## FROM ANALYSIS TO SPARSE SYNTHESIS

# SPARSE REPRESENTATIONS

# PRESENTATION

- Mail: matthieu.kowalski@universite-paris-saclay.fr
- Web site: http://hebergement.universite-paris-saclay.fr/mkowalski/

# REMINDER

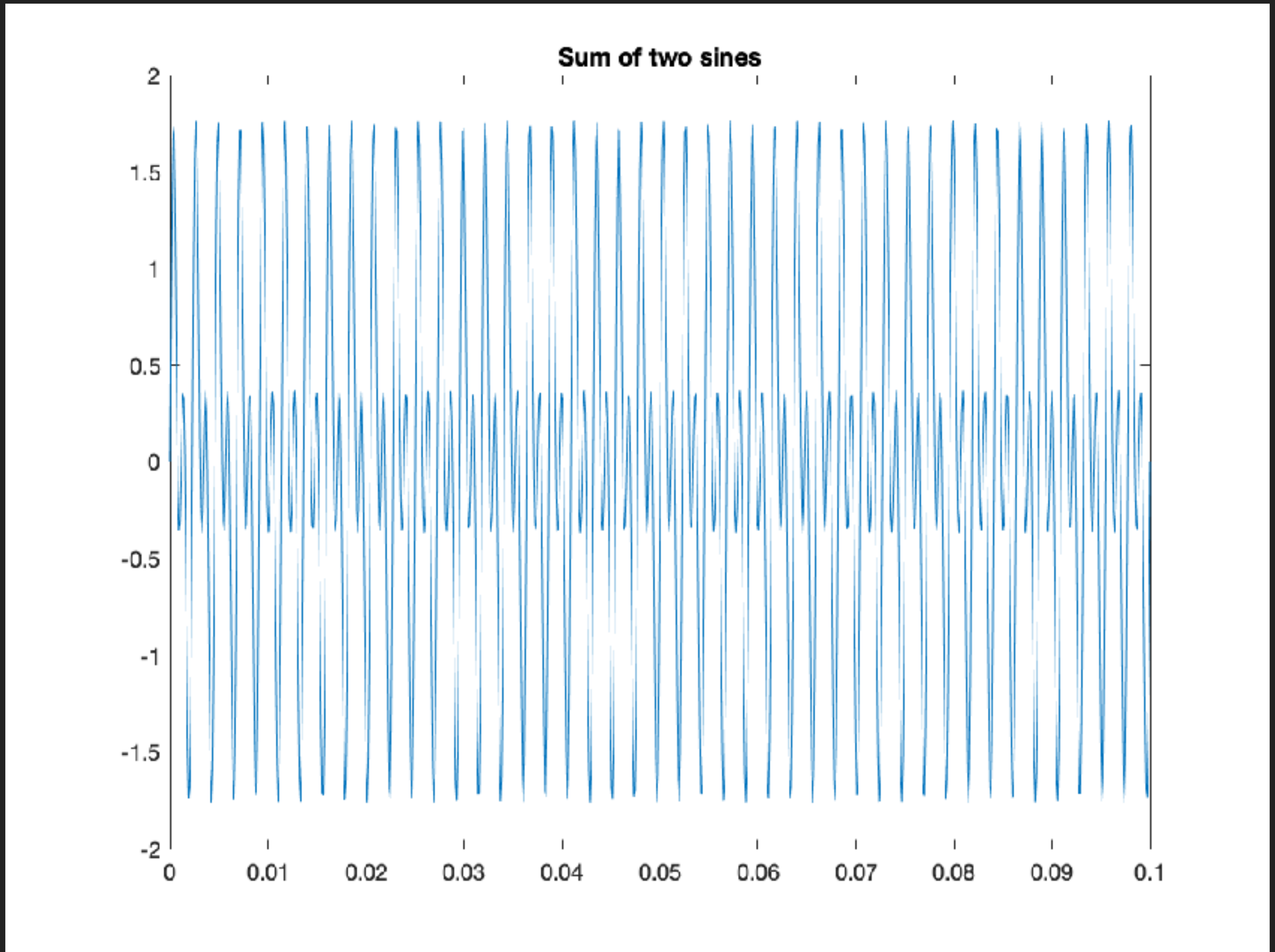
# FOURIER

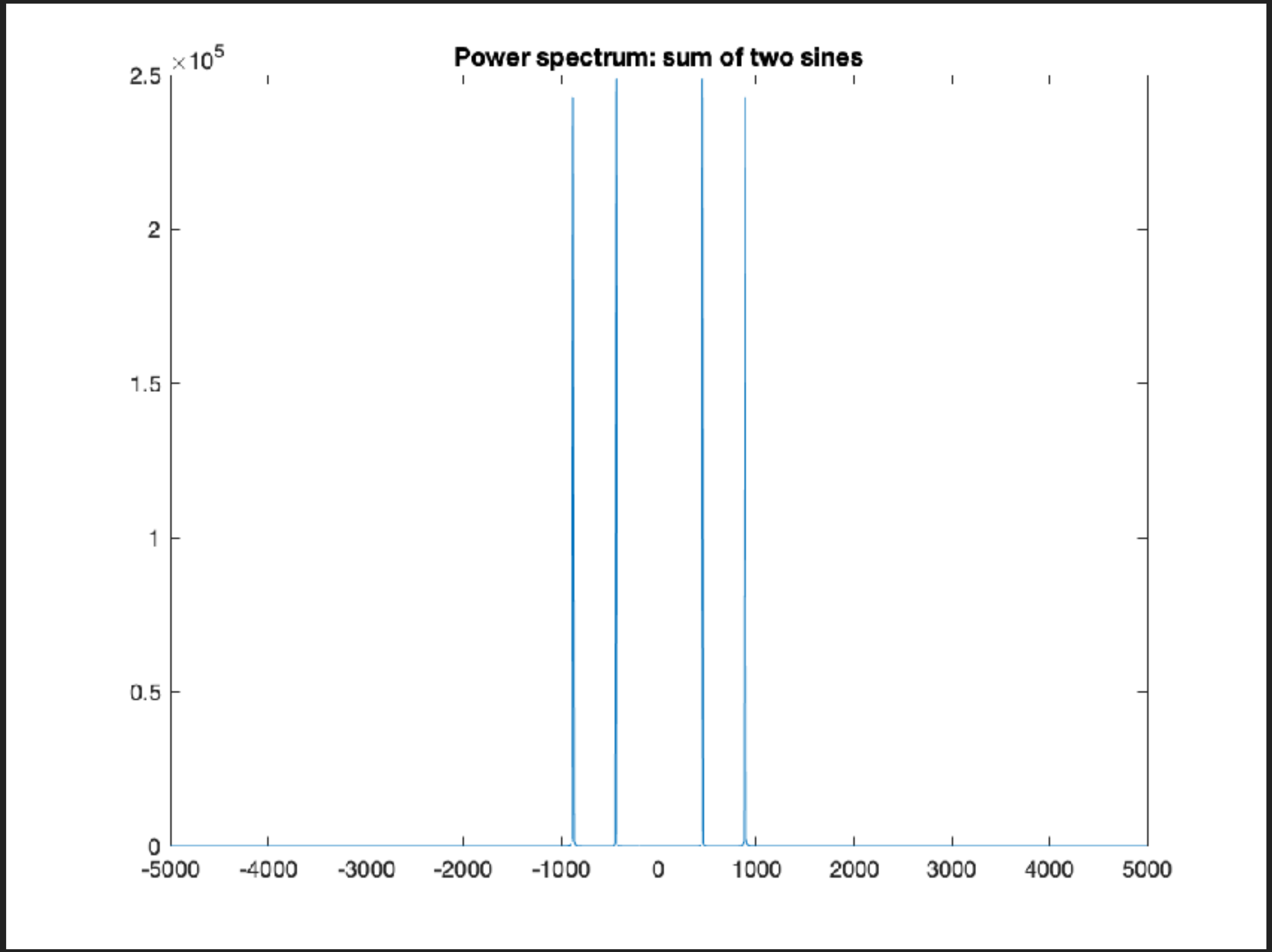
- From time representation to frequency
  - Spectral analysis : frequency content of a function (think about musical notes)
  - Measure the similute (correlation, angle) between pure (complex) sine and a signal
  - Sines are eigen signal of time invariant linear systems (filters)
  - Fourier analysis computes the correlation between the signal $s(t)$ and a pure (complex) sine of frequency $\nu$:

$$\langle s(t), \epsilon_\nu(t) \rangle, \quad \text{with } \epsilon_\nu(t) = e^{i2\pi\nu t}$$

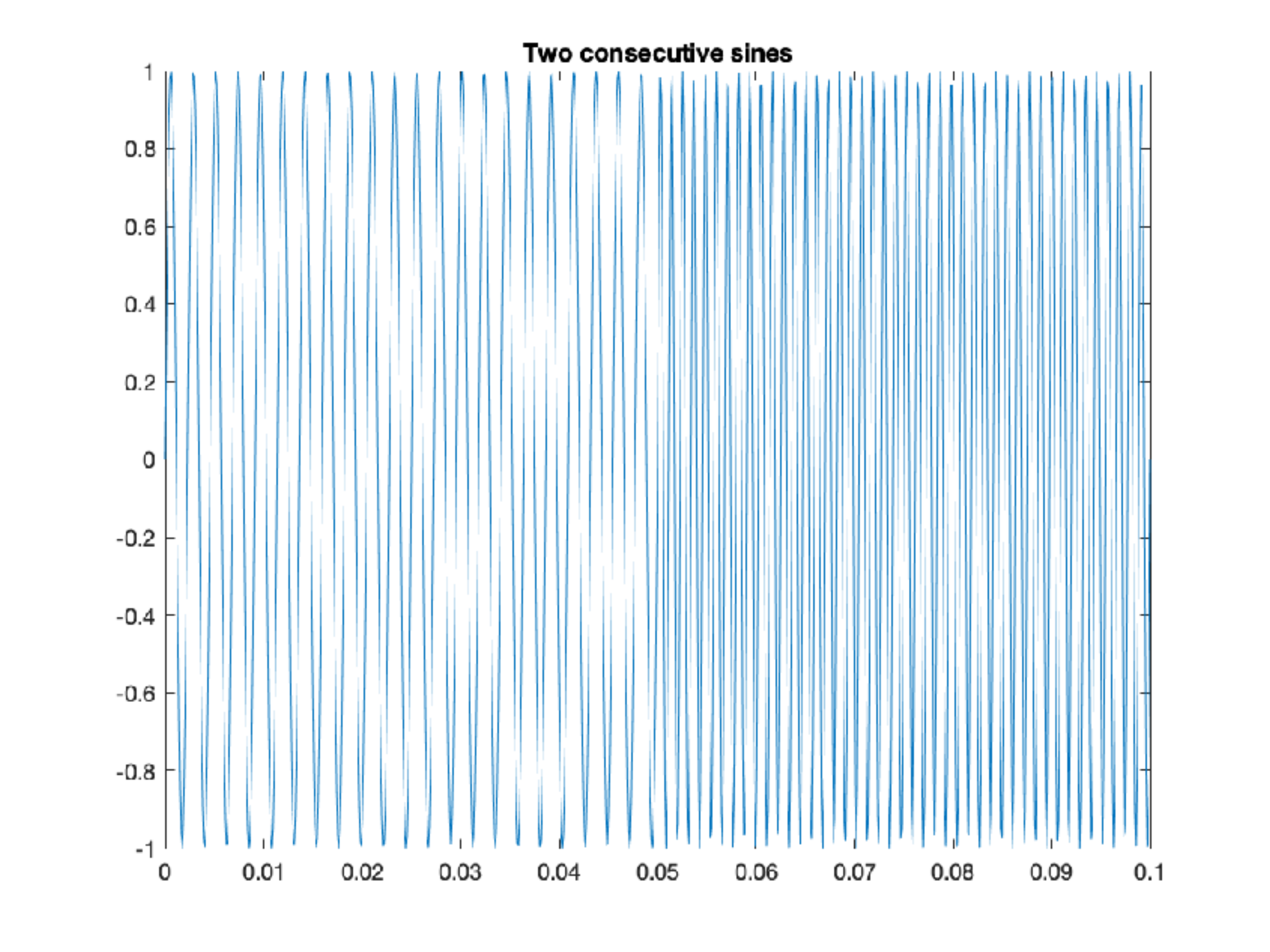
- Limitation of Fourier analysis
  - We obtain a pure frequency content from a pure temporal content
  - What is the difference between a sum of sine, and a succession of sine ?

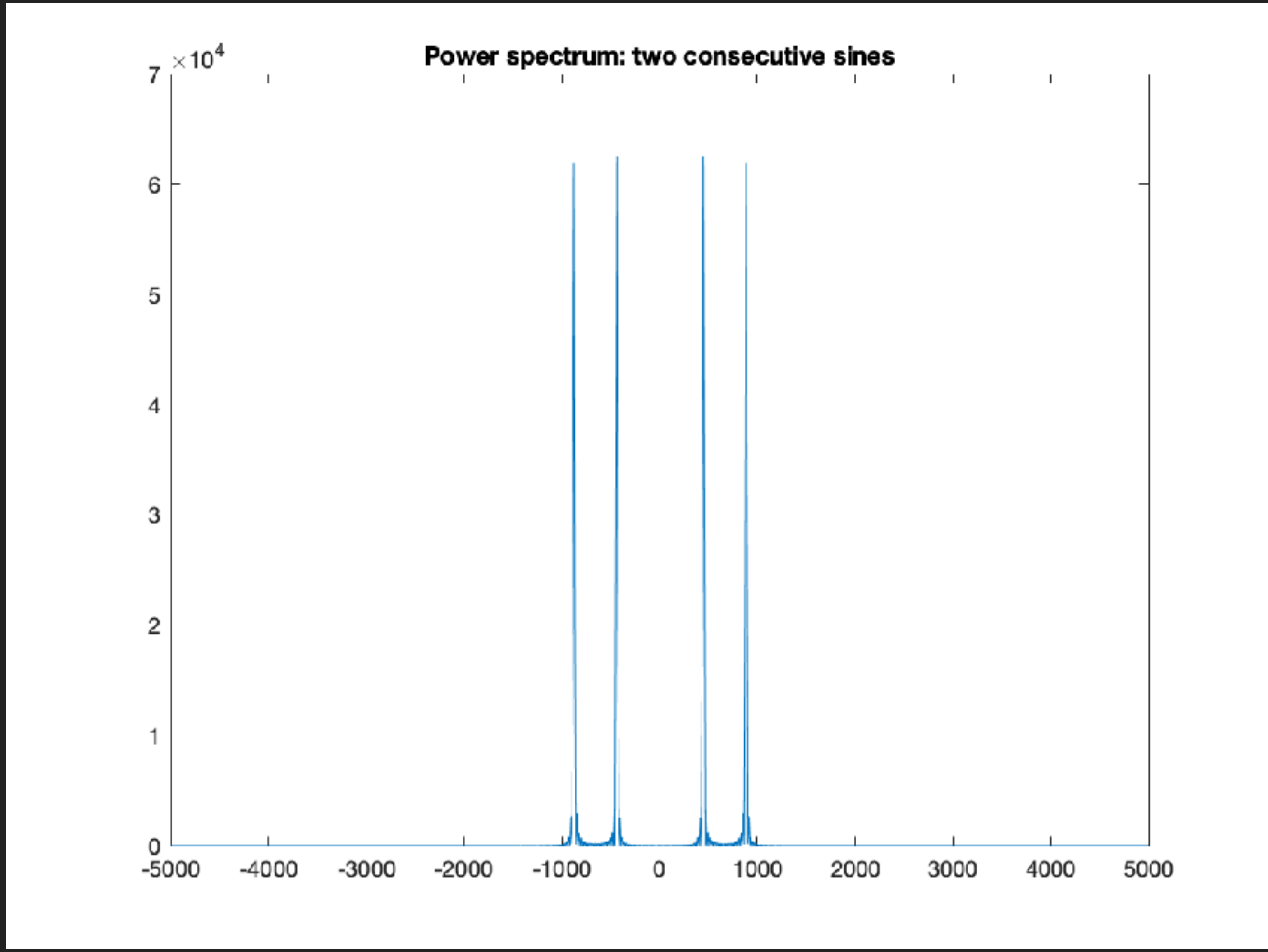
# EXAMPLE: SUM OF 2 SINES



# EXAMPLE: SEQUENCE 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 real smooth window localized around $t = 0$. Let the time-frequency atoms

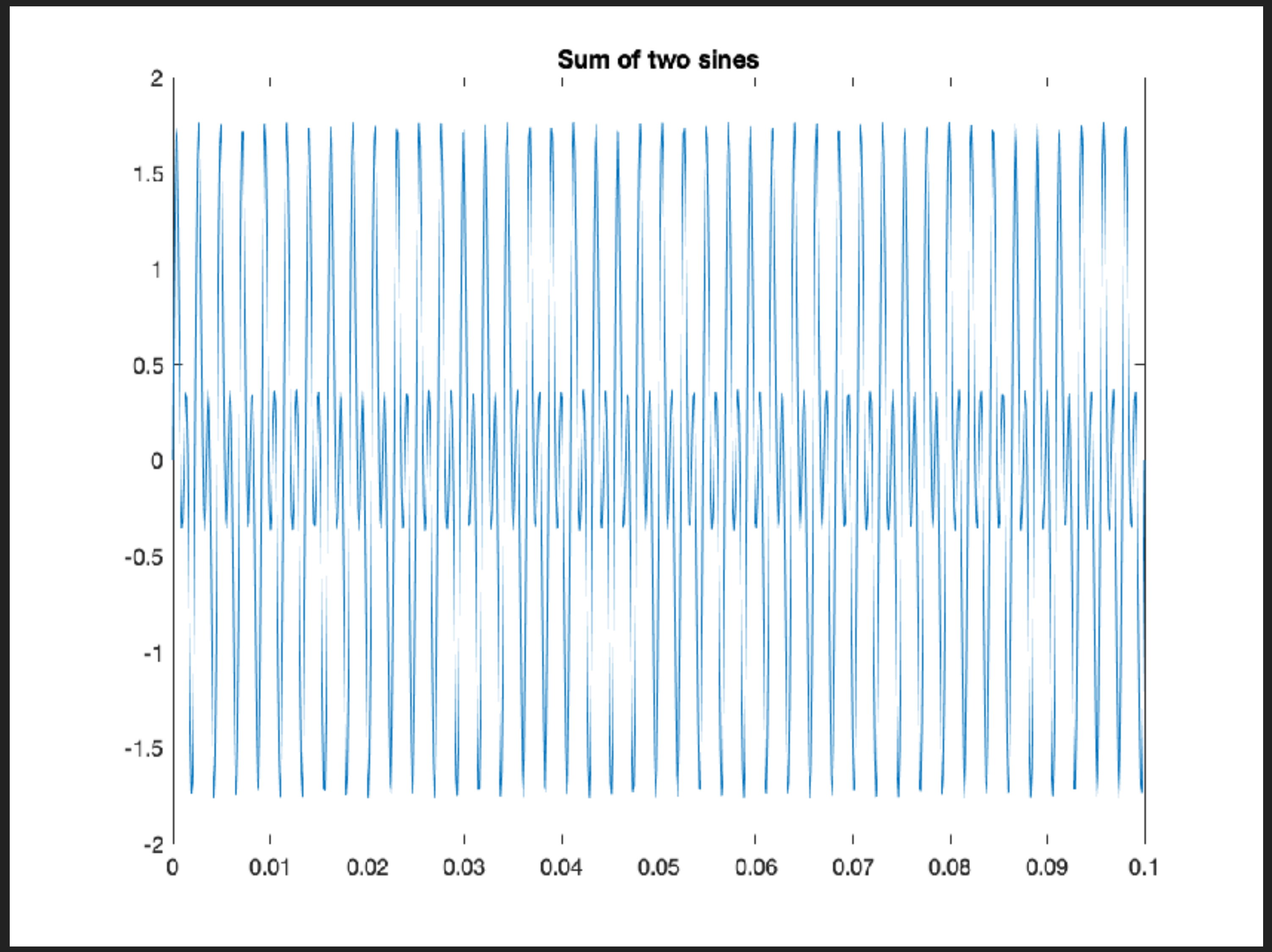
$$\varphi_{\tau,\nu}(t) = w(t-\tau)e^{i2\pi\nu t}$$

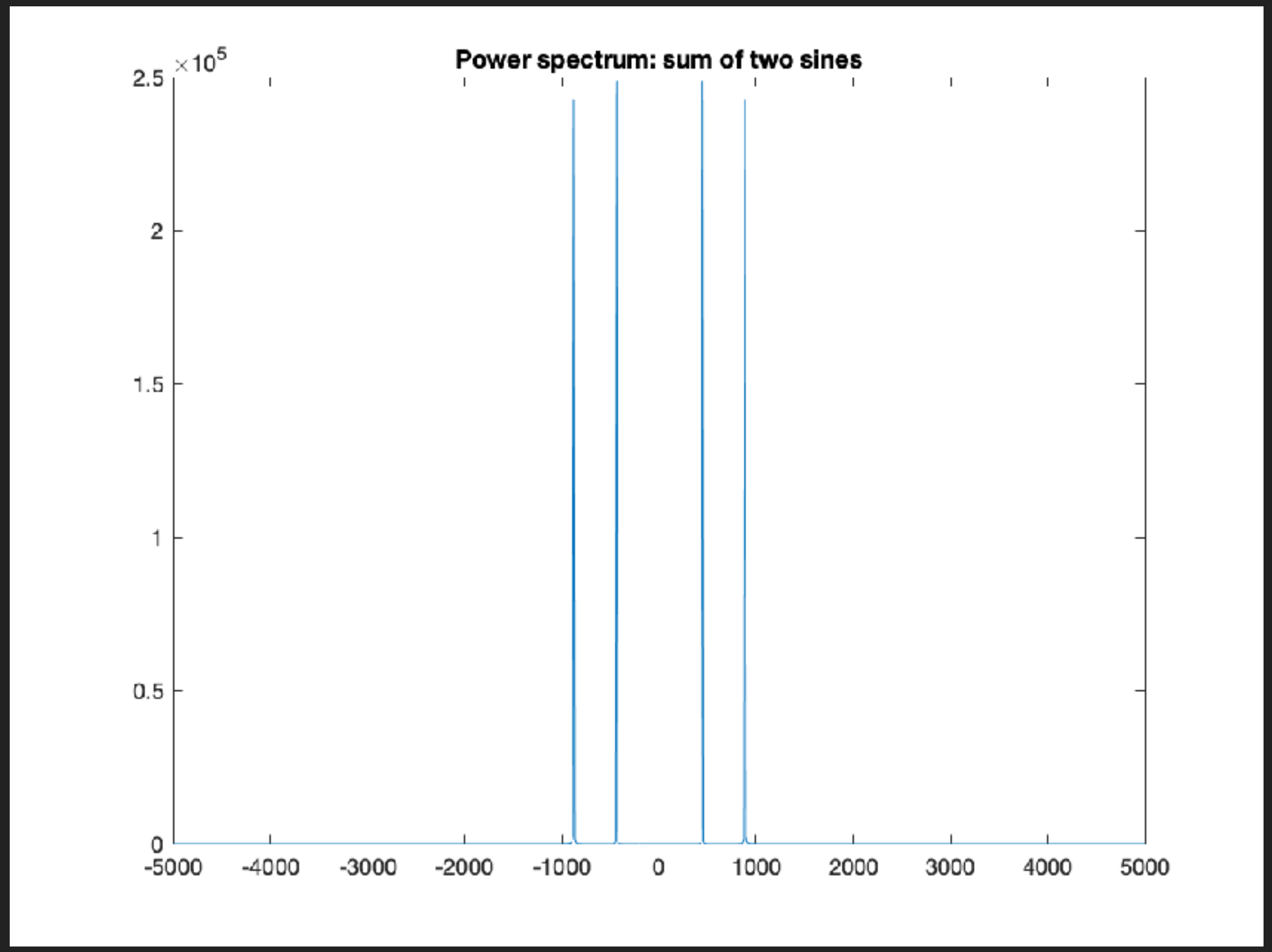
The **stft** transform of a signal $x(t)$ computes the correlation between $x(t)$ and the time-frequency atoms $\varphi_{\tau,\nu}(t) = w(t-\tau)e^{i2\pi\nu t}$:

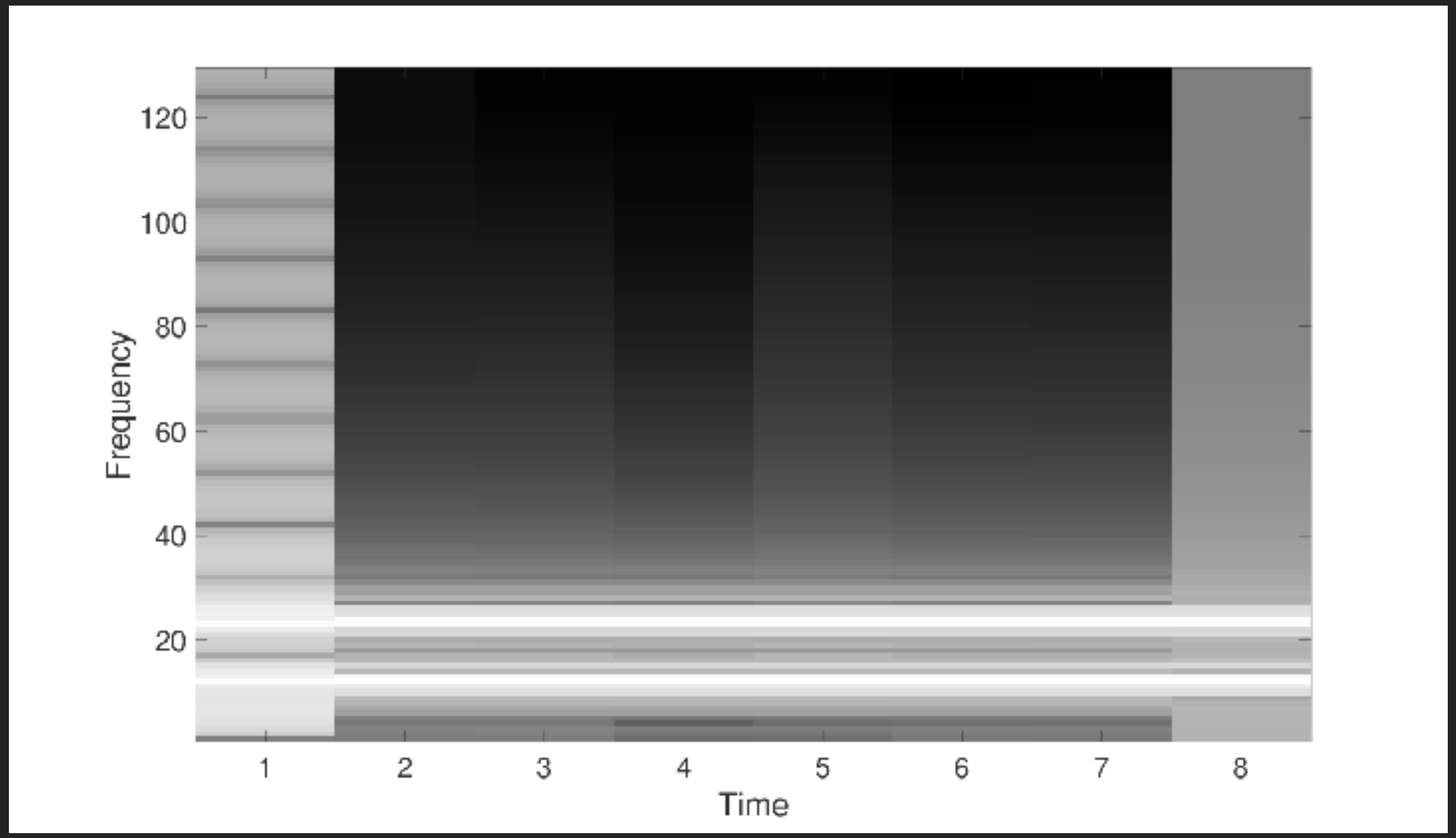
$$X(\tau,\nu) = \langle x(t), \varphi_{\tau,\nu}(t)\rangle = \int_{-\infty}^{+\infty} x(t)w(t-\tau)e^{-i2\pi\nu t}\mathrm{d}t$$

- It is a **time-frequency** transform
- Window choice: Heisenberg's uncertainty principle
- It is invertible: $x(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} X(\tau,\nu)w(t-\tau)e^{-i2\pi\nu\tau}\mathrm{d}\tau\mathrm{d}\nu$
- We have energy preservation

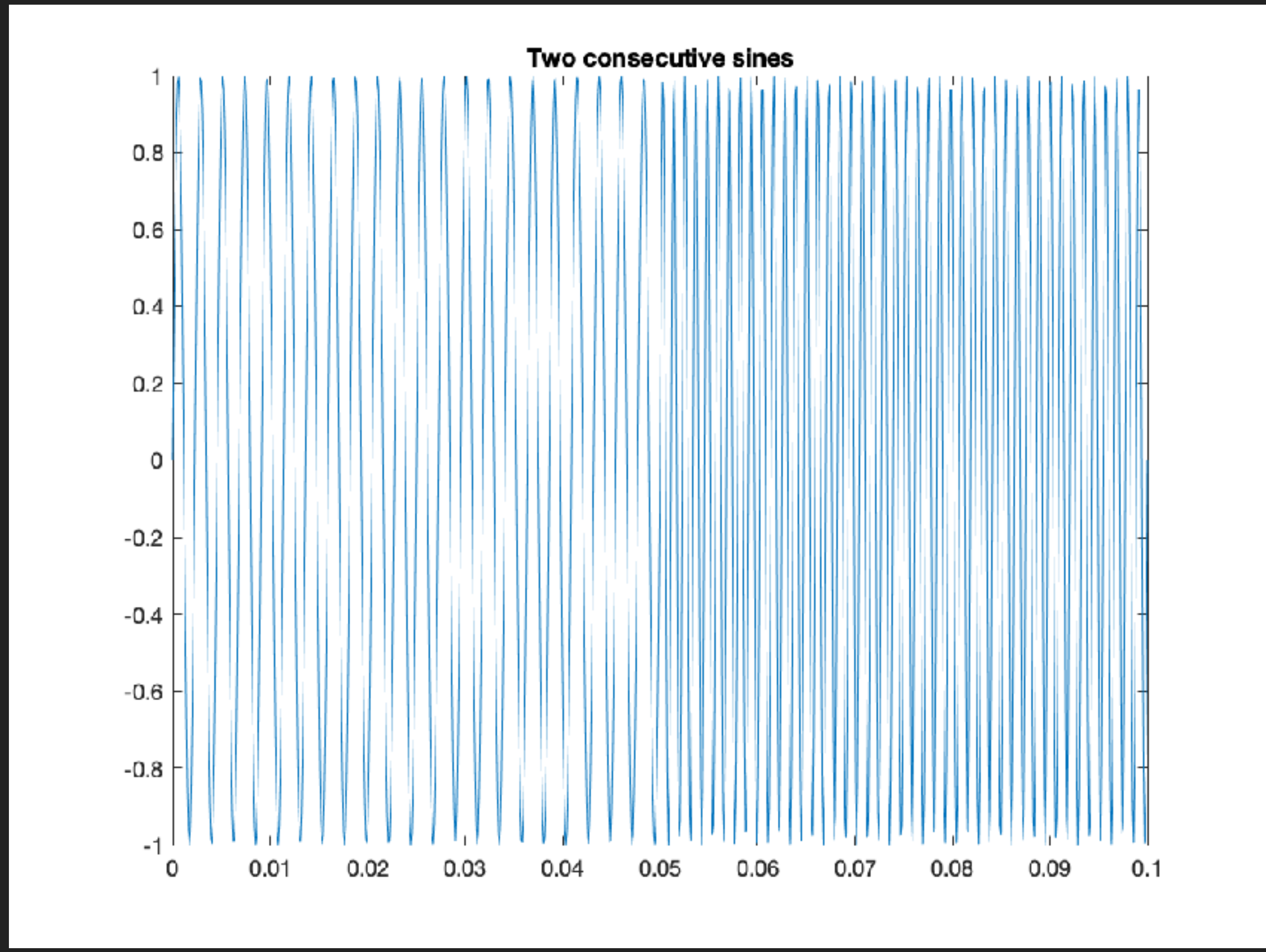
# EXAMPLE: SUM OF 2 SINES

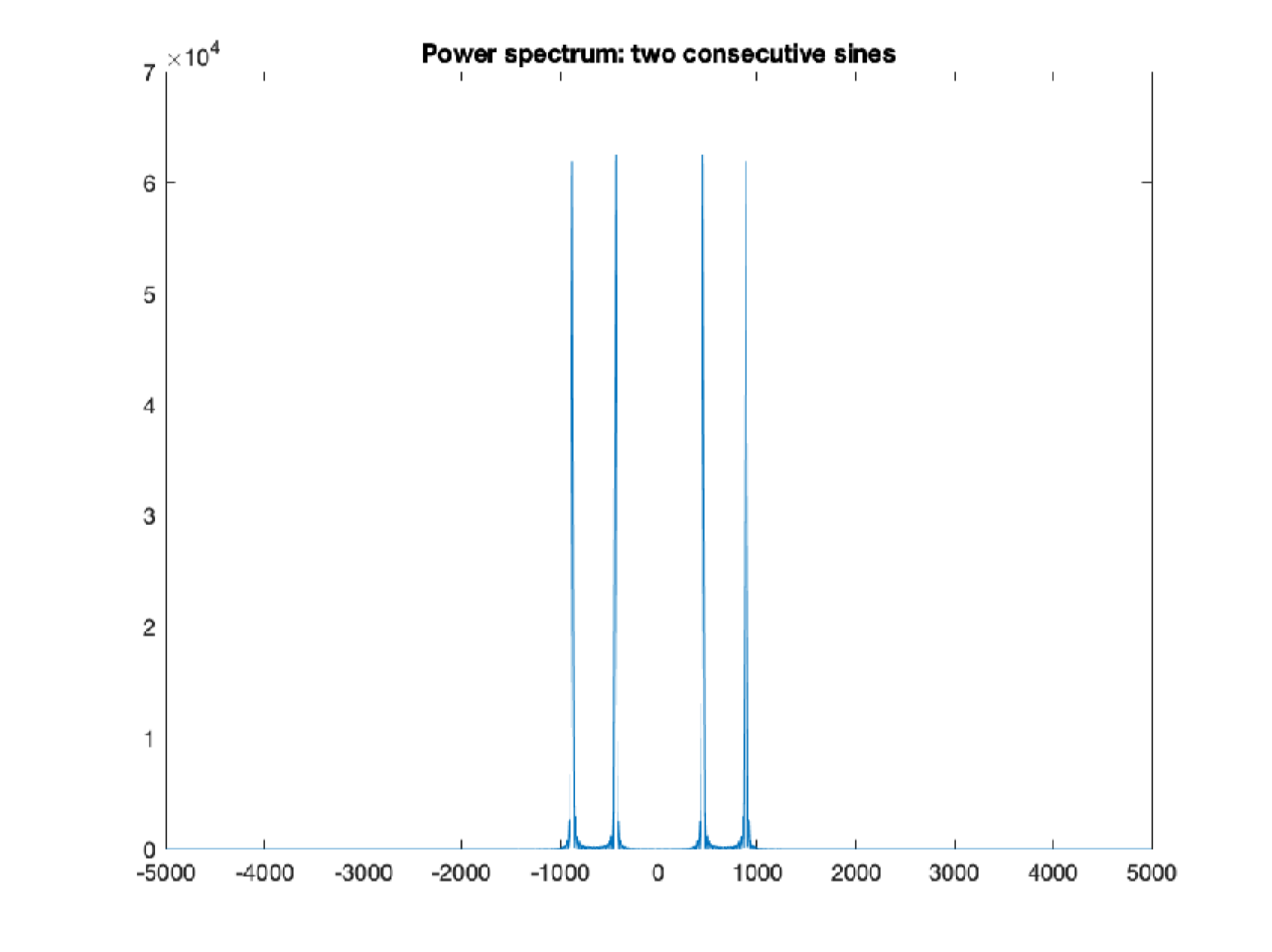


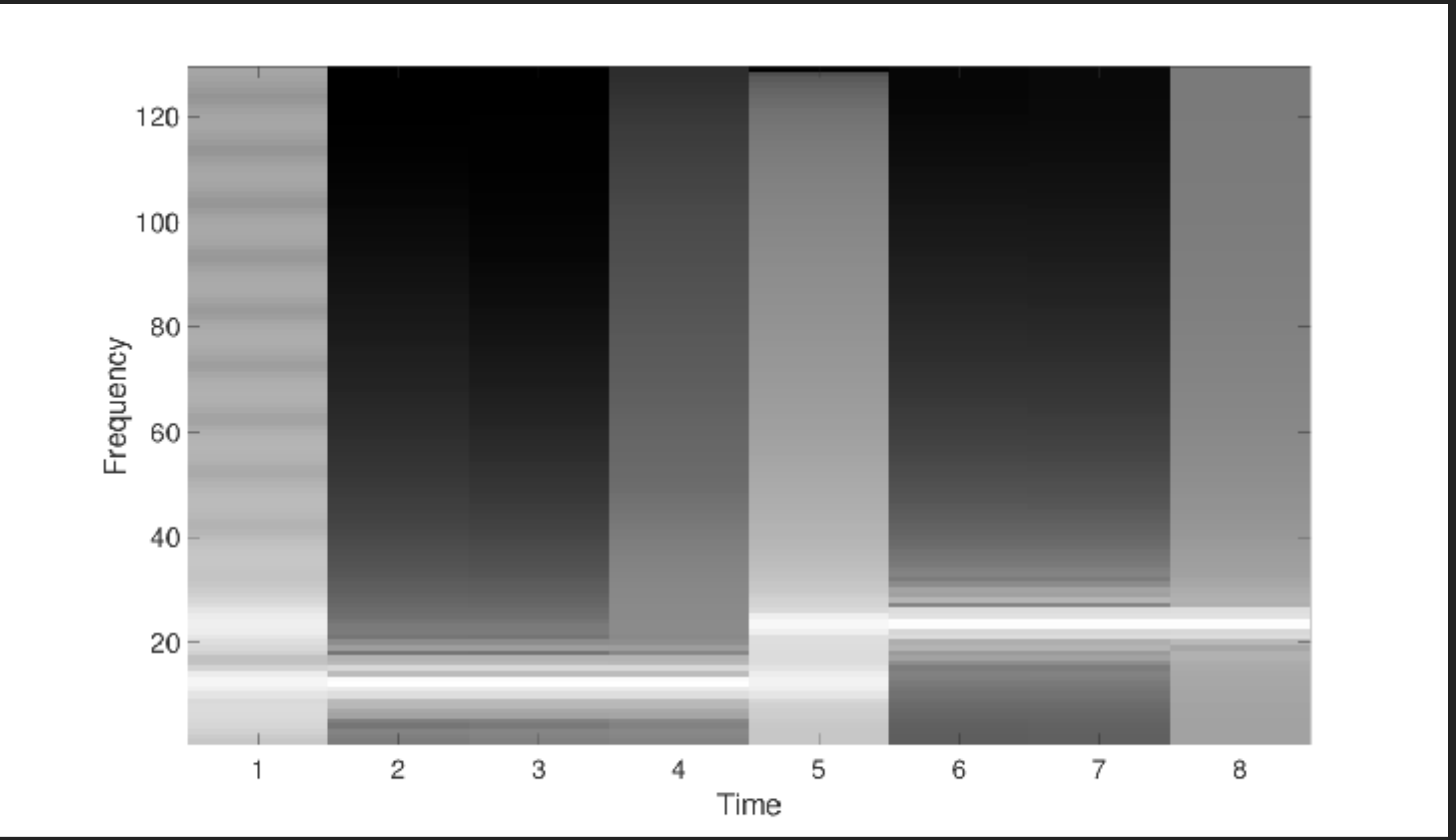



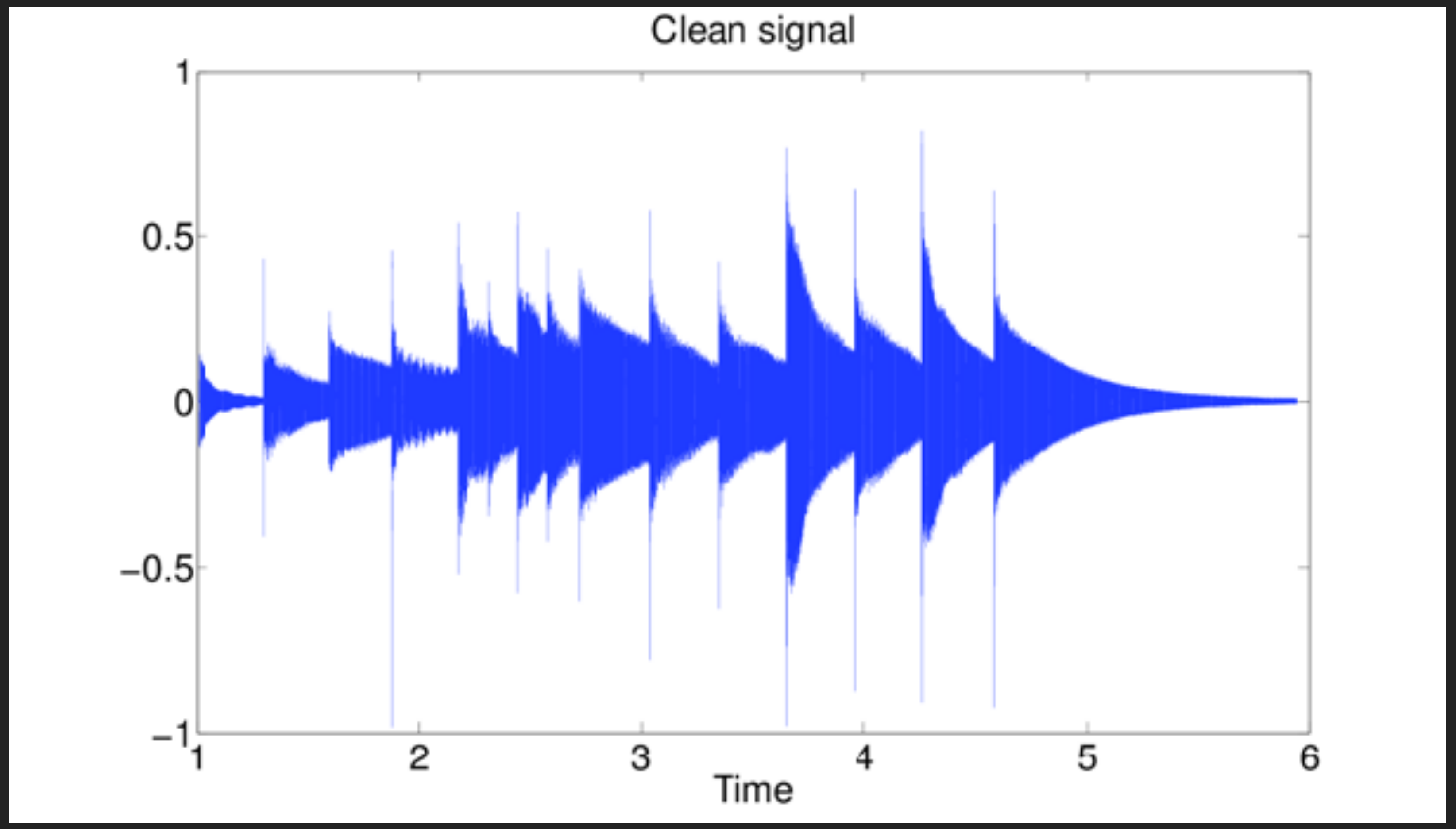
# EXAMPLE: SEQUENCE OF 2 SINES

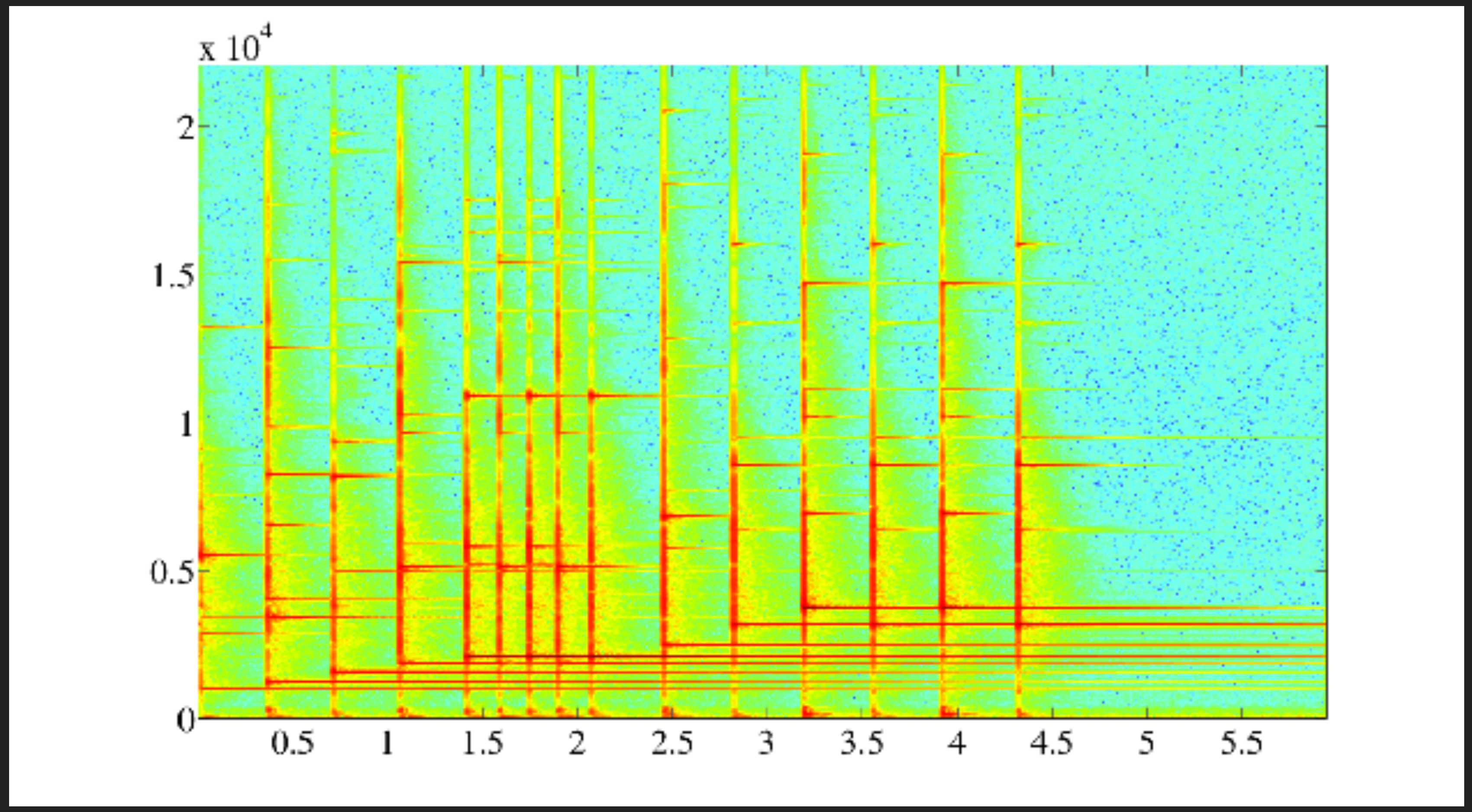


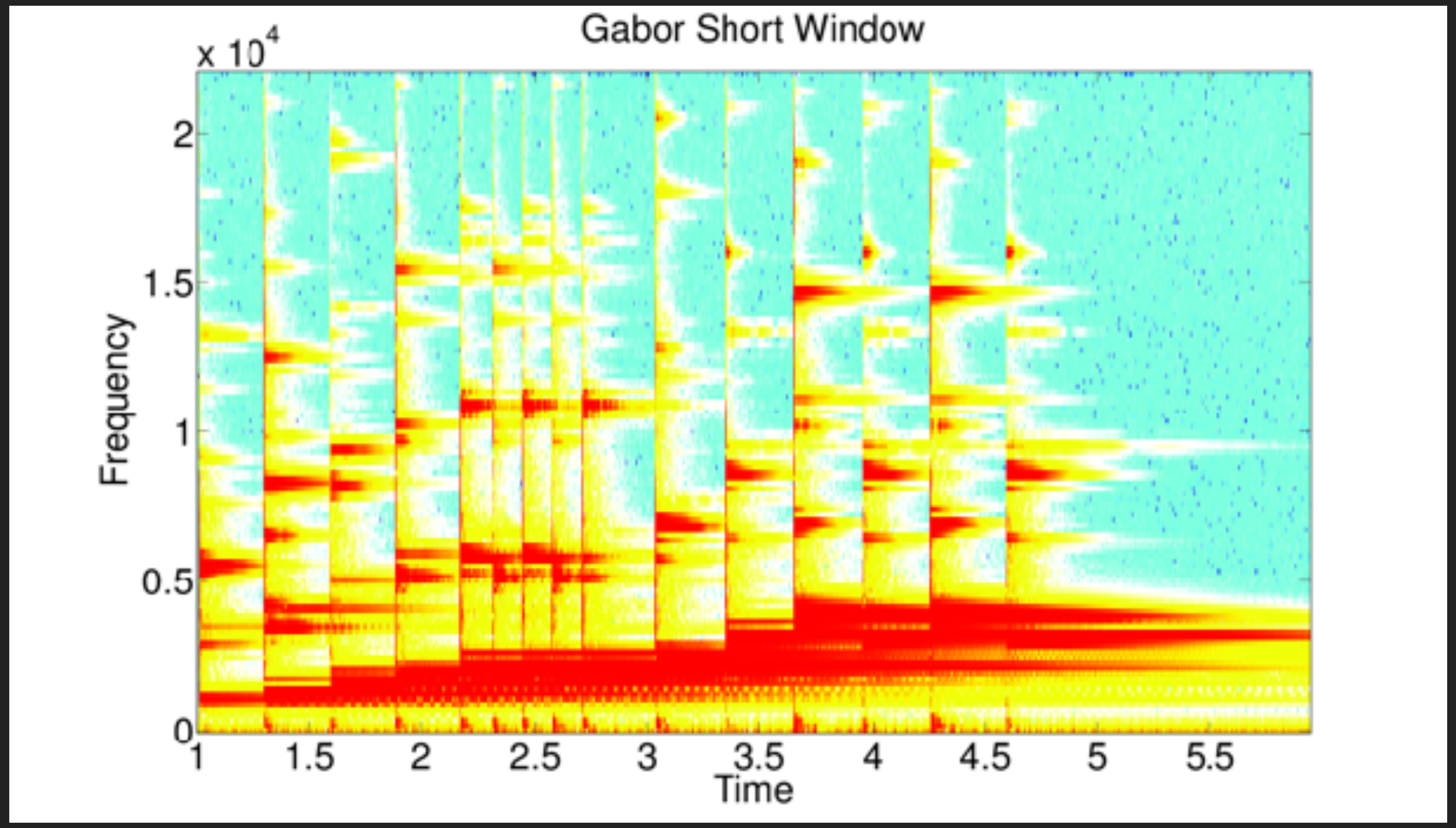



# EXAMPLE: GLOCKENSPIEL



# CONTINUOUS WAVELET TRANSFORM

- Idea: be sensitive to irregularities instead of oscillations

Let $\psi(t)$ be an admissible "mother" wavelet, and its the dilated and translated version

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

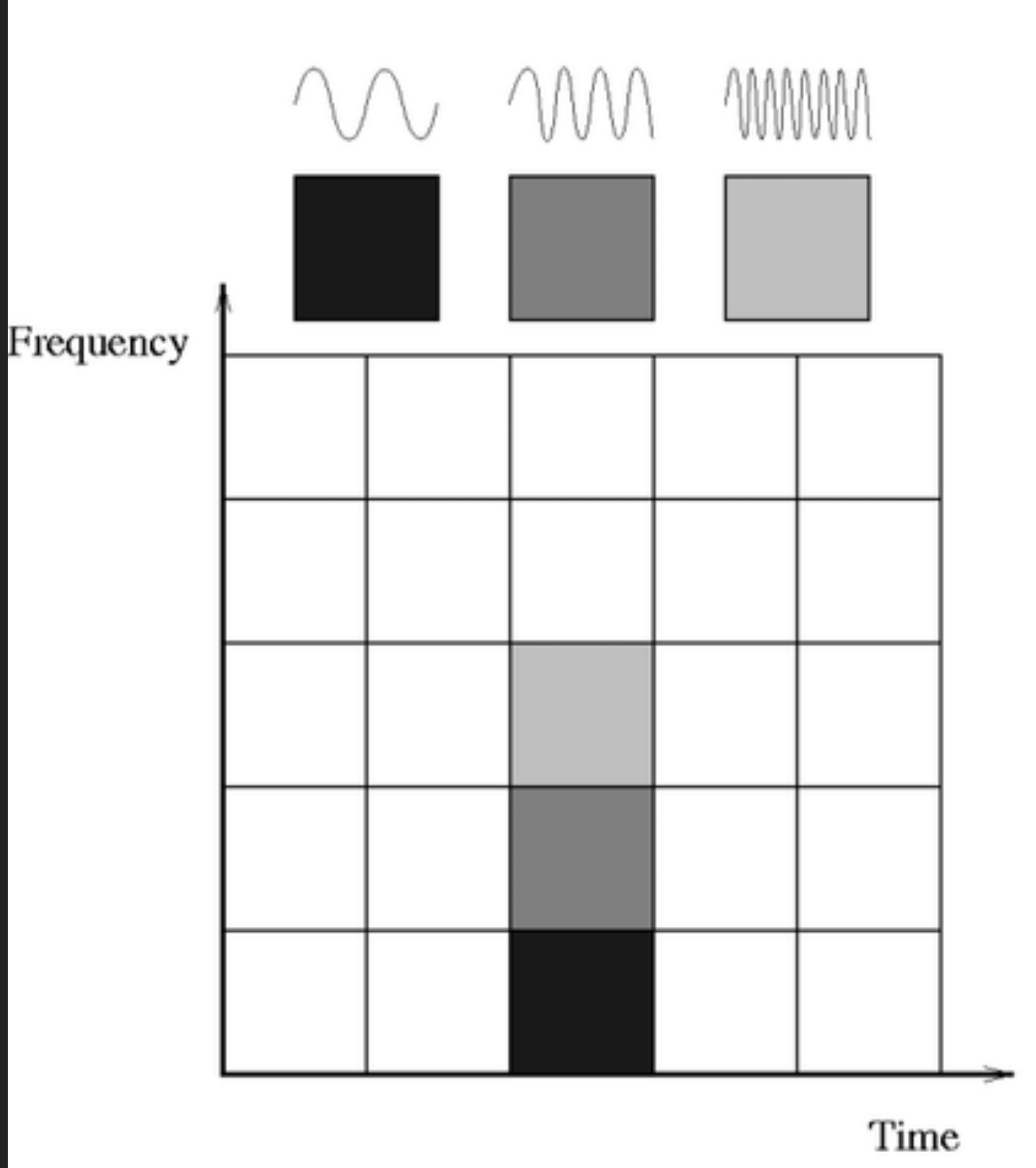
The **continuous** wavelet **transform** is given by:

$$C_x(a,b) = \langle x(t), \psi_{a,b}(t)\rangle = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right)\mathrm{d}t$$

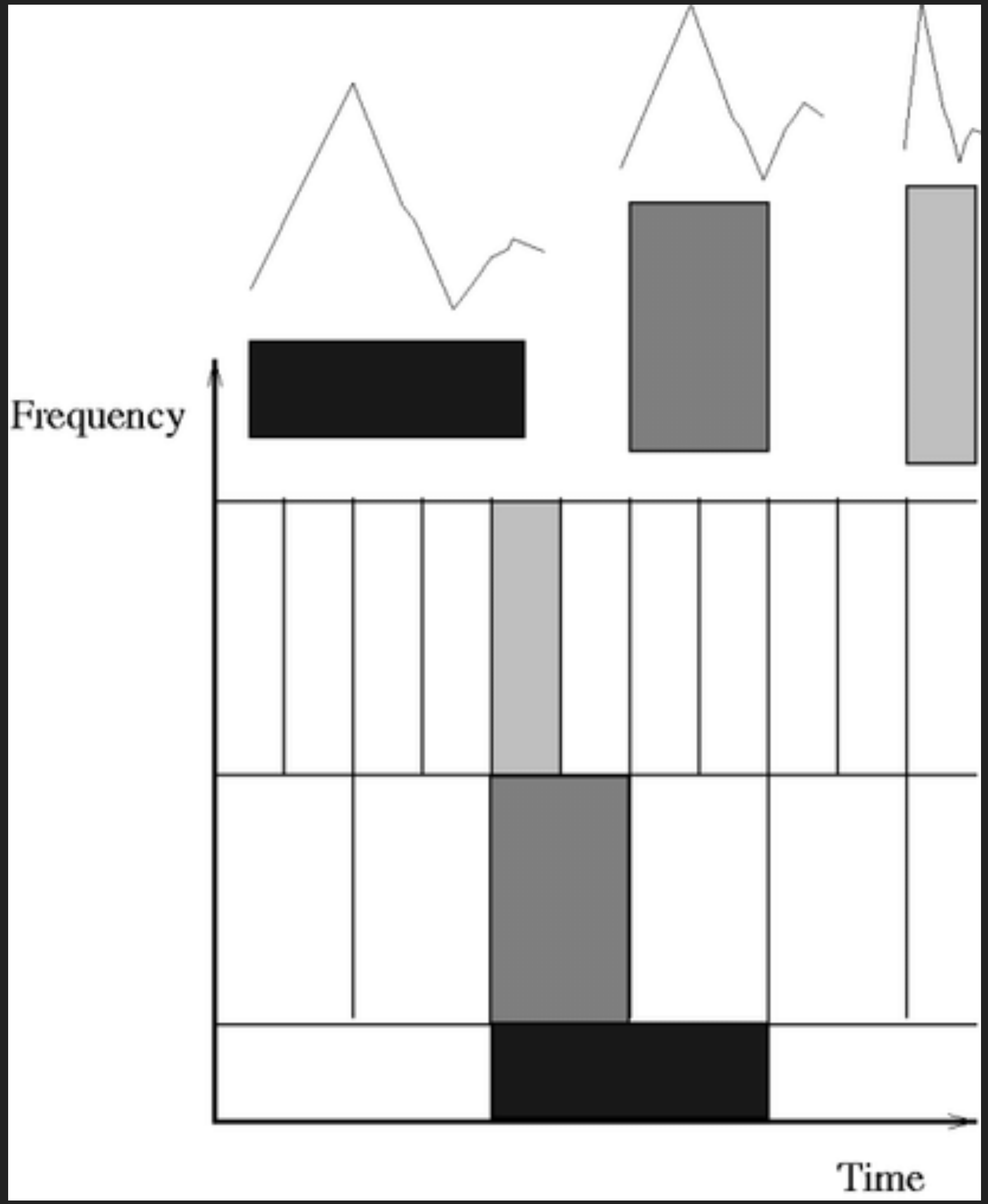
- It is a **times-scale** transform
- It is invertible: $x(t) = \frac{1}{c_\psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} X(a,b)\psi\left(\frac{t-b}{a}\right)\frac{\mathrm{d}a\ \mathrm{d}b}{a^2}$
- We have energy preservation

# TIME-FREQUENCY VS TIME-SCALE

- STFT tilling of the time-frequency plane




- Wavelet tilling of the time-frequency plane



# FROM CONTINUOUS TRANSFORM TO DISCRETE TRANSFORM

- First idea: construction of orthogonal basis
- No STFT orthogonal basis (Balian-Low theorem)
- Time-frequency orthogonal basis: MDCT
- Time-scale orthogonal basis: multi-resolution analysis and dyadic wavelets
- Alternative to orthogonal basis ?

# FRAME

# EXAMPLE

- Let the Kronecker basis $\Delta = \left\{\delta_0(t), \ldots, \delta_{N-1}(t)\right\}$ , $\delta_k(t) = \begin{cases} 1 & t = k \\ 0 & t \neq k \end{cases}$
- Let the Fourier basis $\mathscr{E} = \left\{\epsilon_0(t), \ldots, \epsilon_{N-1}(t)\right\}$, $\epsilon_k(t) = \frac{1}{\sqrt{N}} e^{i2\pi \frac{k}{N} t}$
- Let $x(t) = \delta_k(t) + \epsilon_n(t)$
- $x$ needs N coefficients in $\Delta$, and N coefficients in $\mathscr{E}$ for a perfect representation
- x needs only 2 coefficients in $\mathscr{D} = \Delta \cup \mathscr{E}$ for a perfect representation
- $\mathscr{D}$ is not an orthogonal basis (and is often called a "*dictionary*")

# FRAME: DEFINITION

- A frame is a system of *discrete* representation where inversion is *stable*
- Let a dictionary $\mathscr{D} = \left\{\varphi_n(t),\ \varphi_n(t) \in L^2(\mathbb{R})\right\}$ ($\varphi_n(t)$ being called a atom of $\mathscr{D}$). $\mathscr{D}$ is a frame of $L^2(\mathbb{R})$ iff it exists two constant $A, B > 0$ such that for all $f \in L^2(\mathbb{R})$

$$A\|f\|^2 \leq \sum_{n=-\infty}^{+\infty} \left|\langle f(t), \varphi_n(t)\rangle\right|^2 \leq B\|f\|^2$$

- If $A = B$, the frame is a *tight-frame*
- If $A = B = 1$ the frame is a *Parseval frame*
- If $A = B = 1$ **and** $\|\varphi_n\| = 1\ \forall n$ , the frame is an orthogonal basis

# FRAME: ANALYSIS AND SYNTHESIS OPERATORS

- Analysis operator $\Phi^*$

$$\begin{aligned} \Phi^* \ &: \ L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}) \\ & f(t) \mapsto \Phi^* f = \left\{ \langle f(t), \varphi_n(t) \rangle \right\}_{n\in\mathbb{Z}} \end{aligned}$$

The coefficients $\left\{ \langle f(t), \varphi_n(t) \rangle \right\}_{n\in\mathbb{Z}}$ are called the analysis coefficients of $f(t)$

- Synthesis operator $\Phi$

$$\begin{aligned} \Phi \ &: \ \ell^2(\mathbb{Z}) \to L^2(\mathbb{R}) \\ & \alpha \mapsto \Phi\alpha = \sum_{n=-\infty}^{+\infty} \alpha_n \varphi_n(t) \end{aligned}$$

The coefficients $\alpha_n$ are called synthesis coefficients of $f(t)$

# FRAME: THE FRAME OPERATOR

- The analysis and synthesis operators are **adjoint**
- The frame operator

$$\mathscr{R} = \Phi\Phi^* : \sum_{n=-\infty}^{+\infty} \langle f(t), \varphi_n(t) \rangle \varphi_n(t)$$

- We have $f(t) = \mathscr{R}(f) = \sum_{n=-\infty}^{+\infty} \langle f(t), \varphi_n(t) \rangle \varphi_n(t)$ iff the frame is a **Parseval frame**

# FRAME: THE DUAL FRAME

- Let $\mathscr{D} = \left\{\varphi_n(t),\ \varphi_n(t) \in L^2(\mathbb{R})\right\}$ be a frame with constants $A, B$. The dual frame of $\mathscr{D}$ is given by

$$\tilde{\mathscr{D}} = \left\{\tilde{\varphi}_n(t) = \mathscr{R}^{-1}(\varphi_n)\right\}$$

- $\tilde{\mathscr{D}}$ is also a frame such that

$$\frac{1}{B}\|f\|^2 \leq \sum_{n=-\infty}^{+\infty} \left|\langle f(t), \tilde{\varphi}(t)\rangle\right|^2 \leq \frac{1}{A}\|f\|^2$$

- We have $f(t) = \sum_{n=-\infty}^{+\infty} \langle f, \tilde{\varphi}_n(t)\rangle \varphi_n(t) = \sum_{n=-\infty}^{+\infty} \langle f, \varphi_n(t)\rangle \tilde{\varphi}_n(t)$

- If $\mathscr{D}$ is a tight-frame, then $\tilde{\varphi}_n(t) = \dfrac{1}{A}\varphi_n(t)$

# FRAME: EXAMPLES

- An orthogonal basis is a Parseval-Frame
- An union of 2 orthogonal bases is a tight-frame with constant $A = 2$
- An union of K orthogonal bases is a tight-frame with constant $A = K$
- In finite dimension $\mathbb{C}^M$, every matrix $U \in \mathbb{C}^{MN}$ such that $\text{rank}(U) = M$ is a frame
- Moreover, if $UU^* = A\,\text{Id}_M$, then $U$ is a tight-frame with constant $A$

# GABOR FRAME IN FINITE DIMENSION

- Let $x[t] \in \mathbb{R}^T$, then one can define the full STFT with the real, normalized, analysis window $w[t]$

$$X[\tau,\nu] = \sum_{t=0}^{T-1} x[t]w[t-\tau]e^{-i2\pi\frac{\nu}{T}t} \text{, for all } \tau,\nu = 0..(T-1)$$

- Moreover, we have

$$x[t] = \sum_{\tau=0}^{T-1}\sum_{\nu=0}^{T-1} X[\tau,\nu]w[t-\tau]e^{i2\pi\frac{\nu}{T}t}$$

- Very redundant: $T^2$ time-frequency coefficients

# DISCRETE GABOR FRAME IN PRACTICE

- Let $x[t] \in \mathbb{R}^T$ and Let $w[t] \in \mathbb{R}^L$ be a real, normalized, analysis window (with $L \leq T$)
- Let $t_0, \nu_0 \in \mathbb{N}_*^+$ and let the Gabor atoms $\varphi_{\tau,\nu}[t] \in \mathbb{R}^T$

$$\varphi_{\tau,\nu}[t] = w\left[t - \frac{L}{t_0}\tau\right] e^{i2\pi \frac{\nu}{\nu_0 L} t}$$

- Usual choices are $\nu_0 = 1$ (FFT of size $L$) or $\nu_0 = 2$ (FFT of size $2L$), and $t_0 = 2$ (overlap of $50\,\%$ between 2 windows) or $t_0 = 4$ (overlap of $75\,\%$ between 2 windows)
- The dual frame is still a Gabor frame, with the dual window $\tilde{w}(t)$

# SPARSE SYNTHESIS

# FROM ANALYSIS TO SYNTHESIS

- Let $x[t] \in \mathbb{R}^T$ and $\mathscr{D} = \{\varphi_n[t]\}_{n=0}^N$ be an over complete dictionary ($N \geq T$)
- Let $\Phi \in \mathbb{C}^{TN}$ the matrix associated to the dictionary (the synthesis operator). The $k$-th column of $\Phi$ is then the atom $\varphi_k$
- The analysis of $x$ is given by $\Phi^* x = \left\{\langle x[t], \varphi_n[t]\rangle\right\}_{n=0}^N$

*How is a signal best synthesized?*

# SPARSE SYNTHESIS

- Back to the example: $x(t) = \delta_k(t) + \epsilon_n(t)$
- Goal: how can we synthesize $x$ with the fewest possible coefficients ?
- Ie: find the synthesis coefficients $\alpha$ such that $x[t] = \sum_{n=0}^{N-1} \alpha_n \varphi_n[t]$ such that most of $\alpha_n = 0$
- Definition of the quasi-norm $\ell_0$ : $\|\alpha\|_0 = \#\{\alpha_n \neq 0\}$

# SPARSE SYNTHESIS

- Goal: how can we synthesize $x$ with the fewest possible coefficients ?

$$\min \|\alpha\|_0 \quad \text{s.t. } x = \Phi\alpha$$

- NP Hard ! Can be solved by MILP programming when $N$ is small
- Idea: replace the $\ell_0$ norm by something easier to minimize

# SPARSE SYNTHESIS: THE FRAME METHOD

- Replace the $\ell_0$ norm by the $\ell_2$ norm

$$\min \|\alpha\|_2 \quad \text{s.t. } x = \Phi\alpha$$

- Solution: $\alpha = \Phi^*(\Phi\Phi^*)^{-1}x$
- It is the dual frame
- No sparsity

# SPARSE SYNTHESIS: THE BASIS PURSUIT

- Replace the $\ell_0$ norm by the $\ell_1$ norm: $\|\alpha\|_1 = \sum_{n=0}^{N-1} |\alpha_n|$

$$\min \ \|\alpha\|_1 \quad \text{s.t. } x = \Phi\alpha$$

- Solution obtained by linear programming (the problem is convex and linear)
- Sparsity of the solution
- When this solution is the same as the true $\ell_0$ problem ? (Wait  few classes)

# SPARSE SYNTHESIS: THE MATCHING PURSUIT

- Idea: use a greedy approach
  - Init: $r^{(0)} = x,\ x^{(0)} = 0,\ k = 0$
  - Repeat
    1. Find the optimal atom: $\lambda_k = \underset{\lambda}{\operatorname{argmax}} |\langle r, \varphi_\lambda \rangle|$
    2. Update the approximation: $x^{(k+1)} = x^{(k)} + \langle r^{(k)}, \varphi_{\lambda_k} \rangle \varphi_{\lambda_k}$
    3. And the residual: $r^{(k+1)} = x - x^{(k+1)} = r^{(k)} - \langle r^{(k)}, \varphi_{\lambda_k} \rangle \varphi_{\lambda_k}$

# SPARSE SYNTHESIS: THE MATCHING PURSUIT

After $K$ iterations ($K \geq 0$), we have

- $x = \sum_{k=0}^{K} \langle r^{(k)}, \varphi_{\lambda_k} \rangle \varphi_{\lambda_k} + r^{(K)}$
- $\|r^{(K)}\|^2 = \|r^{(K-1)}\|^2 - |\langle r^{(K-1)}, \varphi_{\lambda_{K-1}} \rangle|^2$

Moreover, we have

$$\lim_{k \to +\infty} \|r^{(k)}\| = 0$$

# SPARSE SYNTHESIS: THE MATCHING PURSUIT

- Shortcoming of the MP:
  - It converges asymptotically
  - An atome $\varphi_\lambda$ can be chosen several times
- Solution: orthogonal matching pursuit

# SPARSE SYNTHESIS: THE OMP

- Idea: orthogonal projection of the signal on the subspace spanned by the selected atoms
- Algorithm:
  - Init: $r^{(0)} = x,\ x^{(0)} = 0,\ k = 0$
  - Repeat
    1. Find the optimal atom: $\lambda_k = \underset{\lambda}{\operatorname{argmax}} \,|\,\langle r, \varphi_\lambda \rangle\,|$
    2. Update the approximation: $x^{(k+1)} = P_{V^k} x = \sum_{j=0}^{k-1} \alpha_j \varphi_{\lambda_j}$ with $\alpha_k = \left(\Phi^*_{\Lambda_k} \Phi_{\Lambda_k}\right)^{-1} \Phi^*_{\Lambda_k} x$ and $\Lambda_k = \{\lambda_j\}_{j=0}^{k}$
    3. And the residual: $r^{(k+1)} = x - x^{(k+1)}$

# SPARSE SYNTHESIS: THE OMP

- Converge in $N$ iterations (remember $x \in \mathbb{R}^N$)
- Orthogonal projection can be costly in computation time

# SPARSE DENOISING

# DENOISING BY SPARSE SYNTHESIS

- Let $x \in \mathbb{R}^T$ be a signal and $\Phi \in \mathbb{R}^{TN}$ a dictionary where $x$ is sparse
- Let $y \in \mathbb{R}^T$ be a noisy observation of $x$:

$$y = x + n$$

With $n \in \mathbb{R}^T$ a gaussian white noise

- Let $\alpha \in \mathbb{R}^N$ some sparse synthesis coefficients of $x$:

$$y = \Phi\alpha + n$$

- How to "denoise" $y$ ? Or, how to estimate the sparse coefficients $\alpha$ ?

# DENOISING BY SPARSE SYNTHESIS

$$y = \Phi\alpha + n$$

- Proposed solutions: solve

$$\min \ \|\alpha\|_0 \quad \text{s.t.} \ \|y - \Phi\alpha\|_2^2 \leq \sigma$$

- One can use MP, OMP, or the convex relaxation by replacing the $\ell_0$ norm by the $\ell_1$ norm
- LASSO or Basis Pursuit Denoising:

$$\min \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

# LASSO

$$\min \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

- It is a convex, non smooth problem
- Can be solved efficiently by proximal descent

# LASSO: THE ORTHOGONAL CASE

$$\min \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$$

- Suppose that $N = T$ and $\Phi^*\Phi = \Phi\Phi^* = \mathrm{Id}_N$.
- The problems reads, with $z = \Phi^* y$

$$\min \frac{1}{2}\|z - \alpha\|_2^2 + \lambda\|\alpha\|_1$$

- It is the so-called proximal operator of $\lambda\|\cdot\|_1$
- Solution: soft-thresholding

$$\alpha_k = \mathcal{S}_\lambda(z_k) = z_k\left(1 - \frac{\lambda}{|z_k|}\right)^+ \quad \text{with } (x)^+ = \max(x,0)$$

# LASSO: ISTA

$$\min \frac{1}{2}\|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1 = \mathcal{F}(\alpha)$$

- Let $L = \|\Phi\|^2$
- Iterative Shrinkage/Thresholding Algorithm (ISTA):

$$\alpha^{(t+1)} = \mathcal{S}_{\lambda/L}\left(\alpha^{(t)} + \frac{1}{L}\Phi^*(y - \Phi\alpha^{(t)})\right)$$

- $\mathcal{F}(\alpha^{(t)}) - \mathcal{F}(\alpha^*) \leq \frac{L}{2}\frac{\|\alpha^{(0)} - \alpha^*\|^2}{t}$
- Fast version: FISTA (just use a relaxation step)

# CONCLUSION

- Stable discretization of continuous transform: frame theory
- Sparse synthesis vs dual frame
- Sparse denoising
- Algorithms: Greedy (MP, OMP), Convex optimization (LASSO and proximal descent)
- Questions: matthieu.kowalski@universite-paris-saclay.fr