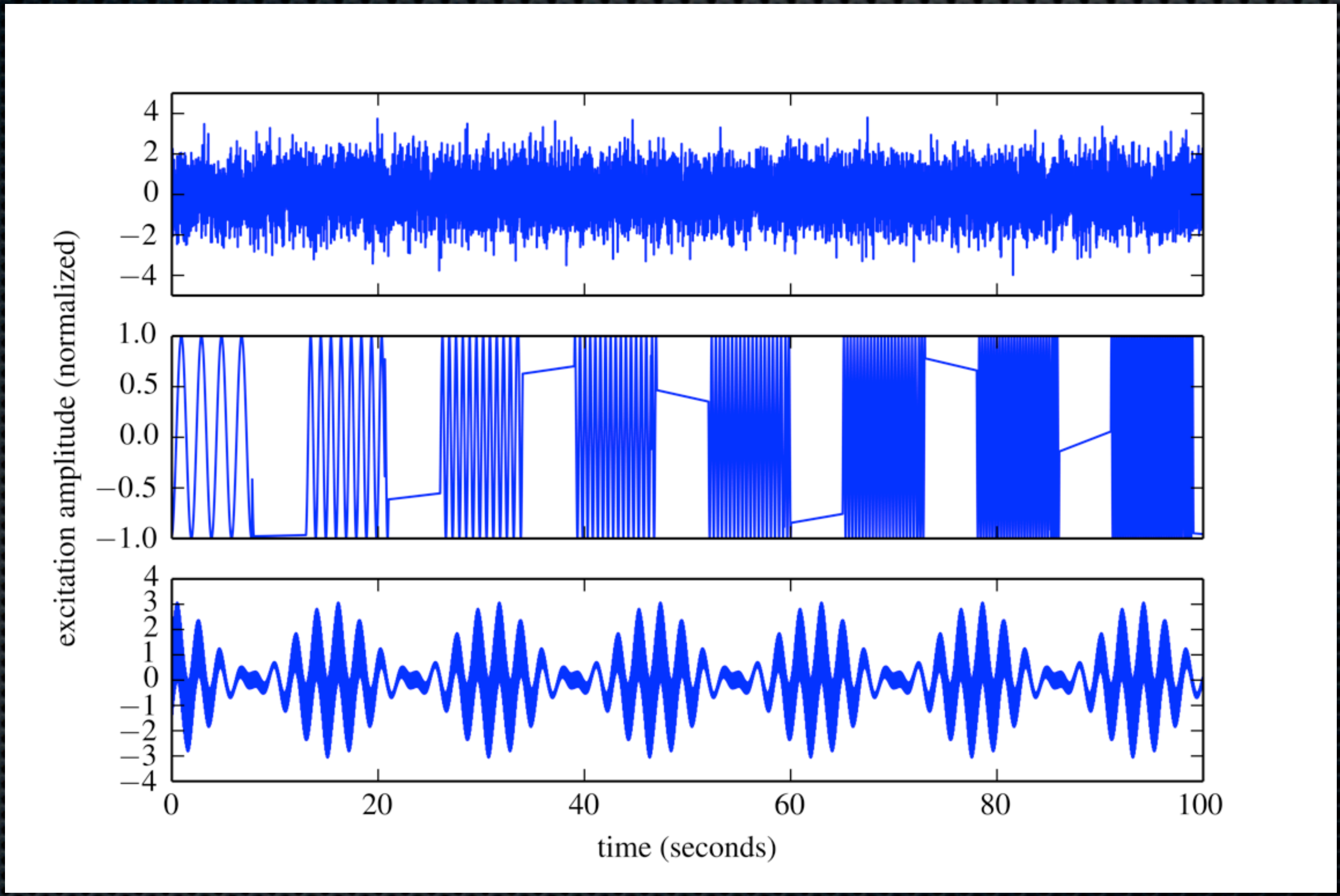
$Q_0 = 20$

$f_0 = .454$ Hz

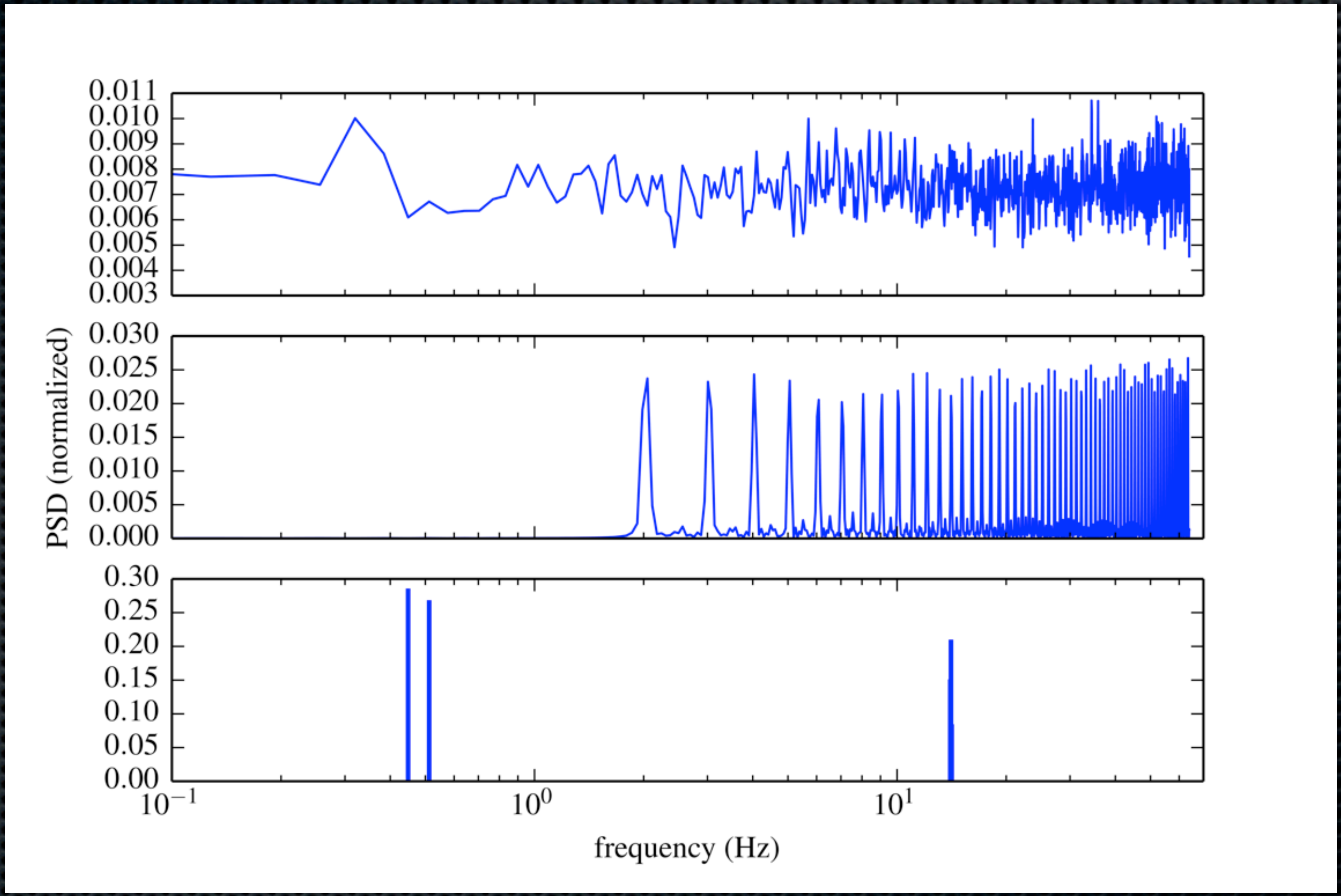
$Q_B = 100$

$f_B = 14.08$ Hz

Time domain excitations

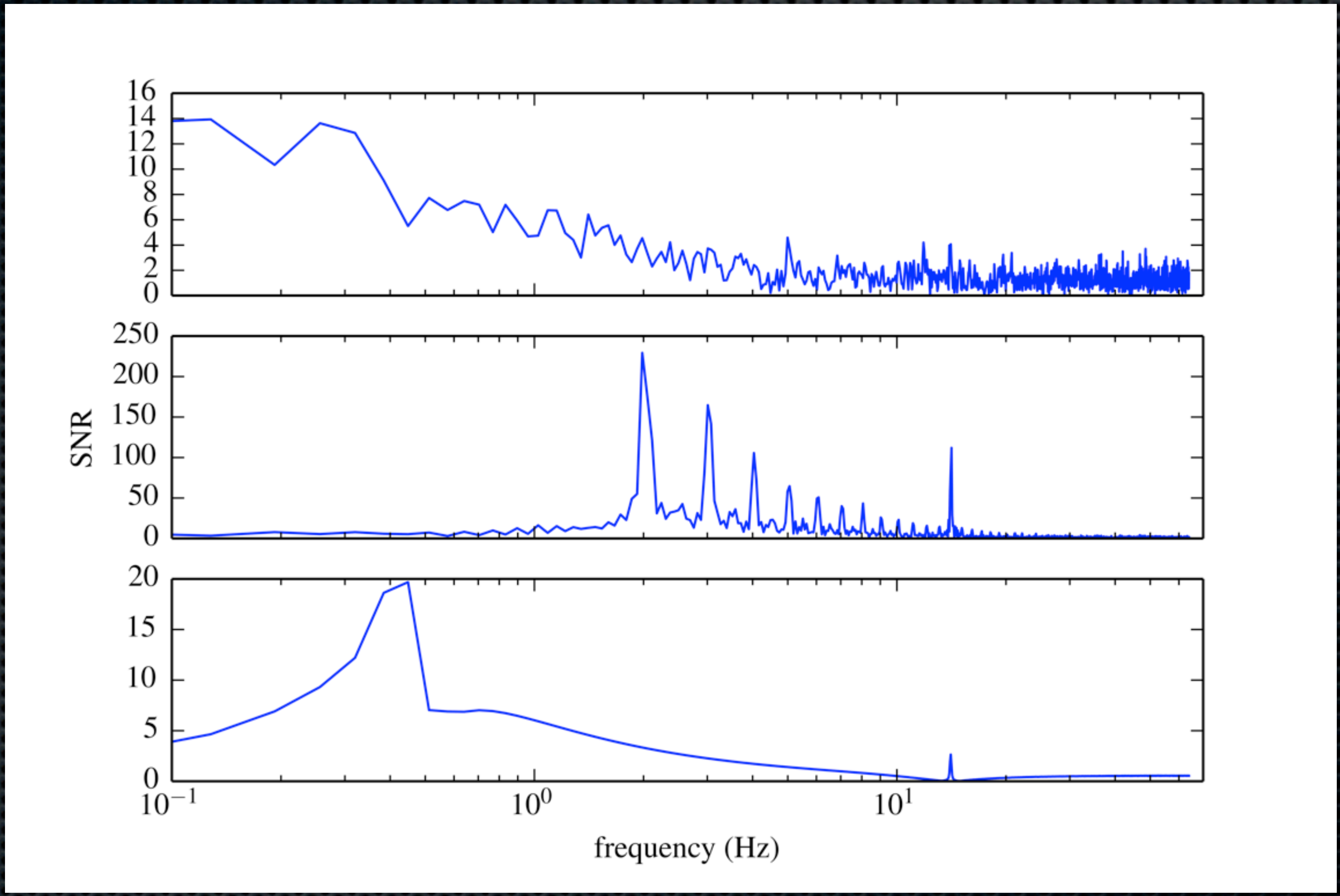


PSDs



0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
PSD (normalized)
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0.30
0.25
0.20
0.15
0.10
0.05
0.00
10^{-1}
10^{0}
10^{1}
frequency (Hz)

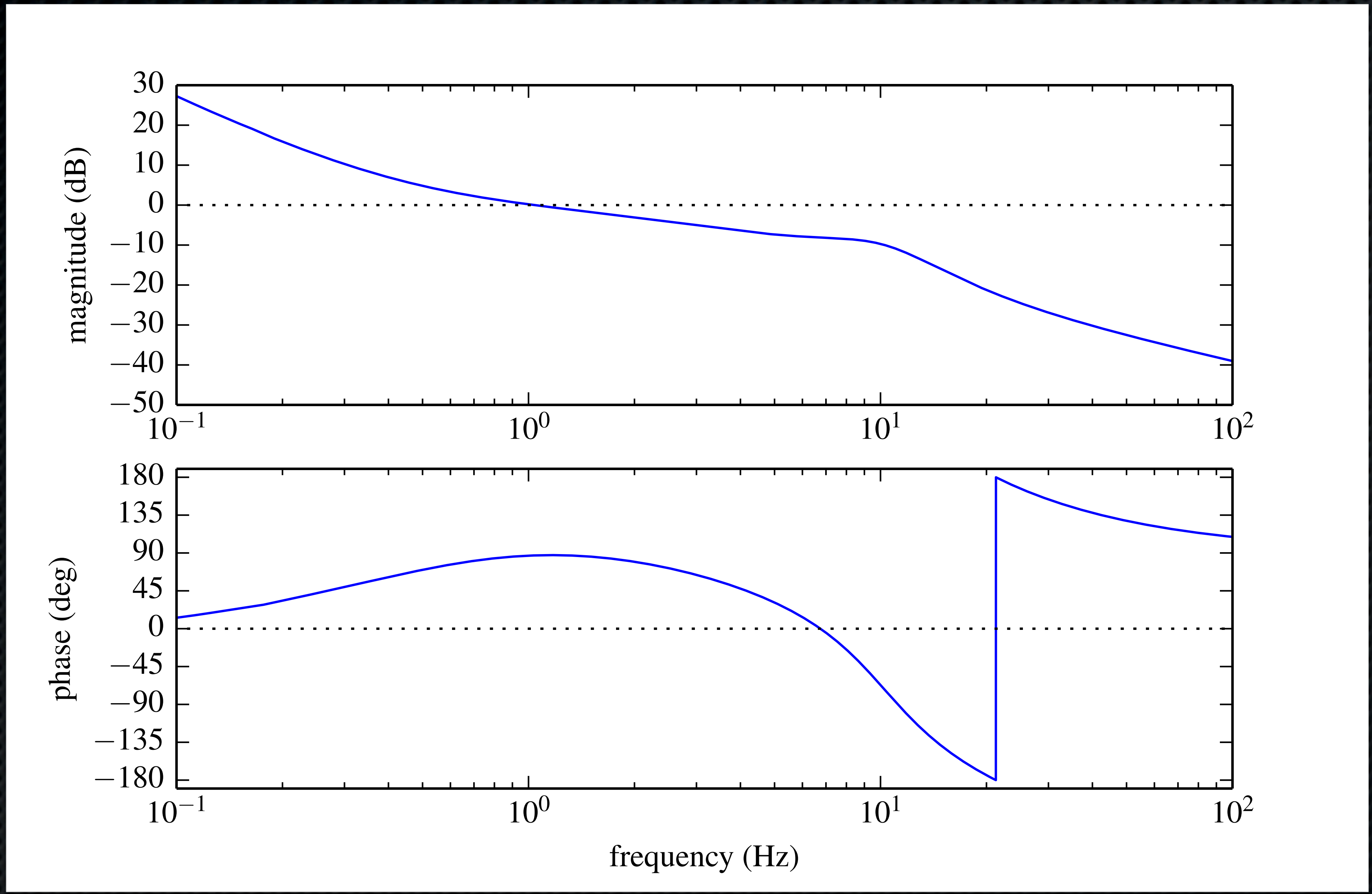
SNR



16
14
12
10
8
6
4
2
0
250
200
150
100
50
0
SNR
20
15
10
5
0
10^{-1}
10^{0}
10^{1}
frequency (Hz)

Results

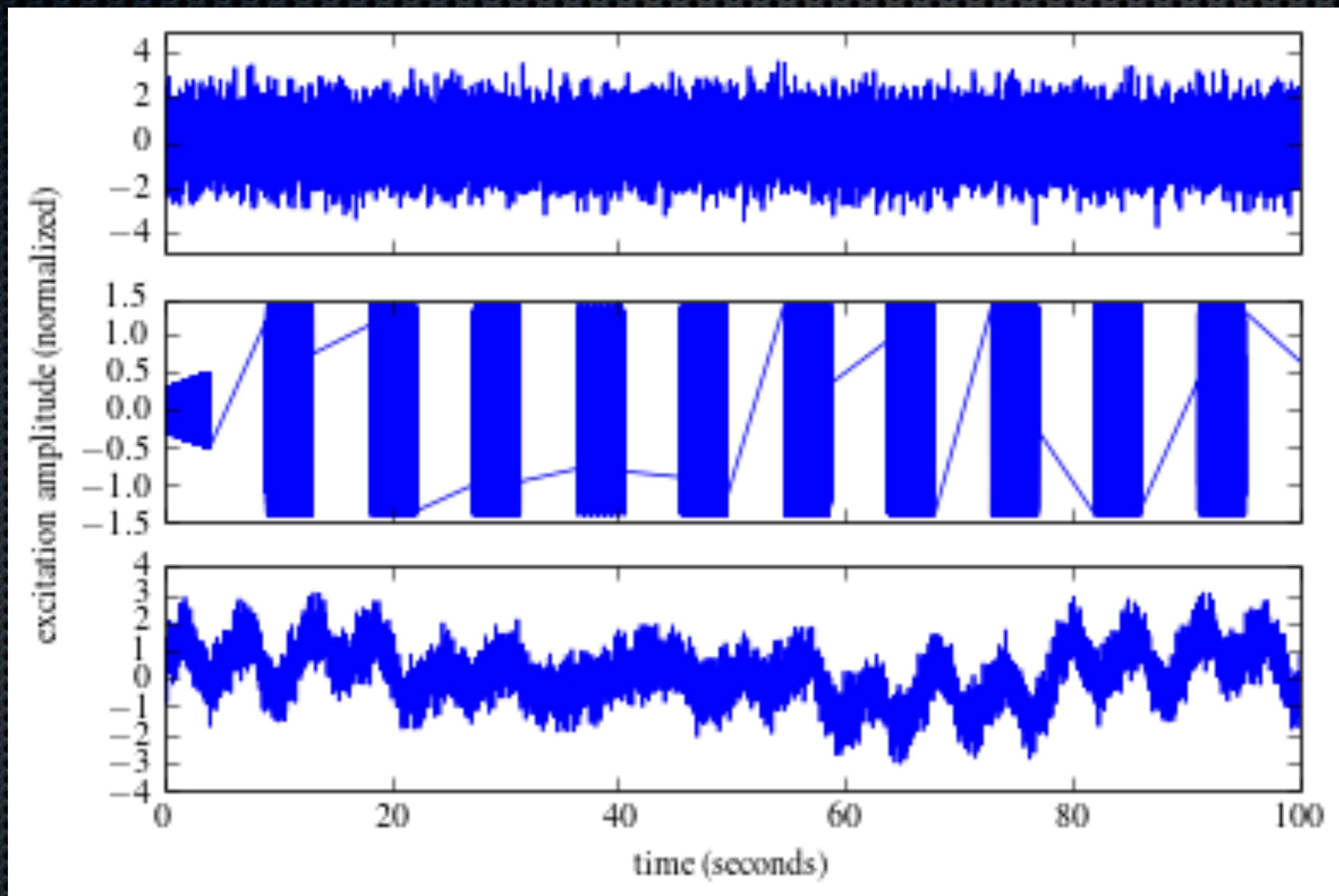
same total excitation power in each case		White noise	Stepped Sine	Optimal
Pendulum With Bounce	time (normalized)	1	2.2	1
	max error	70%	70%	1%
OLG of PDH control servo	time (normalized)	1	2.8	1
	max error	16%	2%	1%



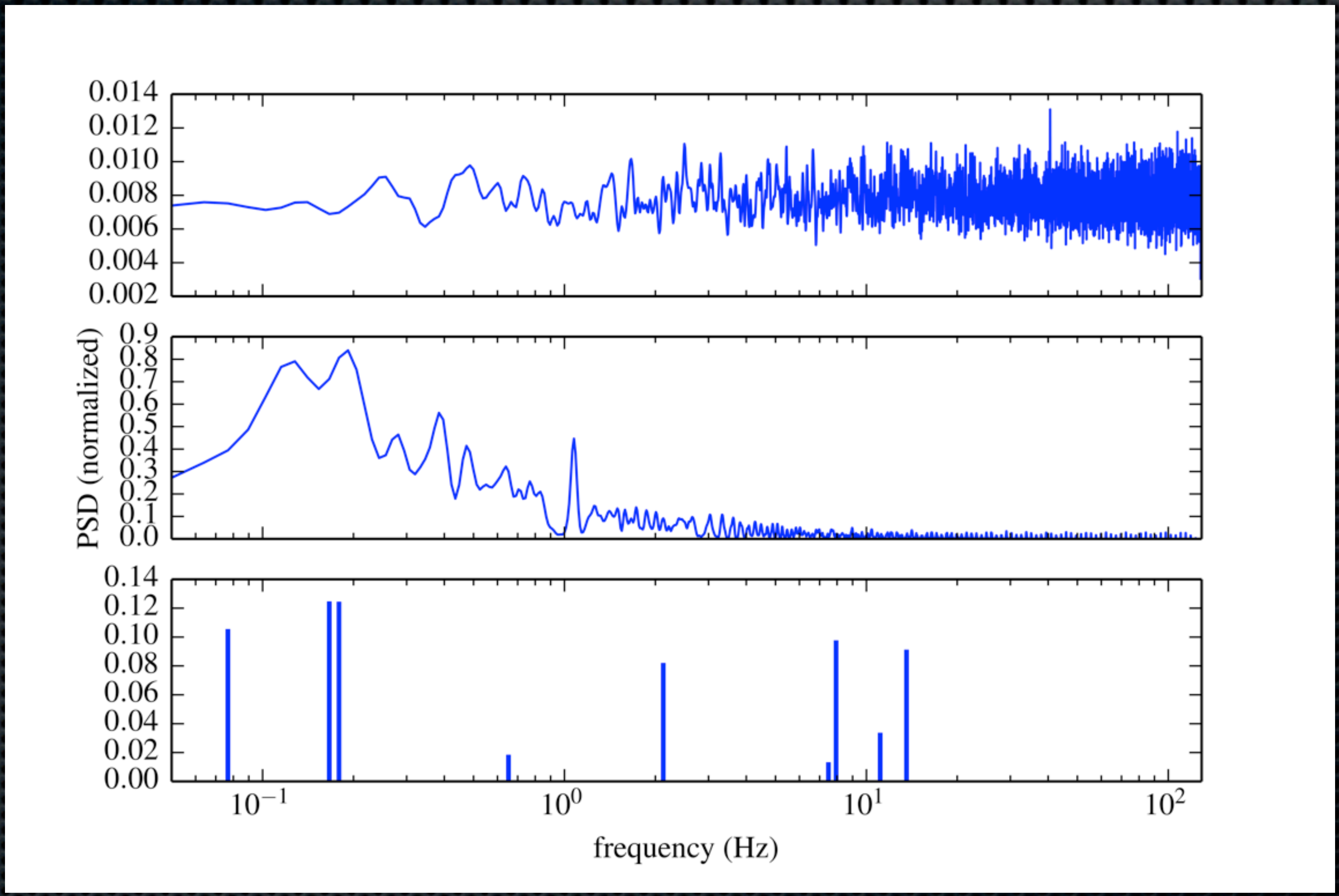
Open-loop gain of PDH control servo (scaled)

7 zeros, 8 poles, and a gain

Time domain excitations

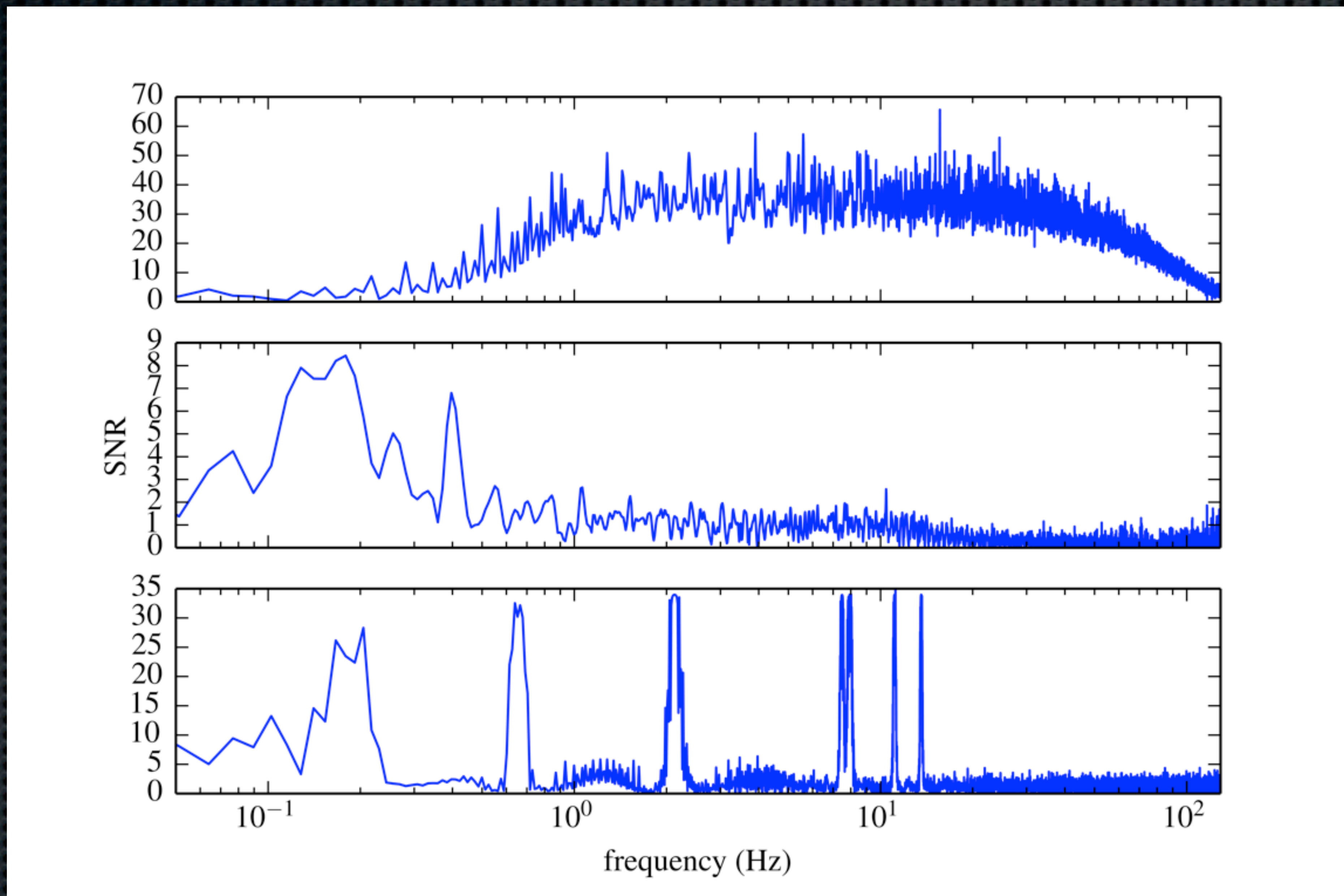


PSDs



0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PSD (normalized)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10^{-1}
10^{0}
10^{1}
10^{2}
frequency (Hz)

SNR



Summary

same total excitation power in each case		White noise	Stepped Sine	Optimal
Pendulum With Bounce	time (normalized)	1	2.2	1
	max error	70%	70%	1%
OLG of PDH control servo	time (normalized)	1	2.8	1
	max error	16%	2%	1%

Robustness (preliminary)

jitter	max error
0.1%	1%
1%	4%
5%	30%
10%	60%

- Pendulum with bounce
- Parameters jittered with a random sign
- Excitation optimized for un-jittered values
- Average max errors

Future work

- Testing on a broader range of systems
- Better frequency selection?
- Parameter estimation
- Work on interface / integration to existing tools
- Calibration line placement
- MIMOs
- Beyond sys id

Amplitude optimization

- Define the *dispersion* function

$$\nu(S_x, \omega) = \mathrm{trace}\left(\mathcal{I}^{-1}(S_x)\mathcal{I}(\omega)\right)$$

- Properties:
  - Minimization is equivalent to maximization of det(Fisher)
  - Has a maximum value less than or equal to the number of parameters (equality at the optimal design)

Algorithm

- Choose a set of frequencies
- Distribute power equal amongst them
- Iterate until max iterations or tolerance is reached:

$$S_x^{i+1} = \frac{\nu(S_x^i, \omega)}{n} S_x^i$$