TIME-FREQUENCY ANALYSIS

M2 AI — SIGNAL PROCESSING

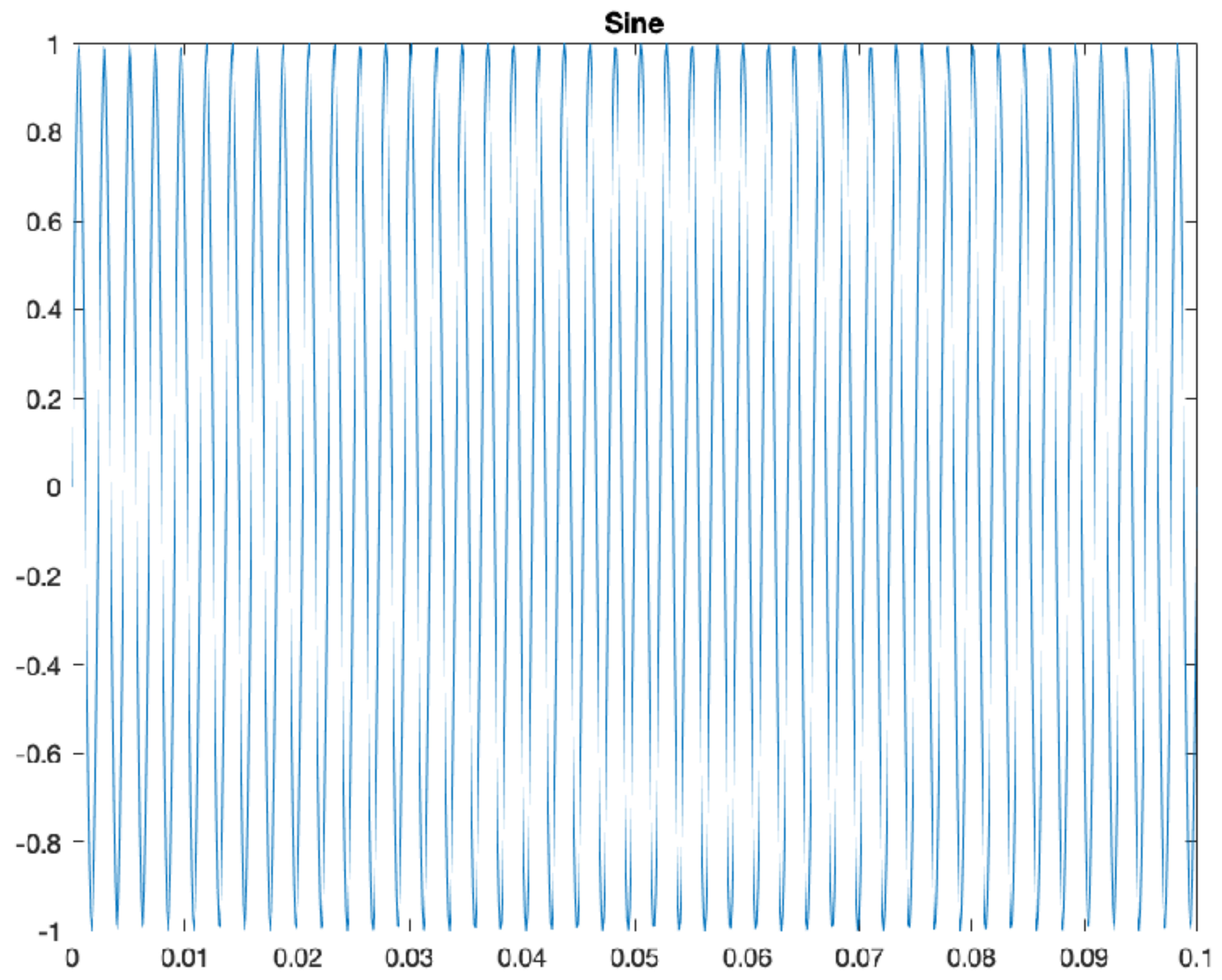
FOURIER REMINDER

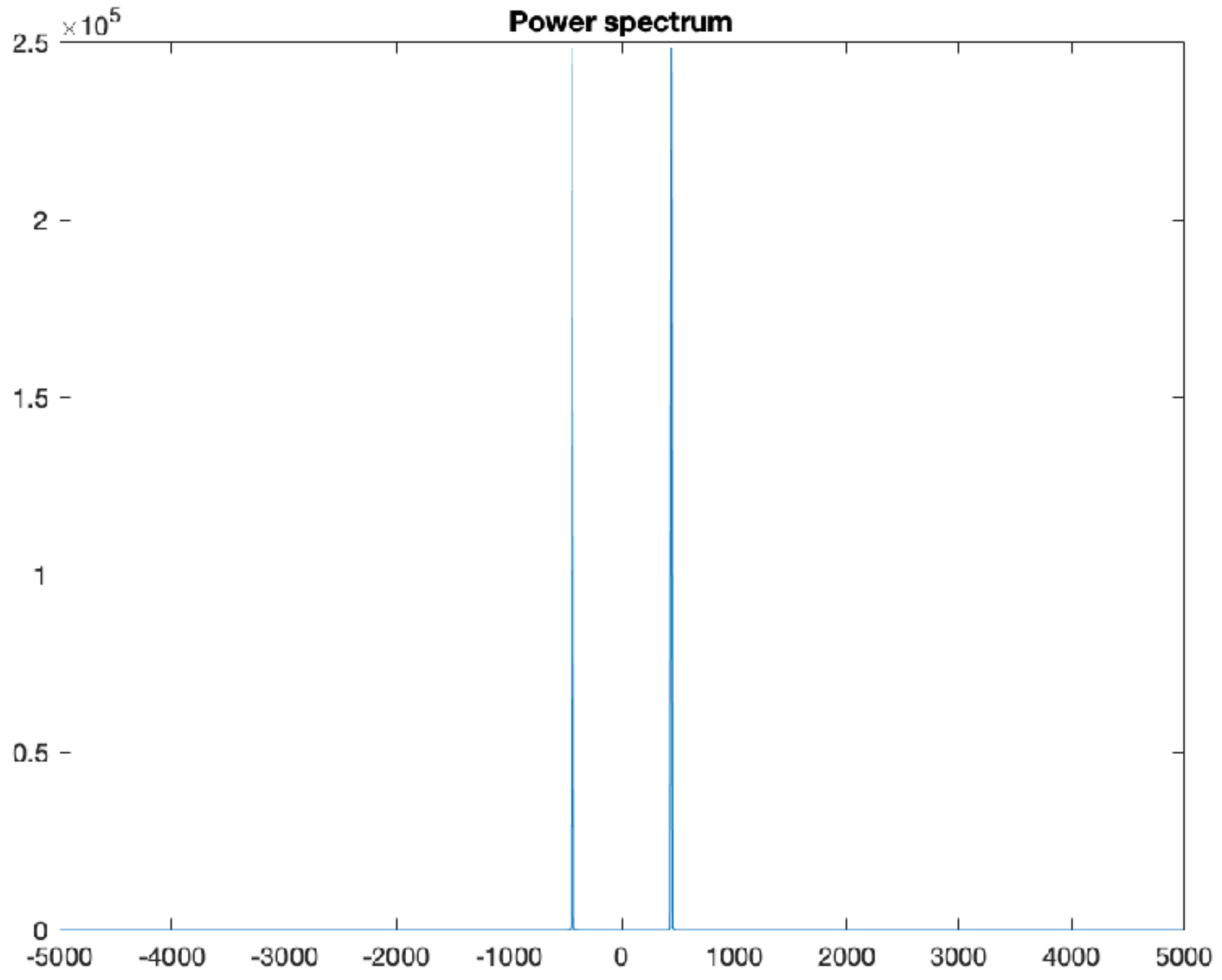
- Spectral analysis: frequency content of a function (Think about musical notes!)
- Measure the similarity (correlation, angle) between pure (complex) sine and a signal
- Sines are eigen signals of time-invariant linear systems (filters)
- Fourier analysis computes the correlation between the signal $s(t)$ a pure sine at various frequencies ν

$$\hat{s}(\nu) = \langle s, e_\nu \rangle$$

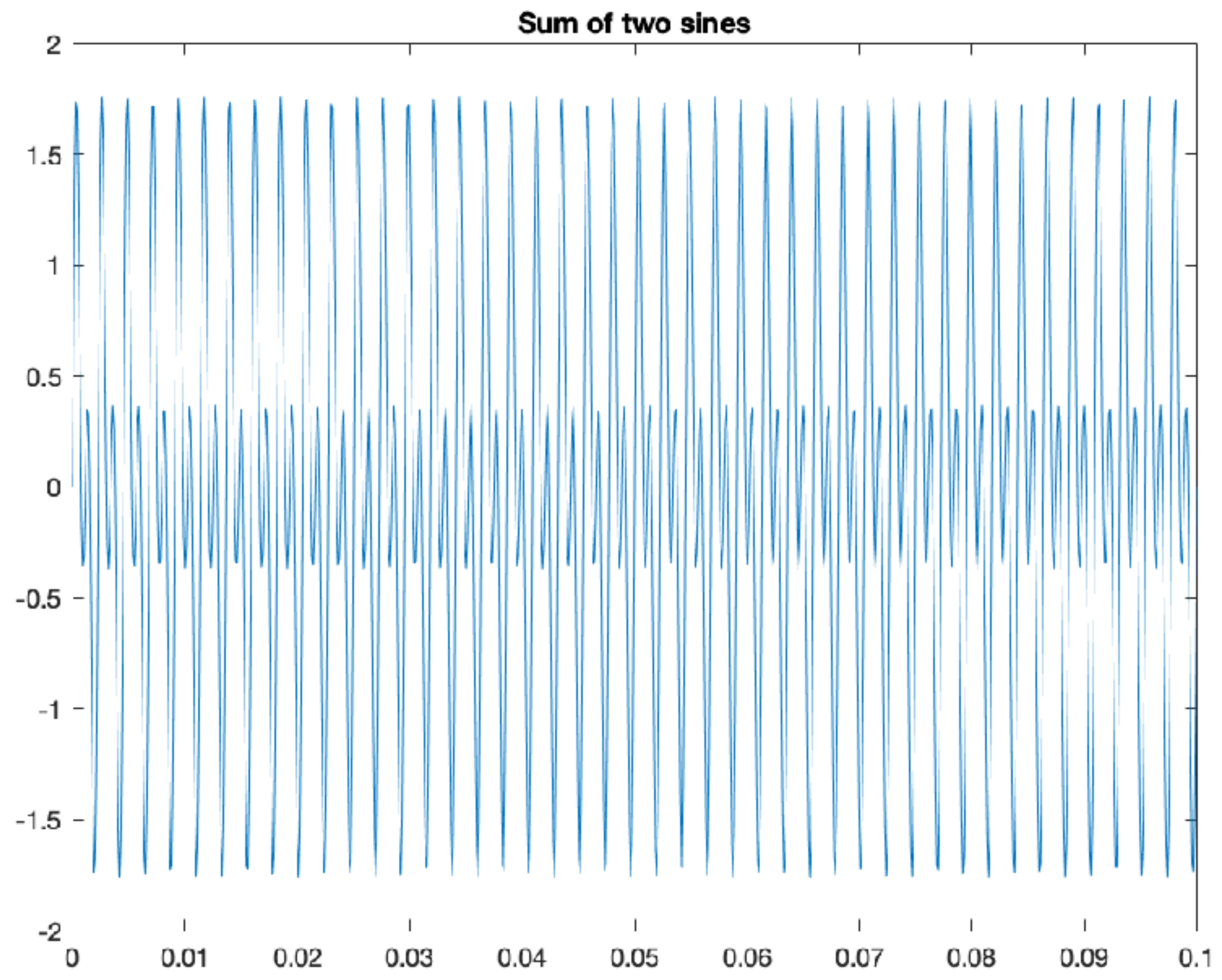
- Limitation of Fourier analysis:
  - We obtain a pure frequency content from a pure temporal content
  - What is the difference between a sum of sine, and a succession of sines?

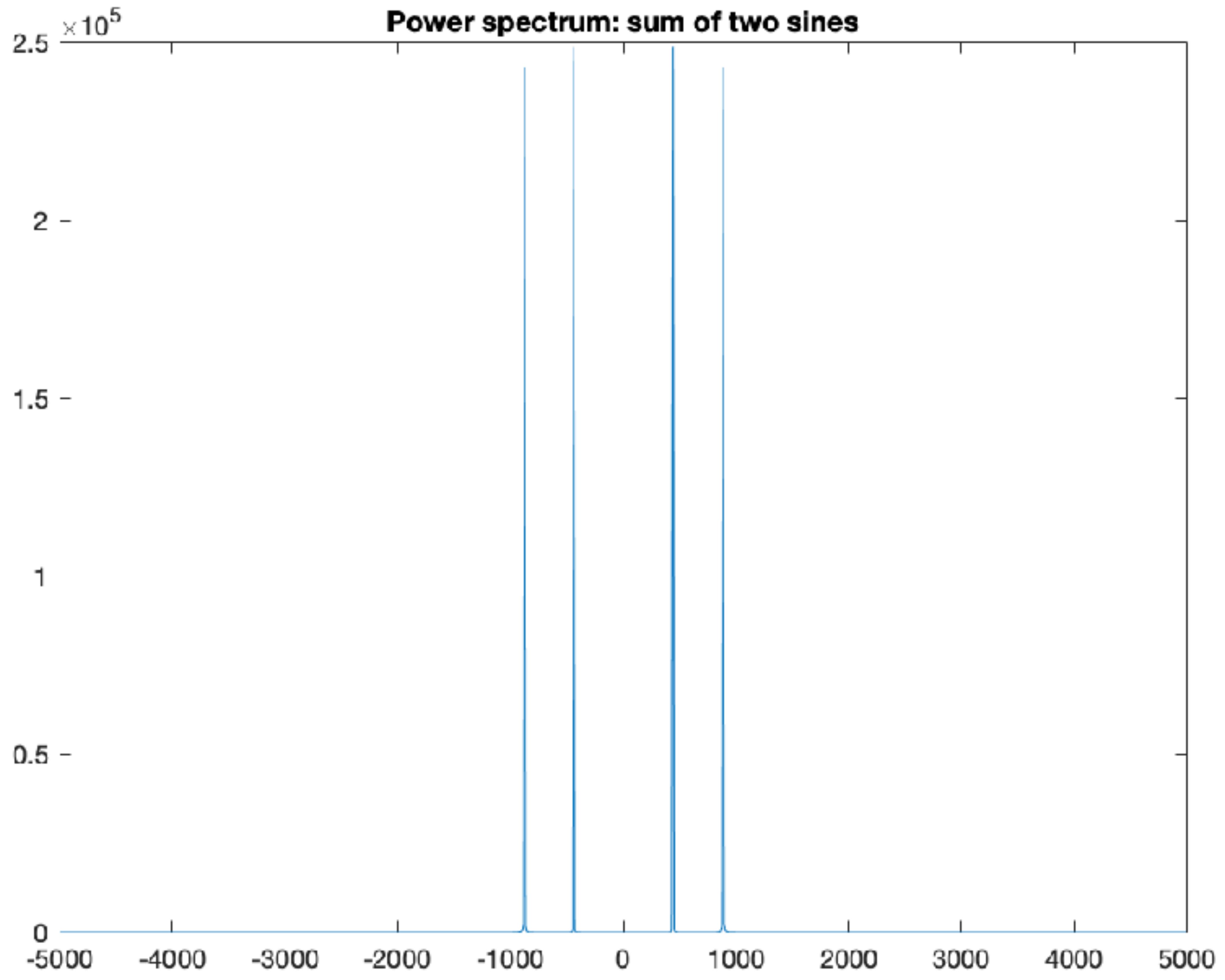
FOURIER EXAMPLE: 1 SINE



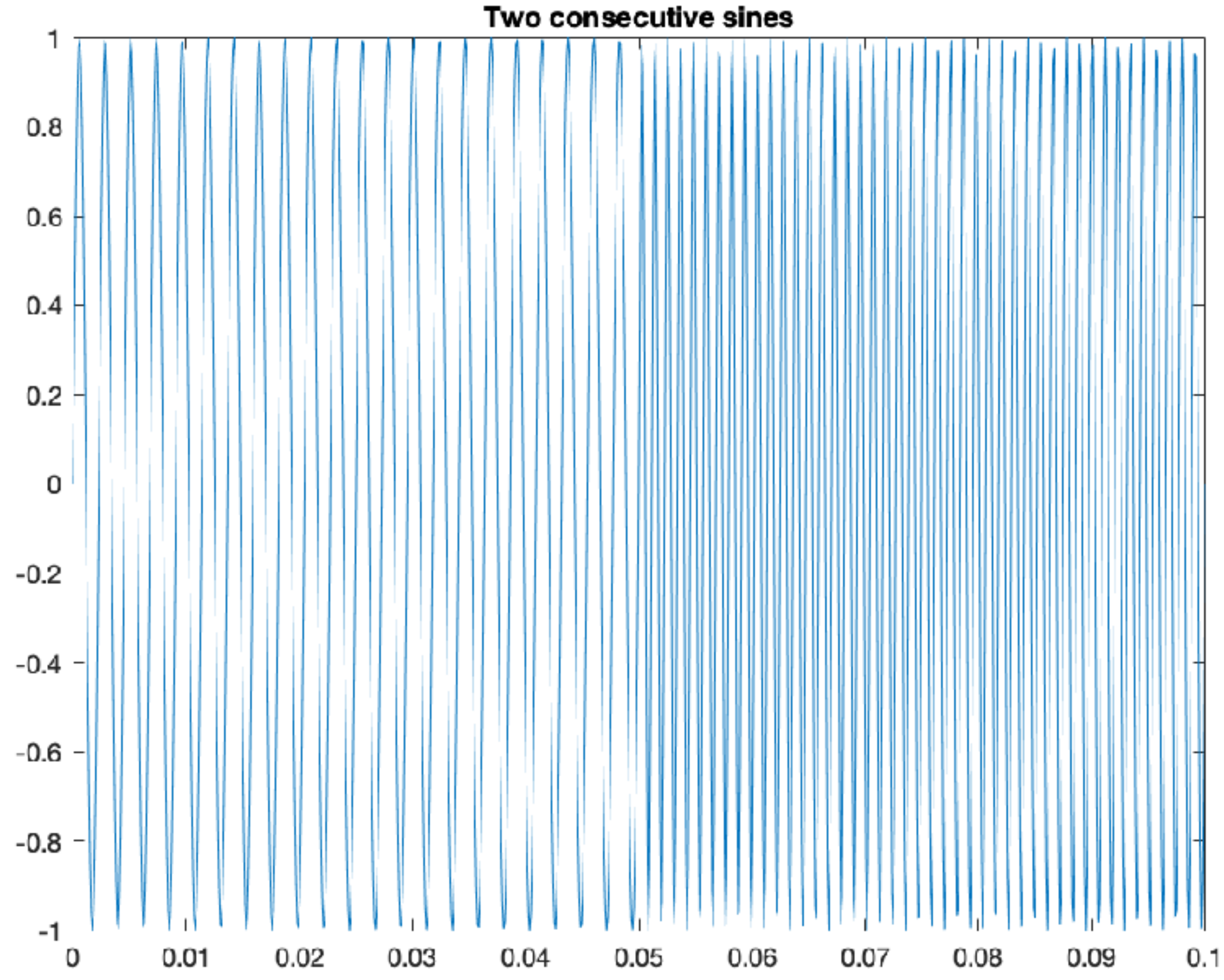


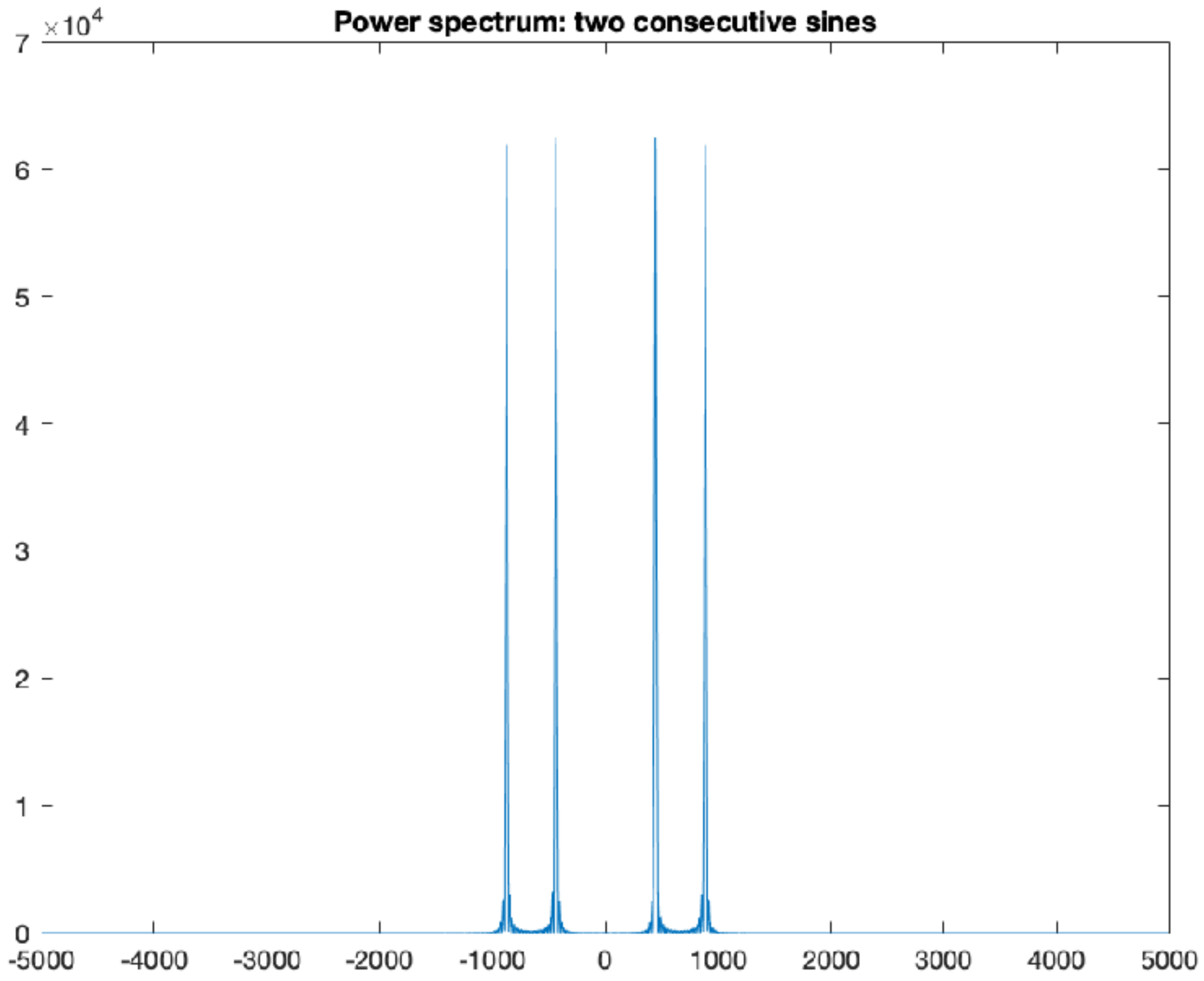
FOURIER EXAMPLE: SUM OF 2 SINES





FOURIER EXAMPLE: SUCCESSION OF 2 SINES





SHORT-TIME FOURIER TRANSFORM (STFT)

- Idea: perform a local spectral analysis of the signal thanks to a sliding window
- Let $w(t)$ be a smooth window localized around $t = 0$. Let the time-frequency atom

$$\varphi_{\tau,\nu}(t) = w(t-\tau)e^{i2\pi\nu t}$$

- The time-frequency transform of a signal x computes its correlation with the time-frequency atom $\varphi_{\tau,\nu}(t)$:

$$X(\tau,\nu) = \langle x, \varphi_{\tau,\nu}\rangle = \int_{-\infty}^{+\infty} x(t)w(t-\tau)e^{i2\pi\nu t}\mathrm{d}t$$

- It corresponds to the Fourier transform of the windowed signal $x(t)w(t-\tau)$
- Parameters of the STFT:
  - The length (and shape) of the window
  - The redundancy in time (hope size between two windows)
  - The redundancy in frequency (length of the frequency transform inside one window)

DISCRETE STFT (GABOR TRANSFORM)

- Let $x \in \mathbb{R}^T$ be a digital signal and let $w \in \mathbb{R}^L$ be a window. The discrete STFT is given by

$$X[\tau,\nu] = \sum_{t=0}^{N-1} x[t]w[t-a\tau]e^{i\frac{2\pi\nu}{M}t}$$

  - $a \geq 1$ control the redundancy in time (the hope size, in samples, between two windows)
  - $M \geq N$ control the redundancy in frequency (usually $M = L$ or $M = 2L$)
- Using the matrix notation, all the time-frequency coefficients $X[\tau,\nu]$ can be computed by the analysis operator Φ: $\boldsymbol{X} = \Phi^* x$
- Each column of Φ is one time-frequency atom.
- The number $K > T$ of columns depends on the time-frequency redundancy,

DISCRETE INVERSE STFT

- We do not have in general:

$$x[t] = \sum_{\tau,\nu} X[\tau,\nu] w[t - a\tau] e^{i\frac{2\pi\nu}{M}t}$$

- With matrix notation:

$$x \neq \Phi\Phi^* x$$

- The invert of a Gabor dictionary Φ is obtained by the canonical dual $\tilde{\Phi}$, which is also a Gabor transform constructed using a dual window $\tilde{w}$

$$x[t] = \sum_{\tau,\nu} X[\tau,\nu] \tilde{w}[t - a\tau] e^{i\frac{2\pi\nu}{M}t}$$

- With matrix notation:

$$x = \tilde{\Phi}\Phi^* x = \Phi\tilde{\Phi}^* x$$

- If the Gabor dictionary is a Parseval Frame (or a normalized tight frame), then $\tilde{\Phi} = \Phi$ and $\|x\|^2 = \|X\|^2$

SYNTHESIS OPERATION

- We have

$$x[t] = \sum_{\tau,\nu} \tilde{X}[\tau,\nu] w[t - a\tau] e^{i\frac{2\pi\nu}{M}t}$$

- With $\tilde{X} = \tilde{\Phi}^* x$, that is

$$x = \Phi \tilde{X}$$

- However, it exists an infinity of synthesis coefficients α such that

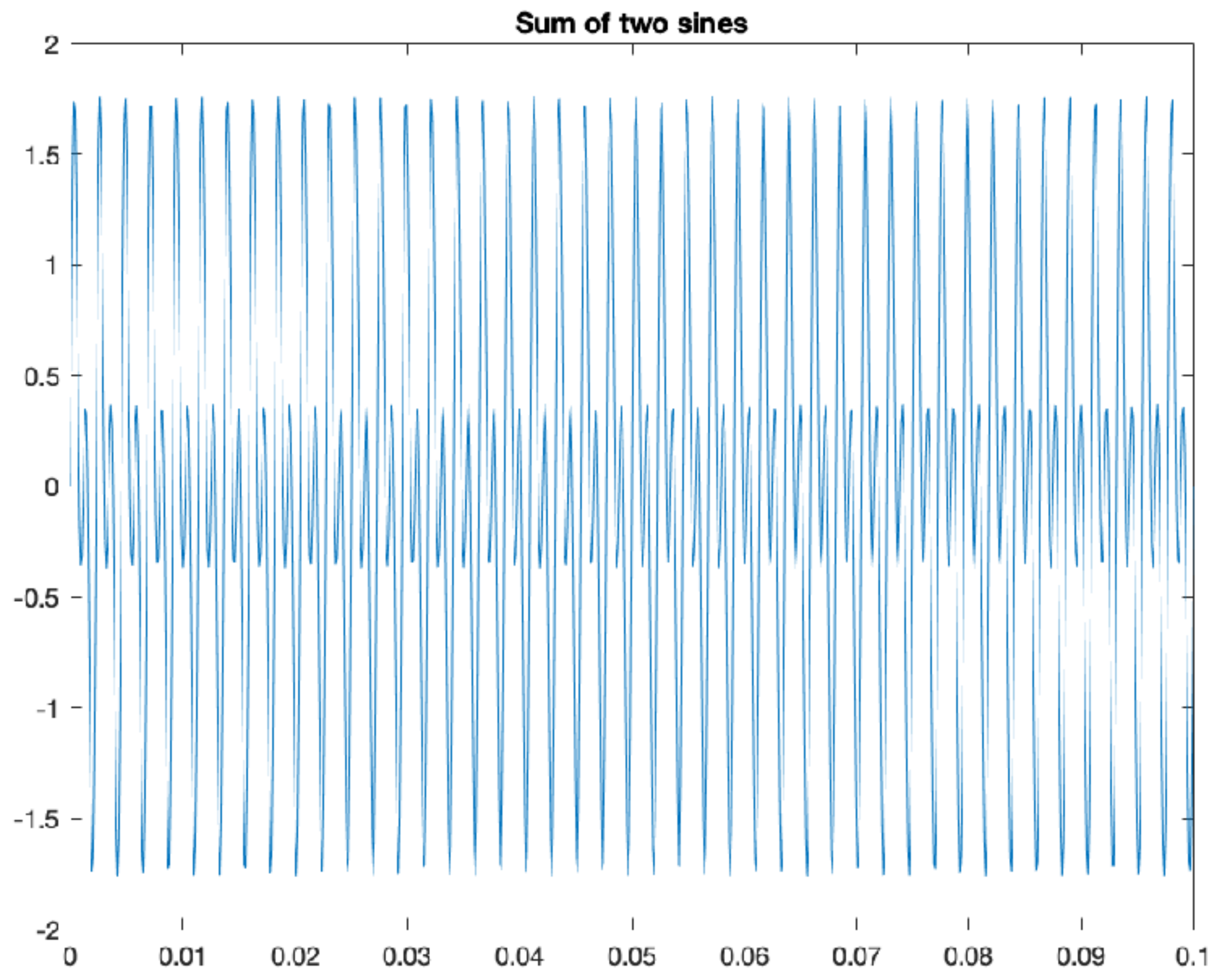
$$x = \Phi \alpha$$

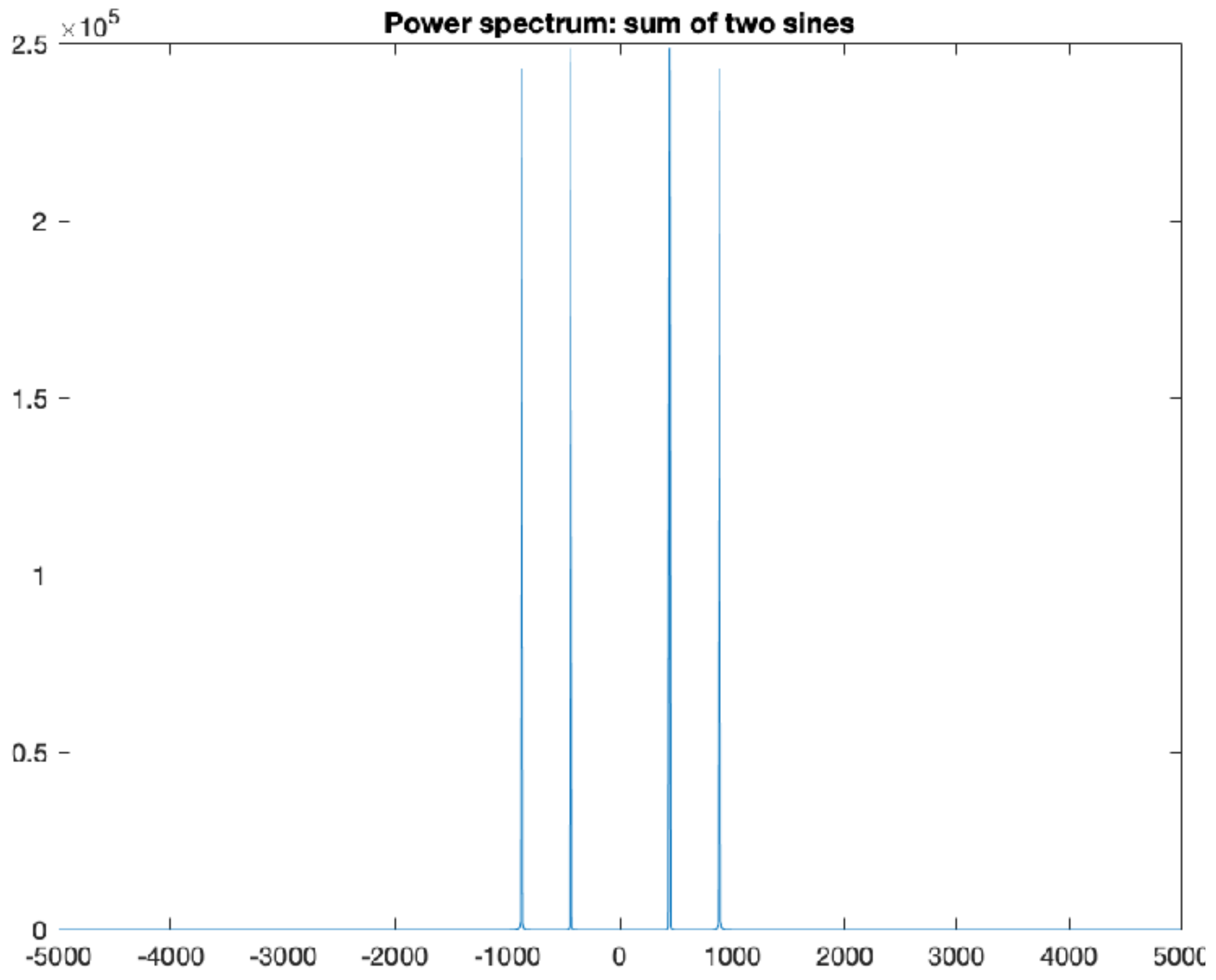
- Beware:
  - In some implementations, the "invert" operator is the "synthesis" operation and must be performed with the appropriate dual window
  - It is more useful to have access to the "synthesis" operator rather than the actual invert operator

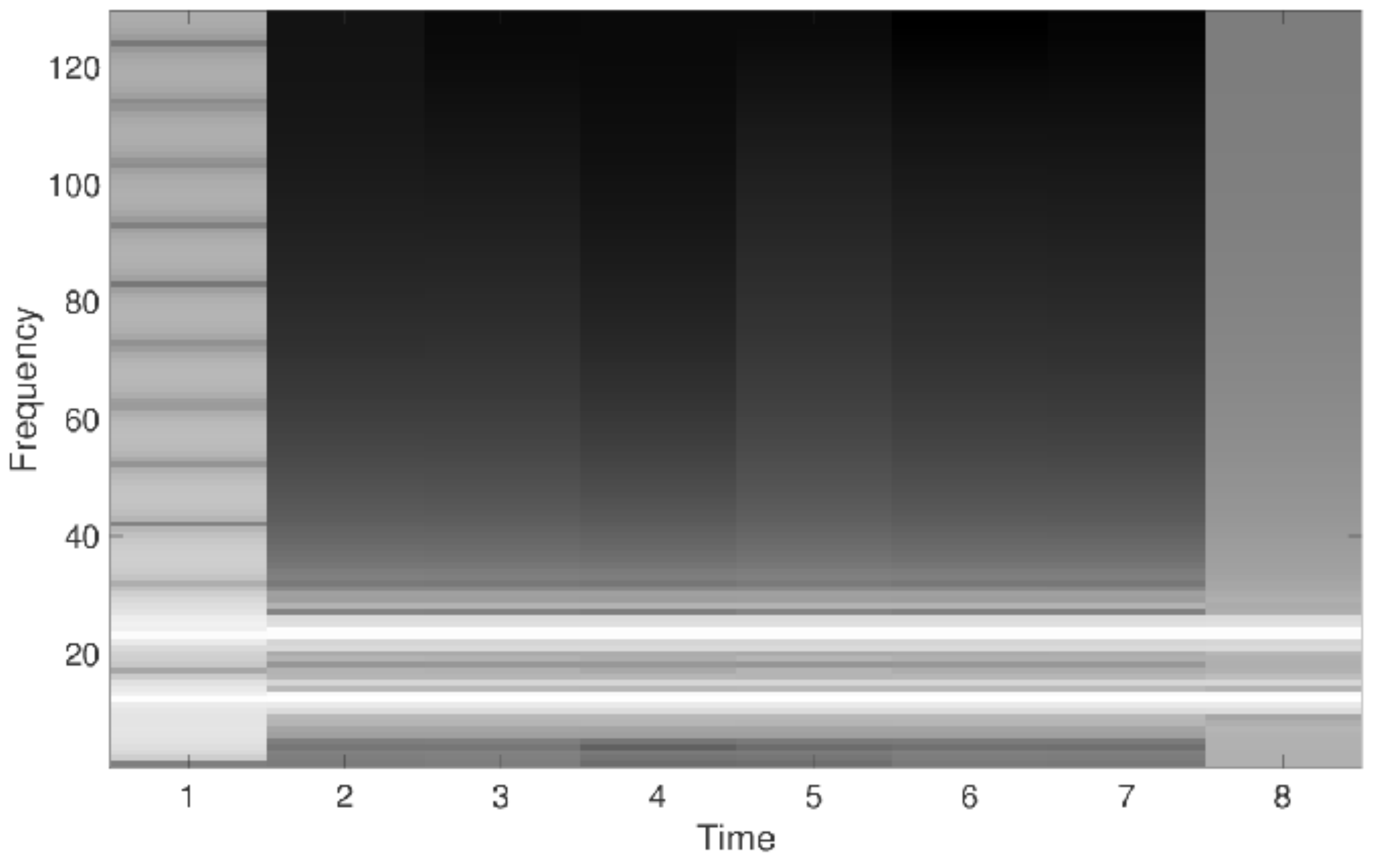
HOW TO CHOOSE THE PARAMETERS

- Heisenberg's uncertainty principle
  - A signal cannot be both well localized in time and in frequency.
  - Consequence: short windows are more adapted to "transient", and long windows to "tonal", "stationary", parts of the signal
- Common choices for a high-fidelity audio signal with a sampling frequency of 44.1 kHz with a window of size L
  - Shape of the window: Hann, Hamming, Gaussian
  - Length of the window: between 256 samples to 4096 samples. Common choice: 1024 samples
  - Redundancy in time: overlap of 50% or 75% between two consecutive window
  - Redundancy in frequency: FFT of size L or $2L$

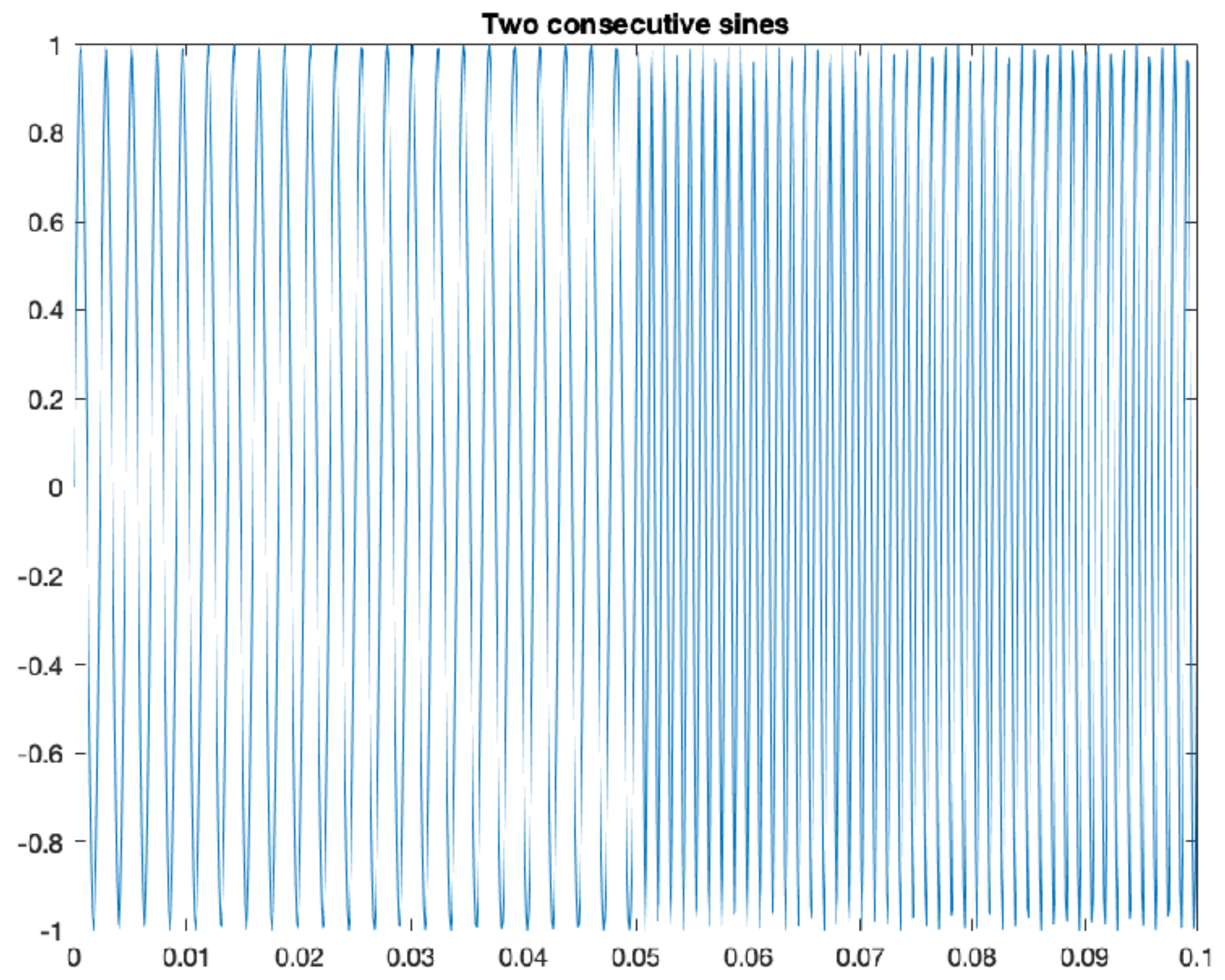
STFT EXAMPLE: SUM OF 2 SINES

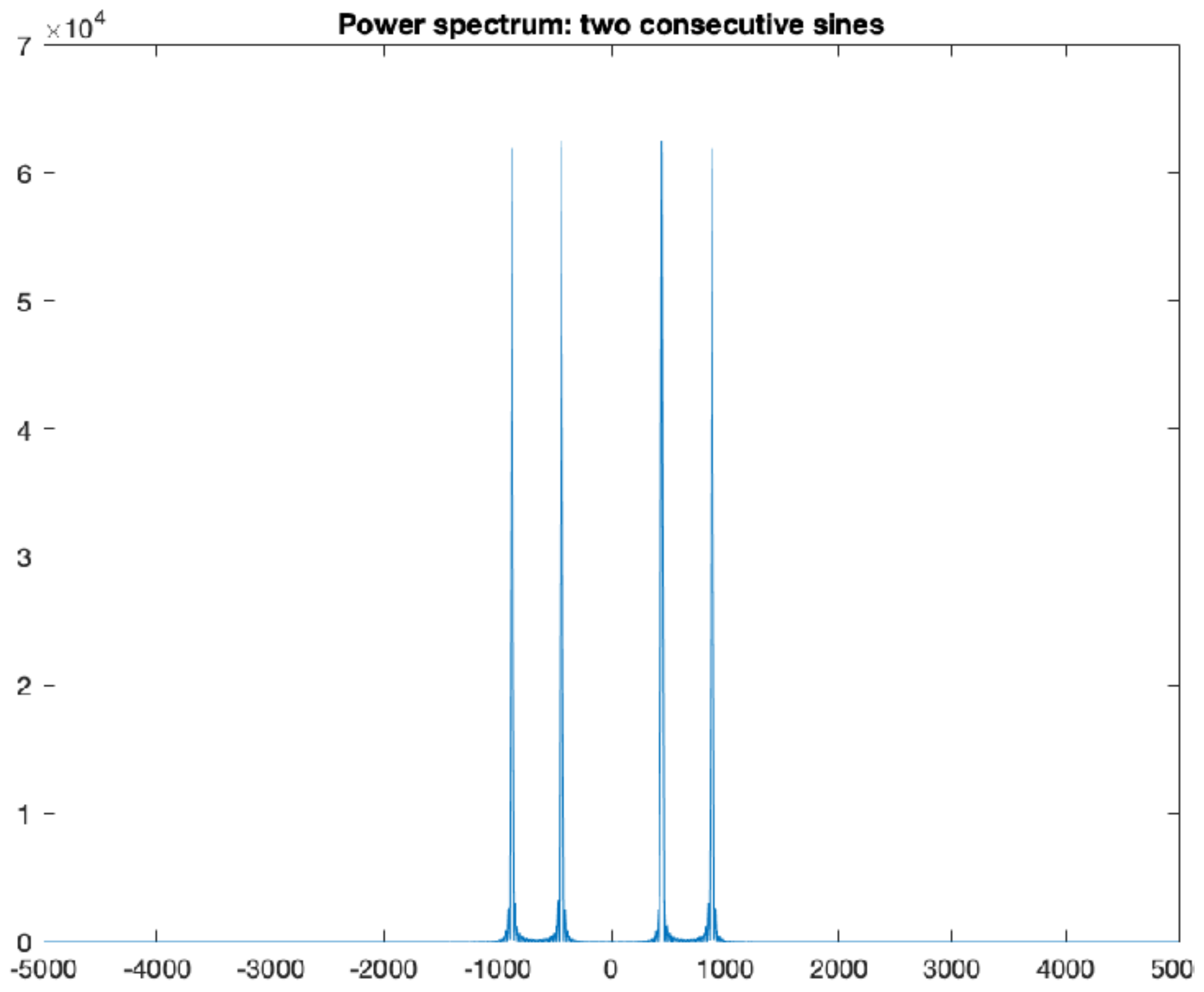


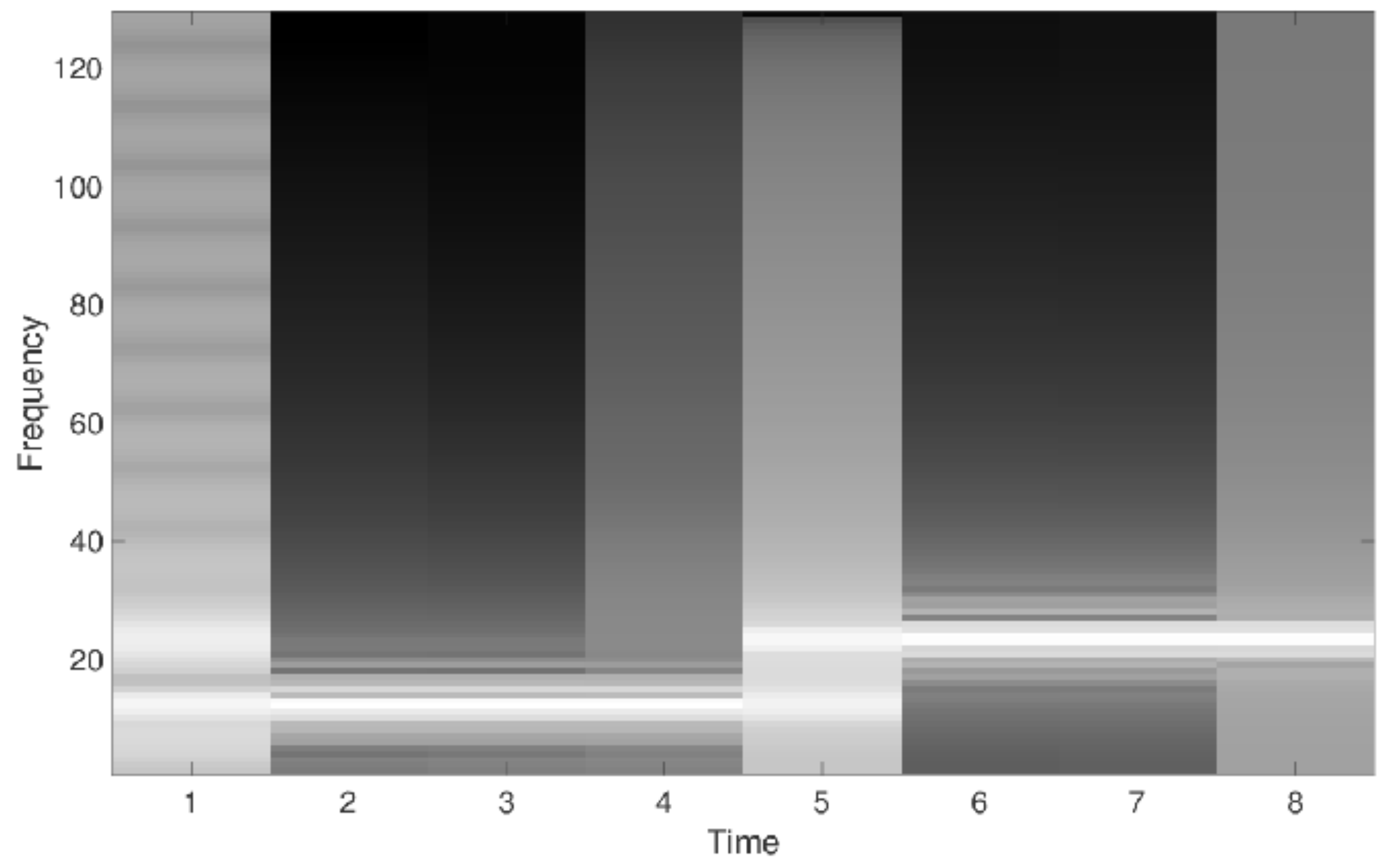




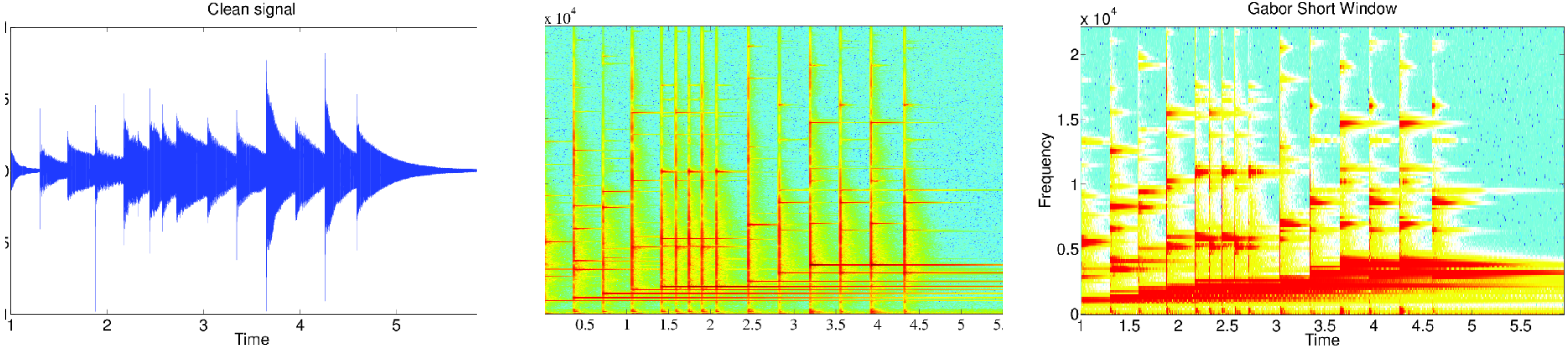
STFT EXAMPLE: SUCCESSION OF 2 SINES







STFT EXAMPLE: GLOCKENSPIEL



DENOISING IN THE TIME–FREQUENCY DOMAIN

- Let y be a noisy measure of a "clean" signal x corrupted by some additive noise n:

$$y = x + n$$

- In the STFT domain, we have

$$Y[\tau, \nu] = X[\tau, \nu] + N[\tau, \nu]$$
$$\mathbb{E}\left\{ |Y[\tau, \nu|^2 \right\} = |X[\tau, \nu]|^2 + S_n[\nu]$$

- Proposed estimator
  - Hard Thresholding

$$X(\tau, \nu) = \begin{cases} Y(\tau, \nu) & \text{if } |Y(\tau, \nu)| > \lambda |S_n(\nu)| \\ 0 & \text{if } |Y(\tau, \nu)| \leq \lambda |S_n(\nu)| \end{cases}$$

  - Spectral subtraction

$$X(\tau, \nu) = Y(\tau, \nu)\left(1 - \frac{\lambda^2 |S_n(\nu)|^2}{|Y(\tau, \nu)|^2}\right)^+$$

TO DO: DENOISING IN THE STFT DOMAIN

- Data
  - The 3 noises of the random chapter
  - “Clean” music signal
- Todo
  - Simulate a noisy version of the music using the noises at various SNR Level
  - Implement the denoising by hard thresholding and spectral subtraction
  - Denoise the different given noisy version of the clean signal
  - Discuss the parameters