Transform Methods & Signal Processing

lecture 12

Matthew Roughan
<matthew.roughan@adelaide.edu.au>

Discipline of Applied Mathematics
School of Mathematical Sciences
University of Adelaide

July 27, 2009

Self-similarity in the frequency domain

Self-similarity

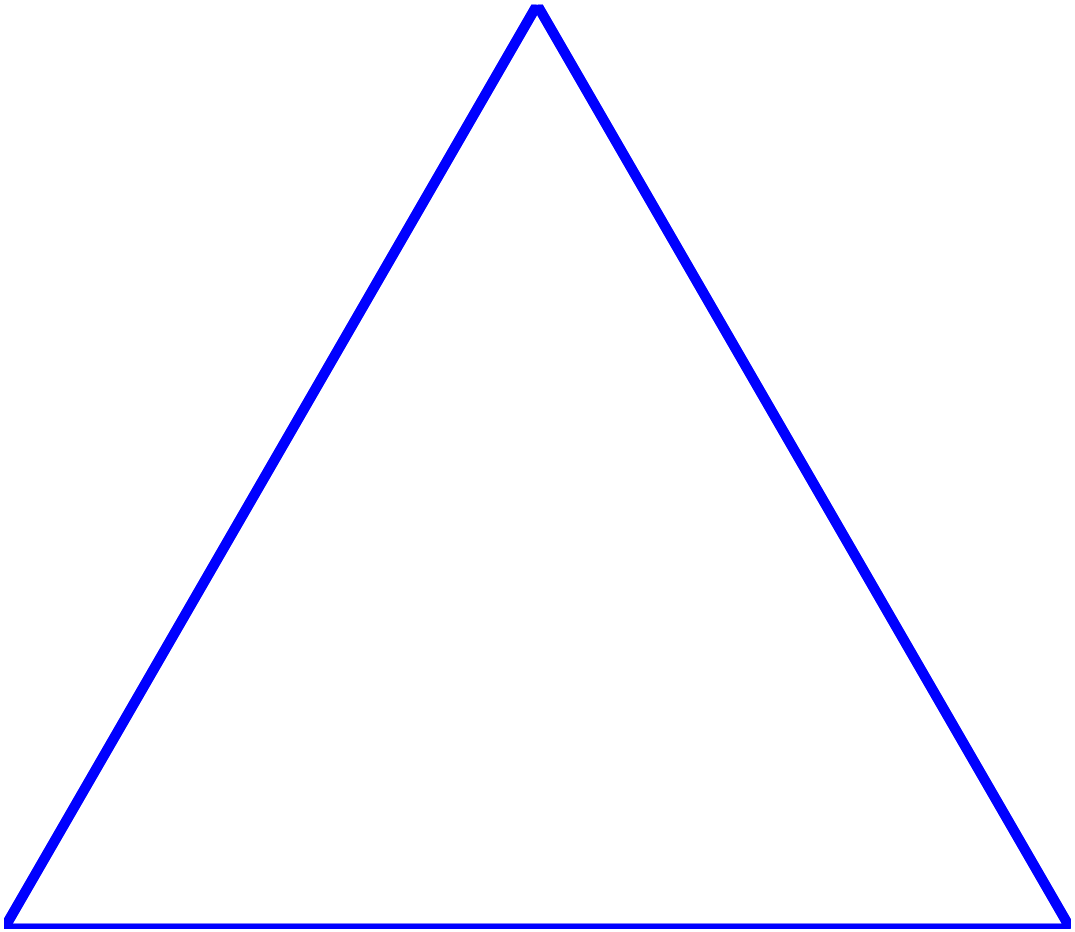
So, Nat'ralists observe, a flea
Hath smaller fleas that on him prey;
And these have smaller still to bite 'em
And so proceed ad infinitum

Jonathon Swift, 1733

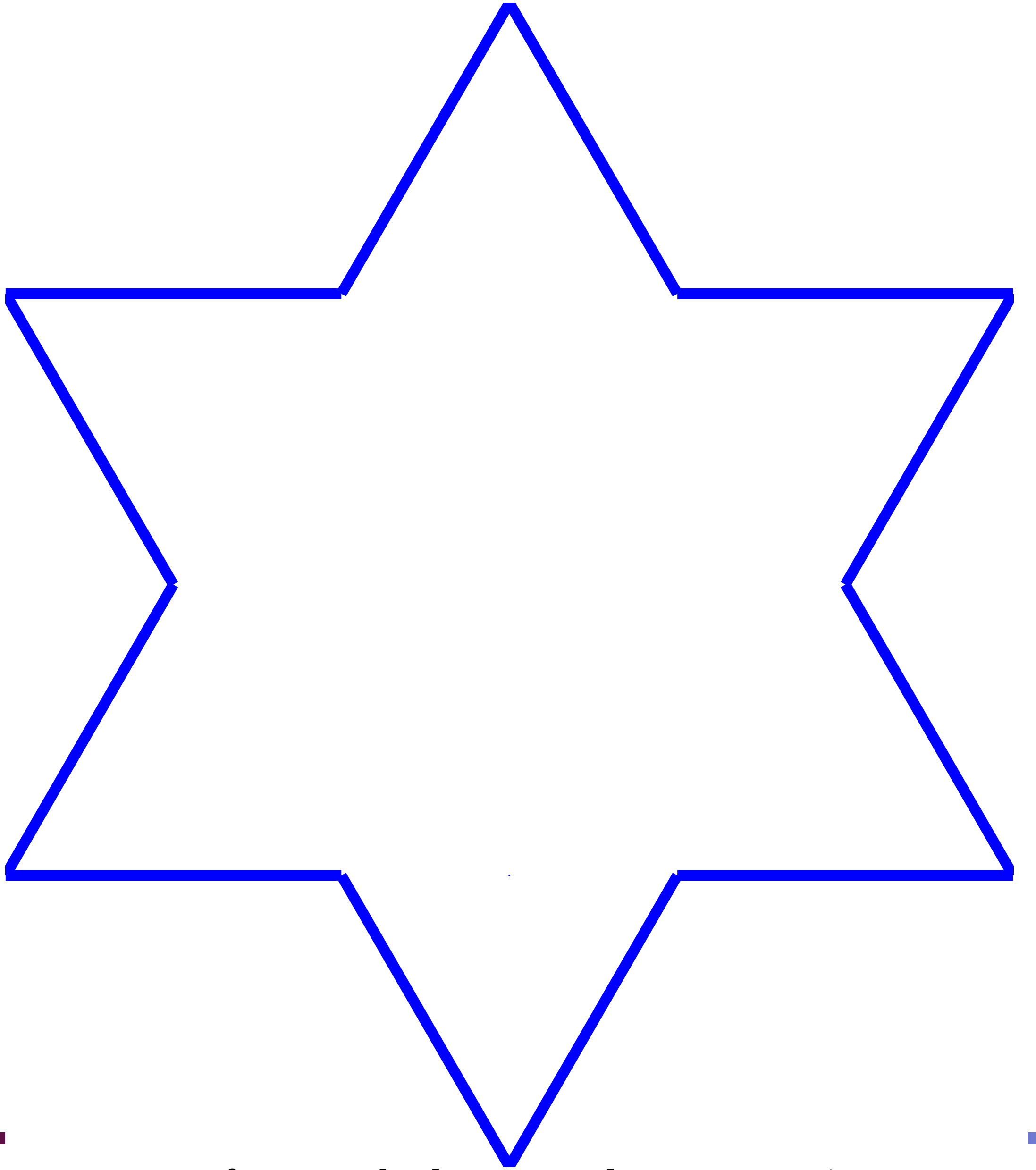
Great fleas have little fleas upon their backs to bite 'em,
And little fleas have lesser fleas, and so ad infinitum.
And the great fleas themselves, in turn, have greater fleas to go on;
While these again have greater still, and greater still, and so on.

De Morgan: A Budget of Paradoxes, p. 377.

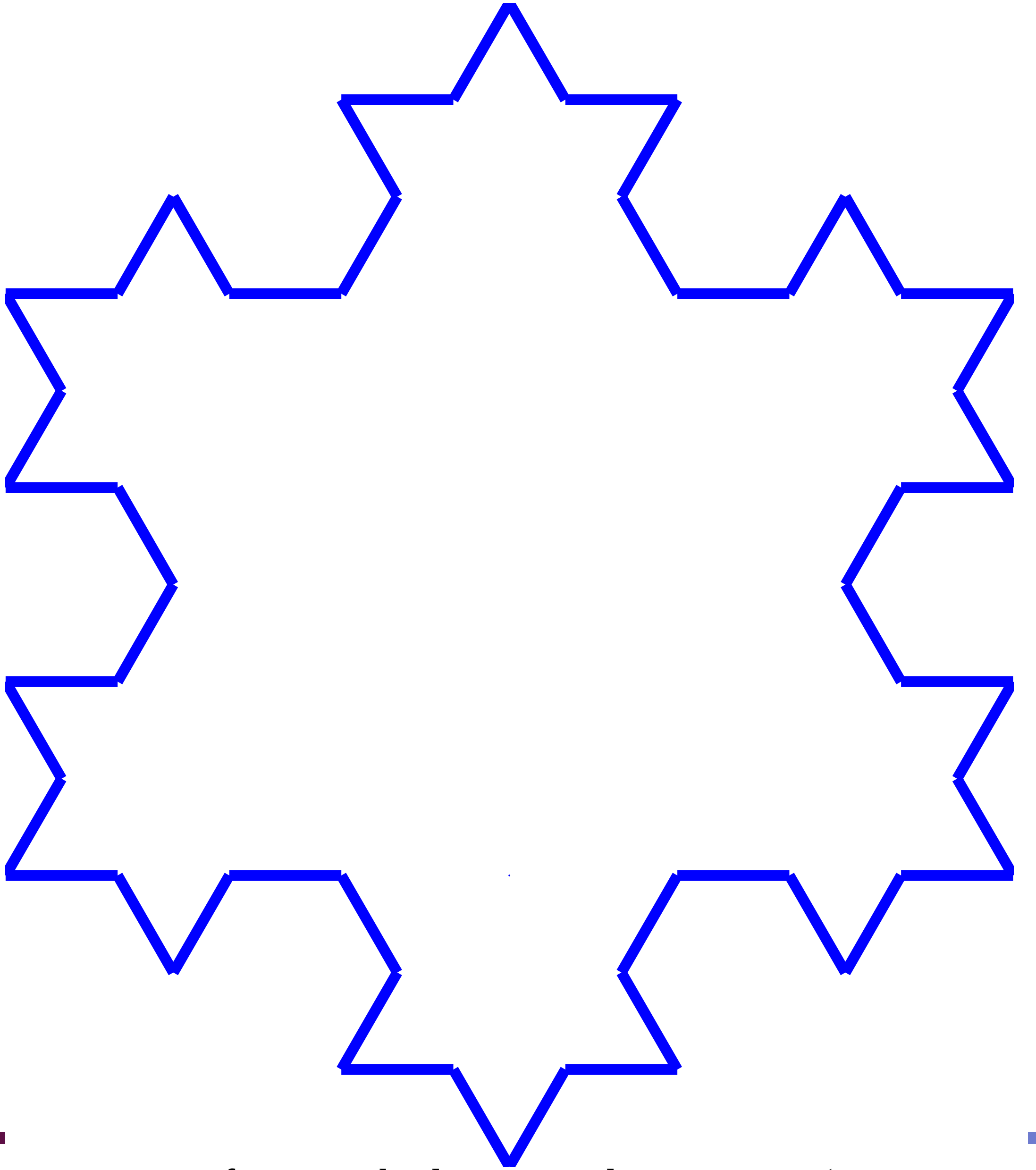
Self-similarity: Koch Snowflake



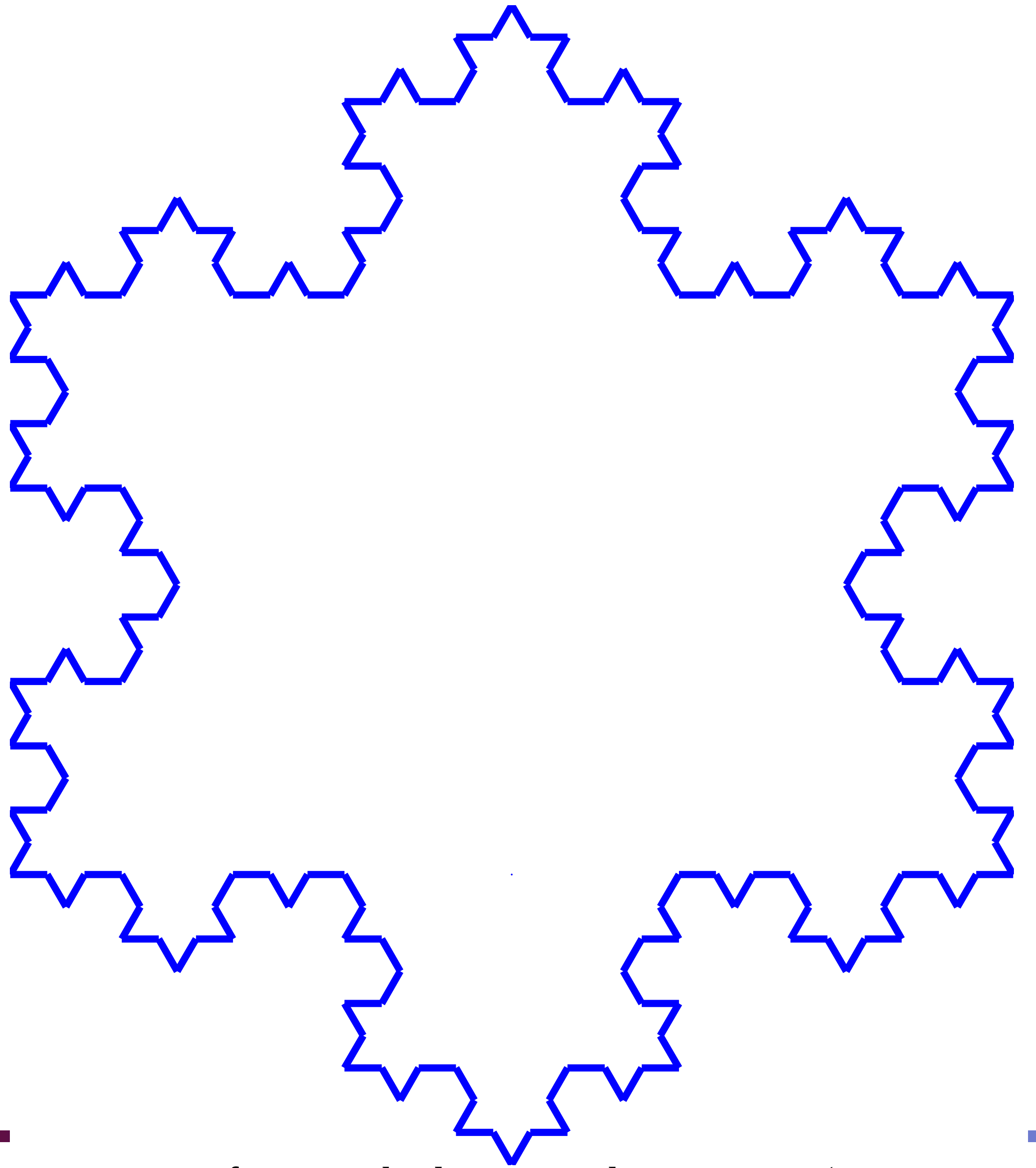
Self-similarity: Koch Snowflake



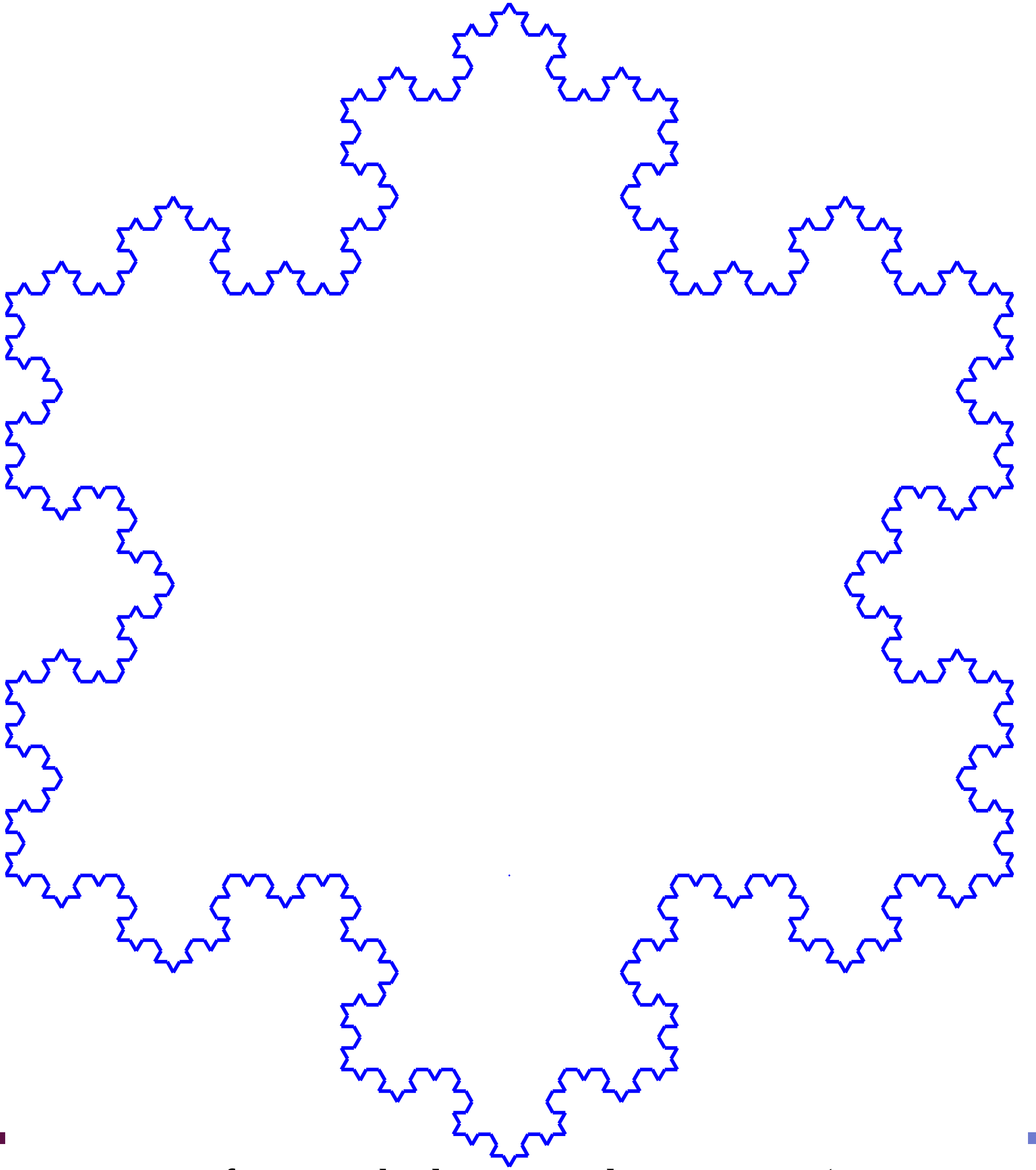
Self-similarity: Koch Snowflake



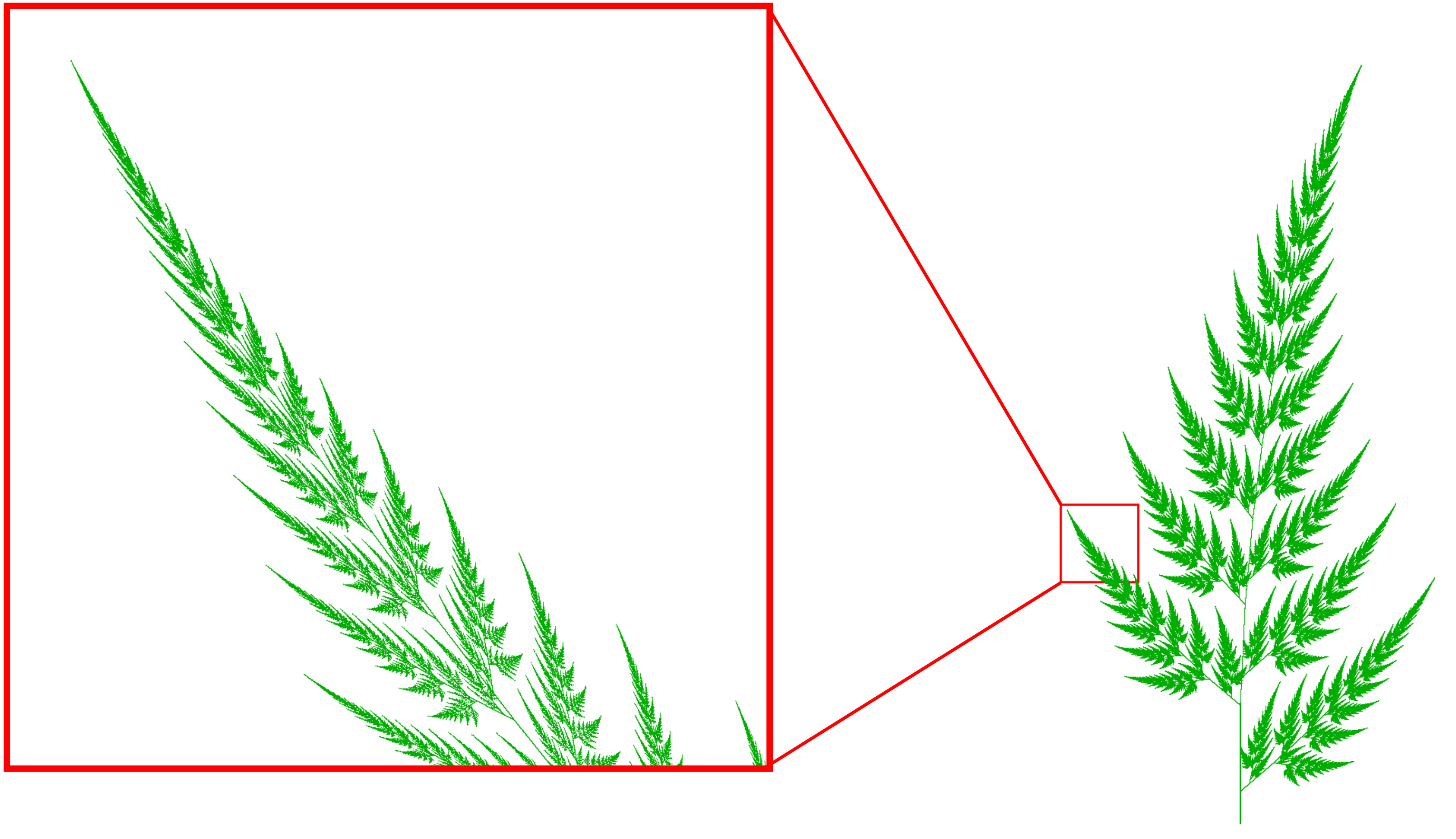
Self-similarity: Koch Snowflake



Self-similarity: Koch Snowflake

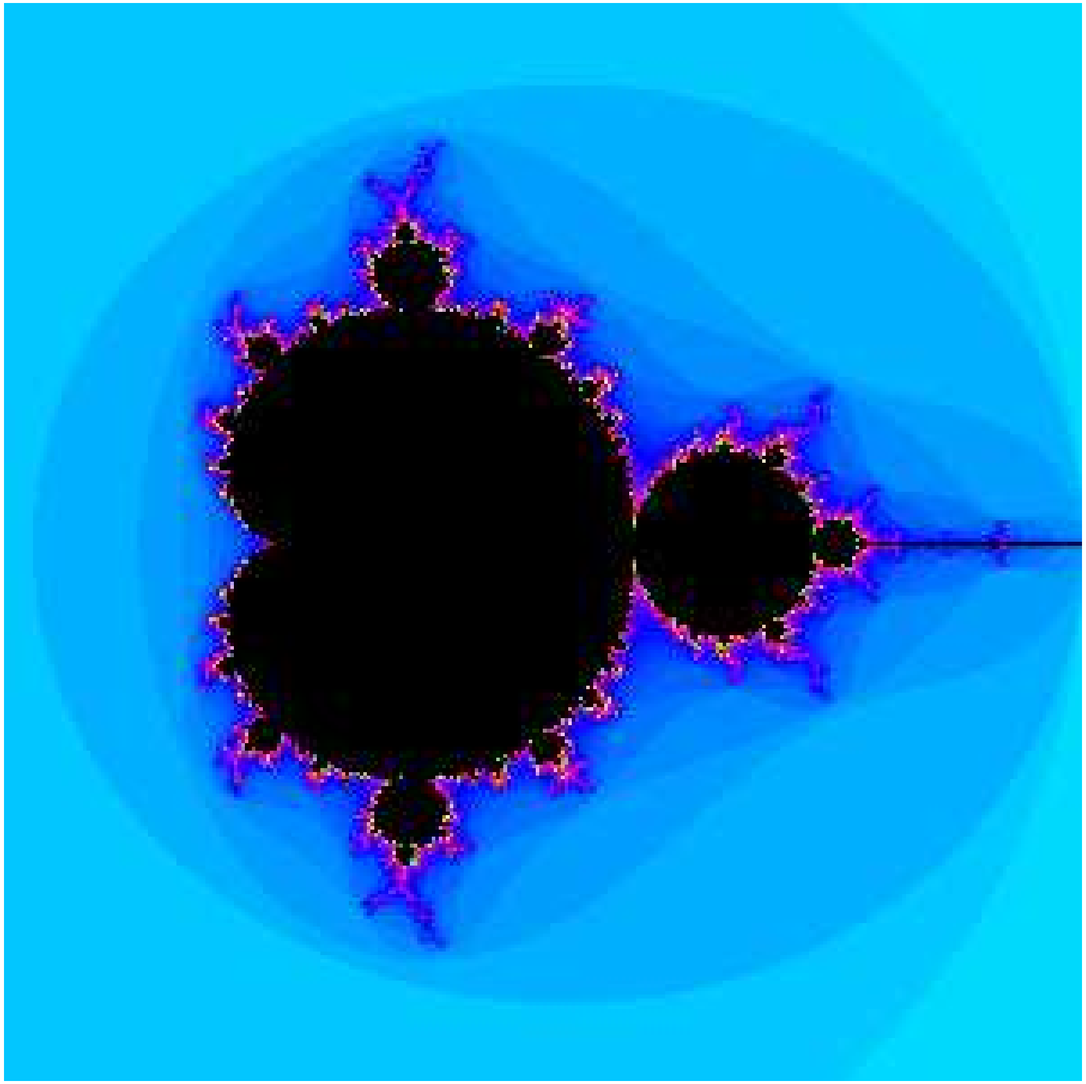


Self-similarity: IFS Fern



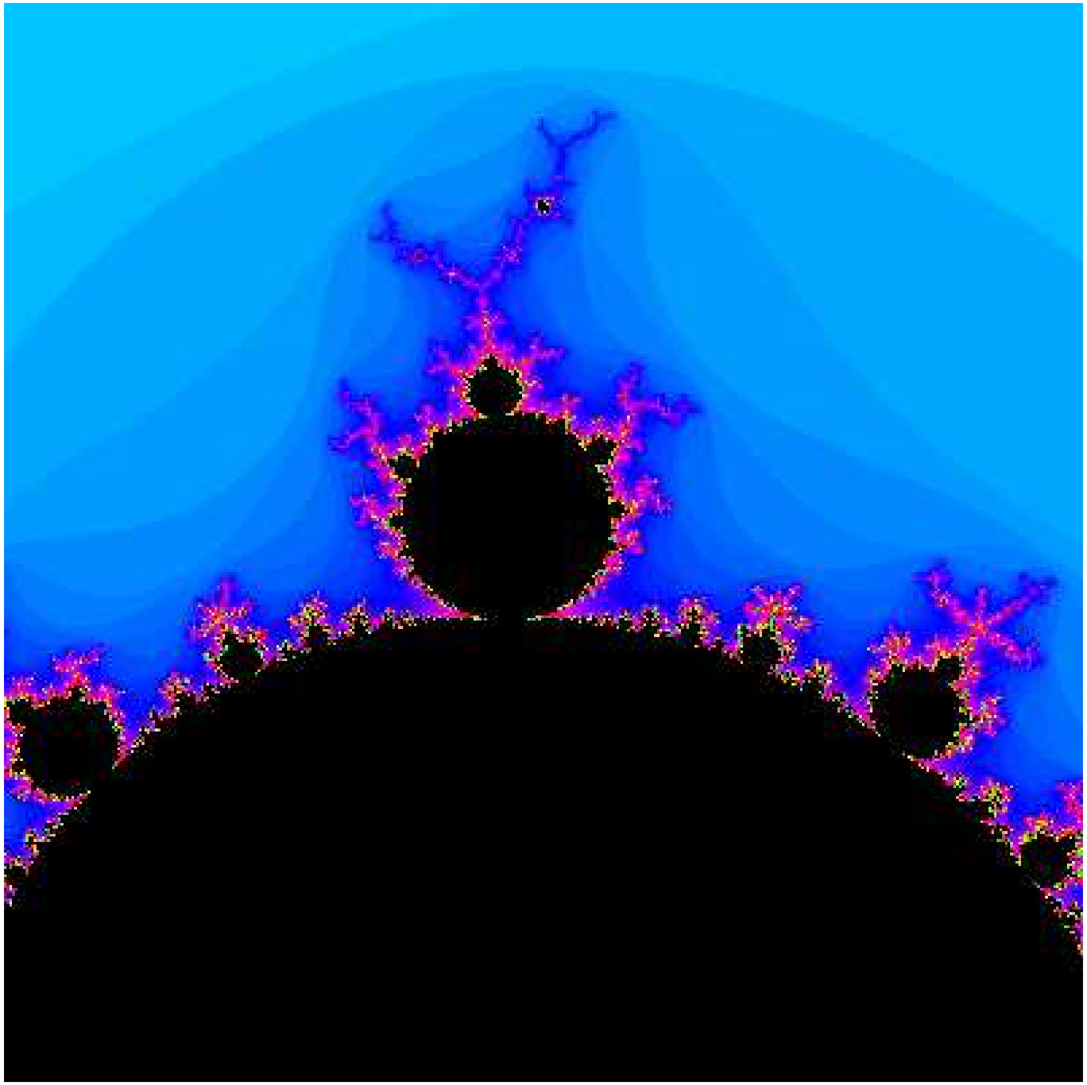
C code from
http://astronomy.swin.edu.au/~pbourke/fractals/

Mandelbrot set I



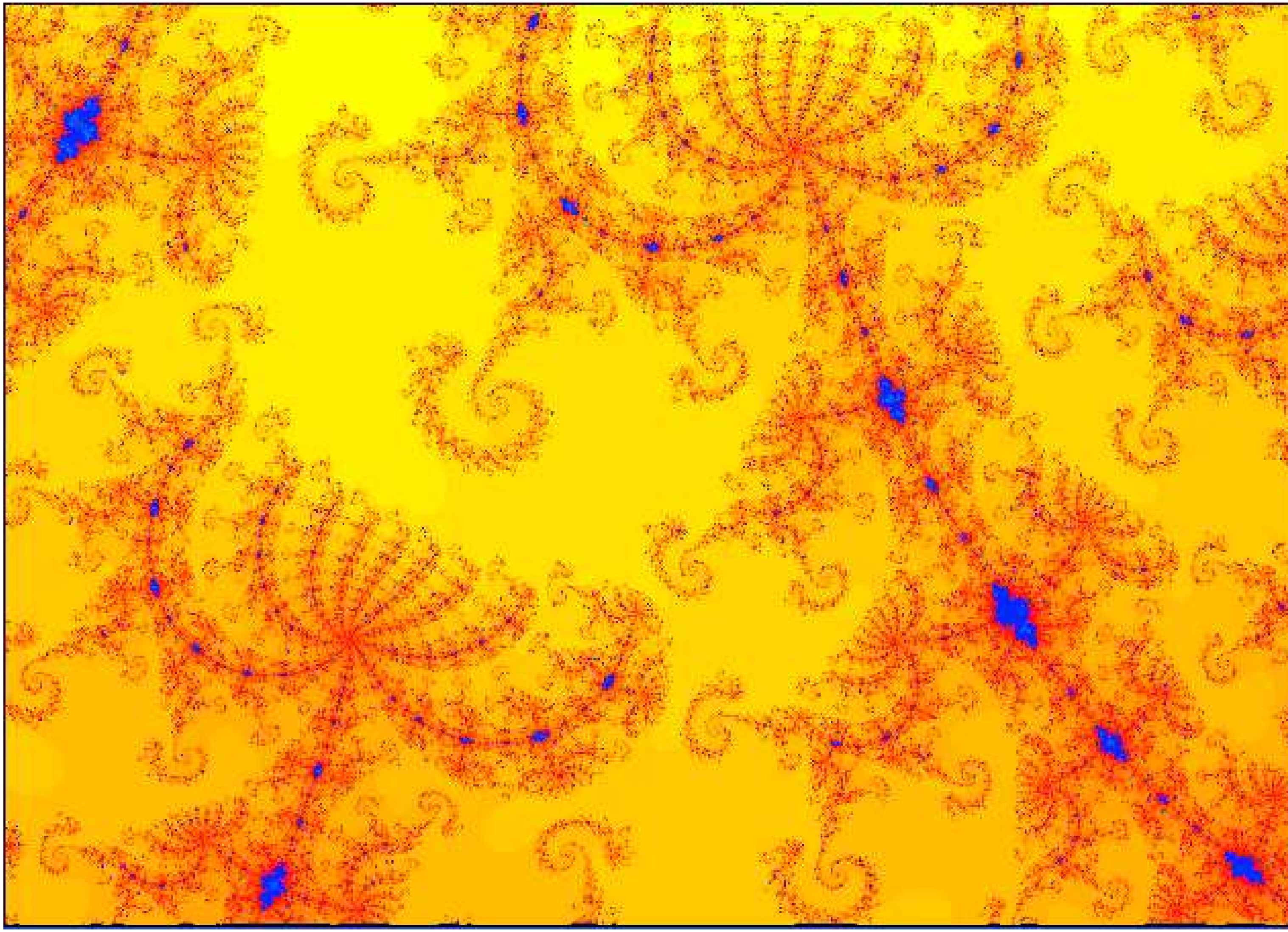
http://aleph0.clarku.edu/~djoyce/julia/julia.html

Mandelbrot set II



http://aleph0.clarku.edu/~djoyce/julia/julia.html

Mandelbrot set III



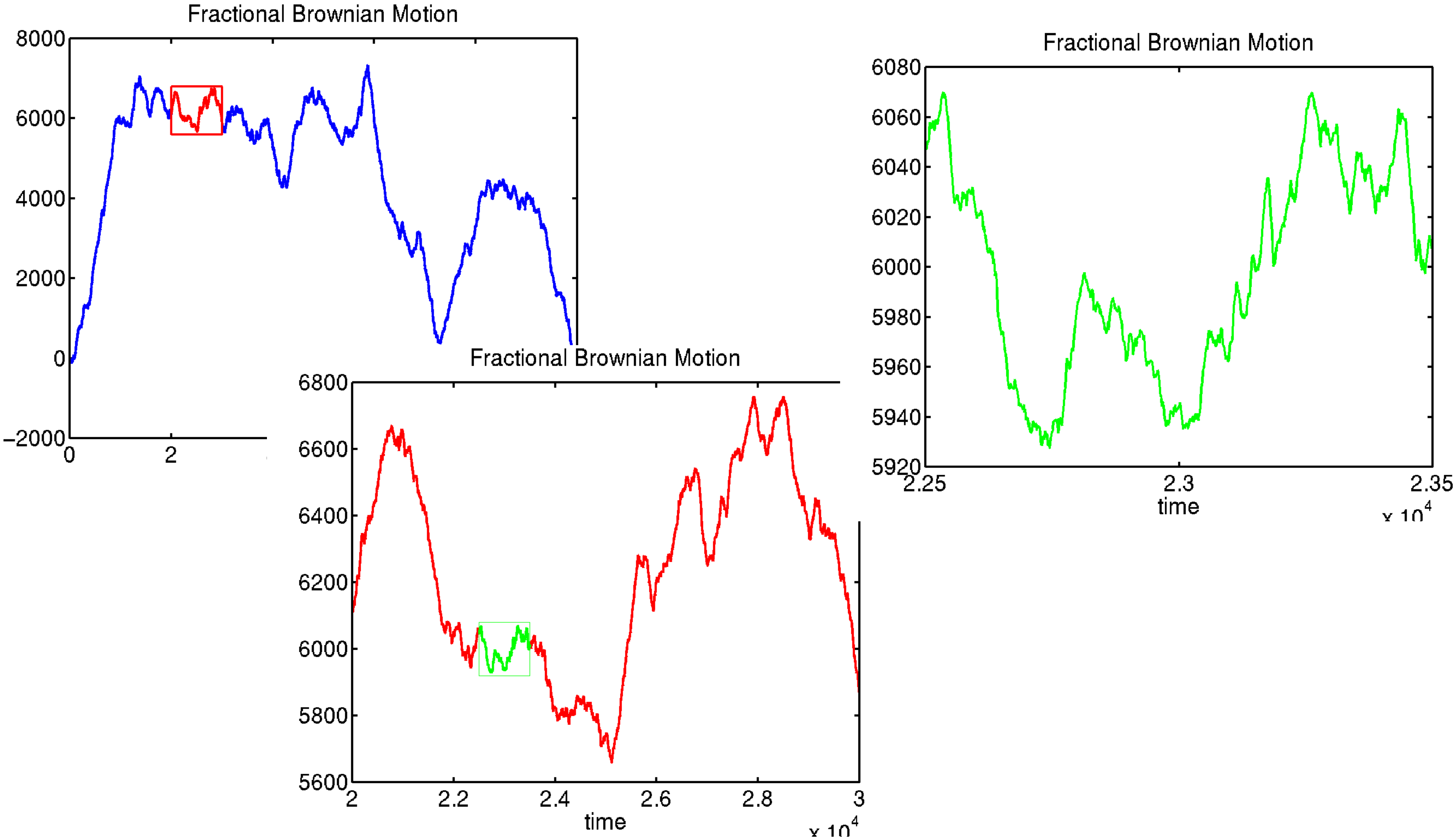
http://www.softsource.com/softsource/fractal.html

Statistical Self-similarity

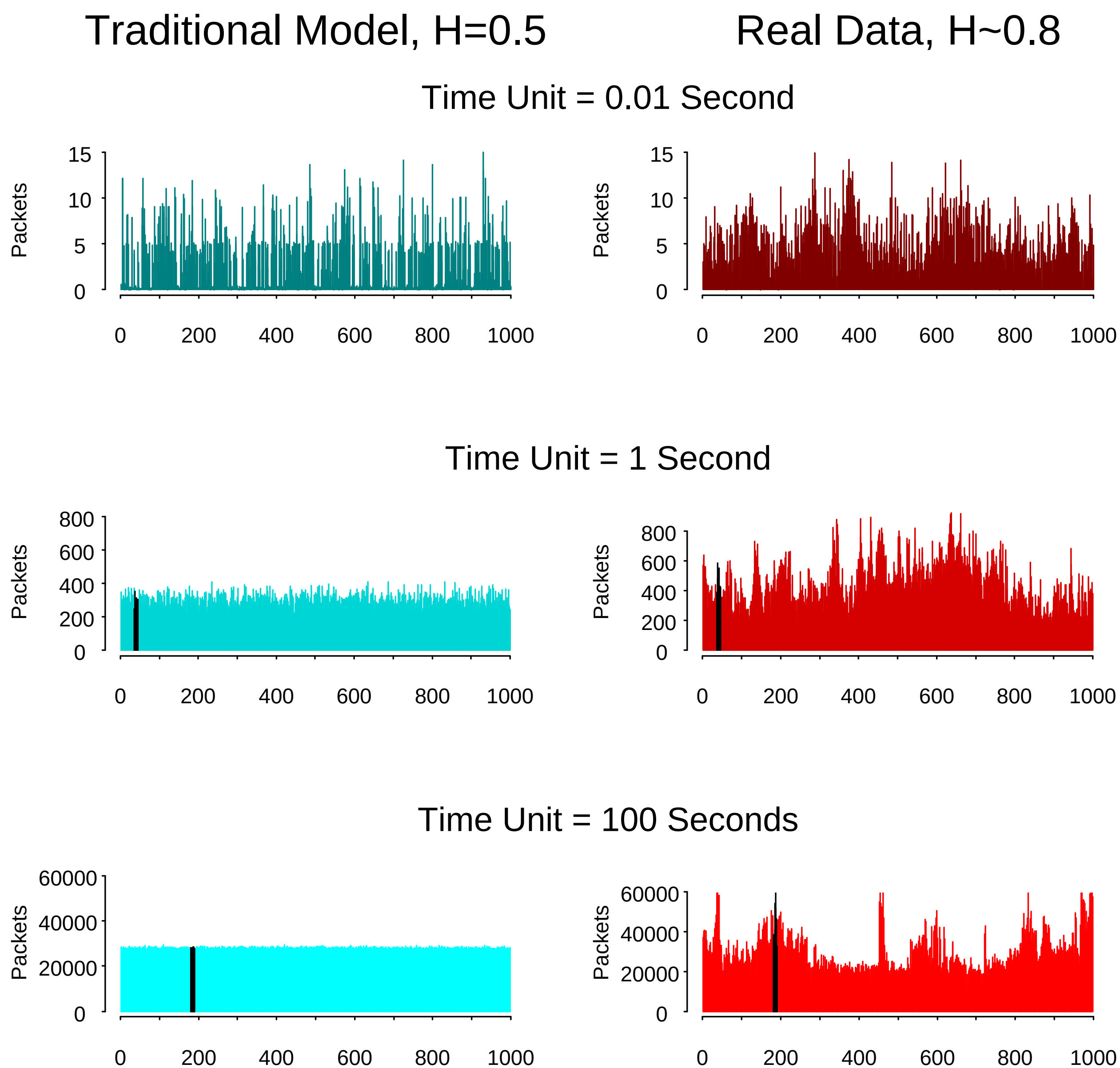
Statistical Self-similarity (SS)

- this is not a course on fractals
- Fractals (such as above) are deterministic
- we are interested in statistical properties of traffic
- look for **statistical** self-similarity

Statistical Self-similarity



Ethernet traffic



Statistical Self-similarity

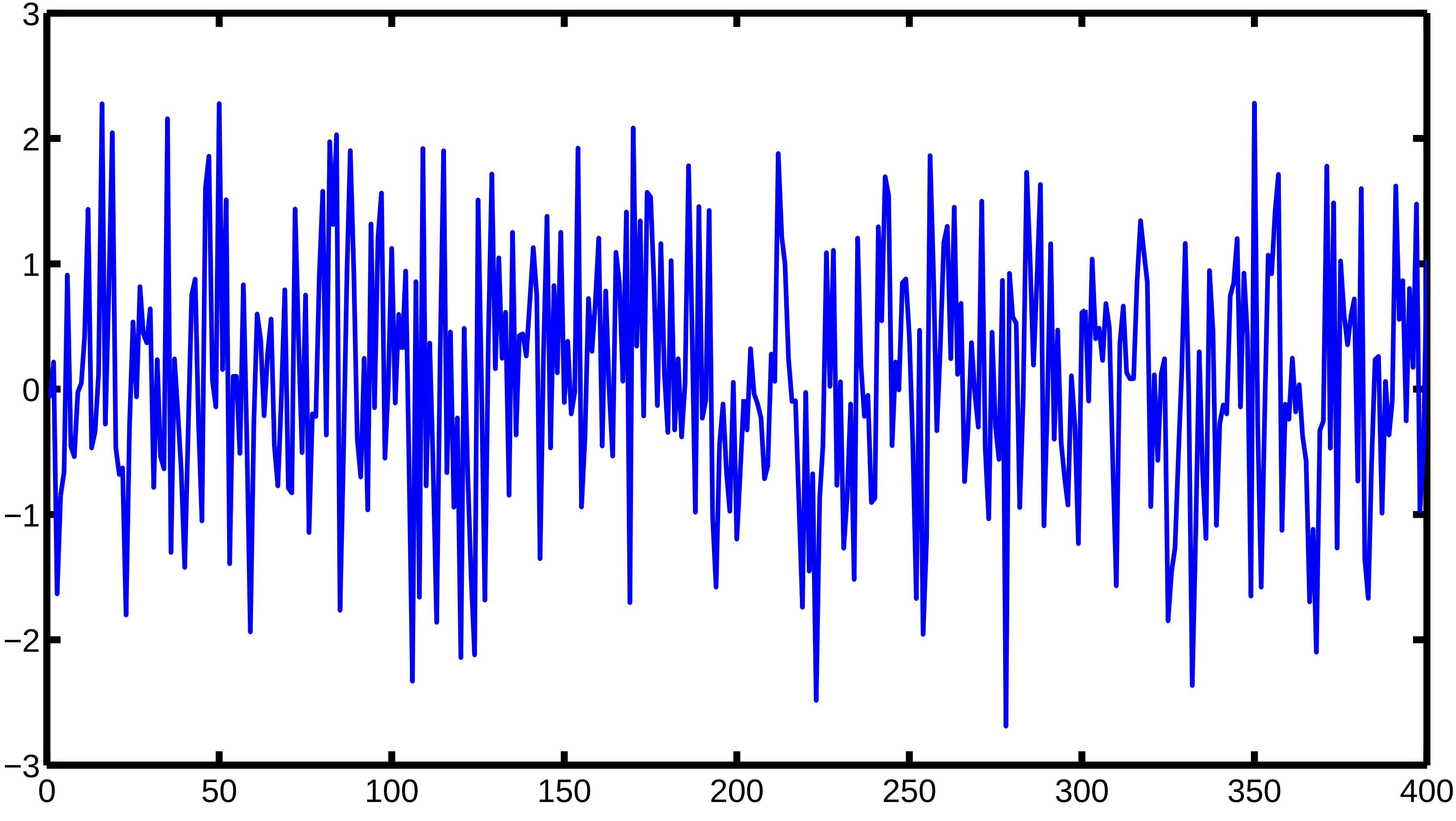
SS block aggregation definition
(another definition exists)

We define he aggregated time series $\{X_k^{(m)}\}$ at level m by

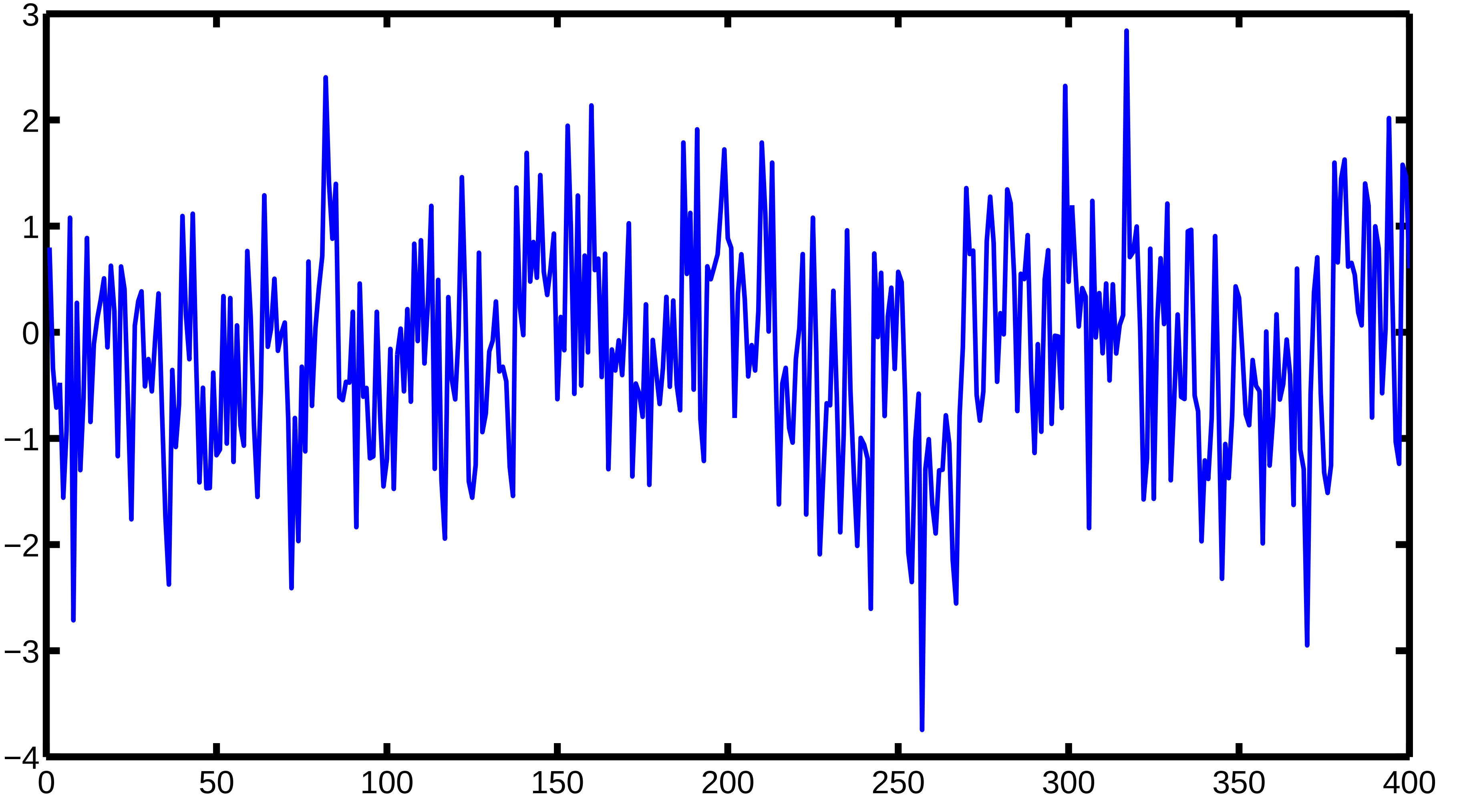
$$X_k^{(m)} := \frac{X_{(k-1)m+1} + \cdots + X_{km}}{m}.$$

A stationary time series $X = \{X_1, X_2, \ldots\}$ is called self-similar with Hurst parameter H if, for all m, the aggregated process $m^{1-H} X^{(m)}$ has the same distributions as X.

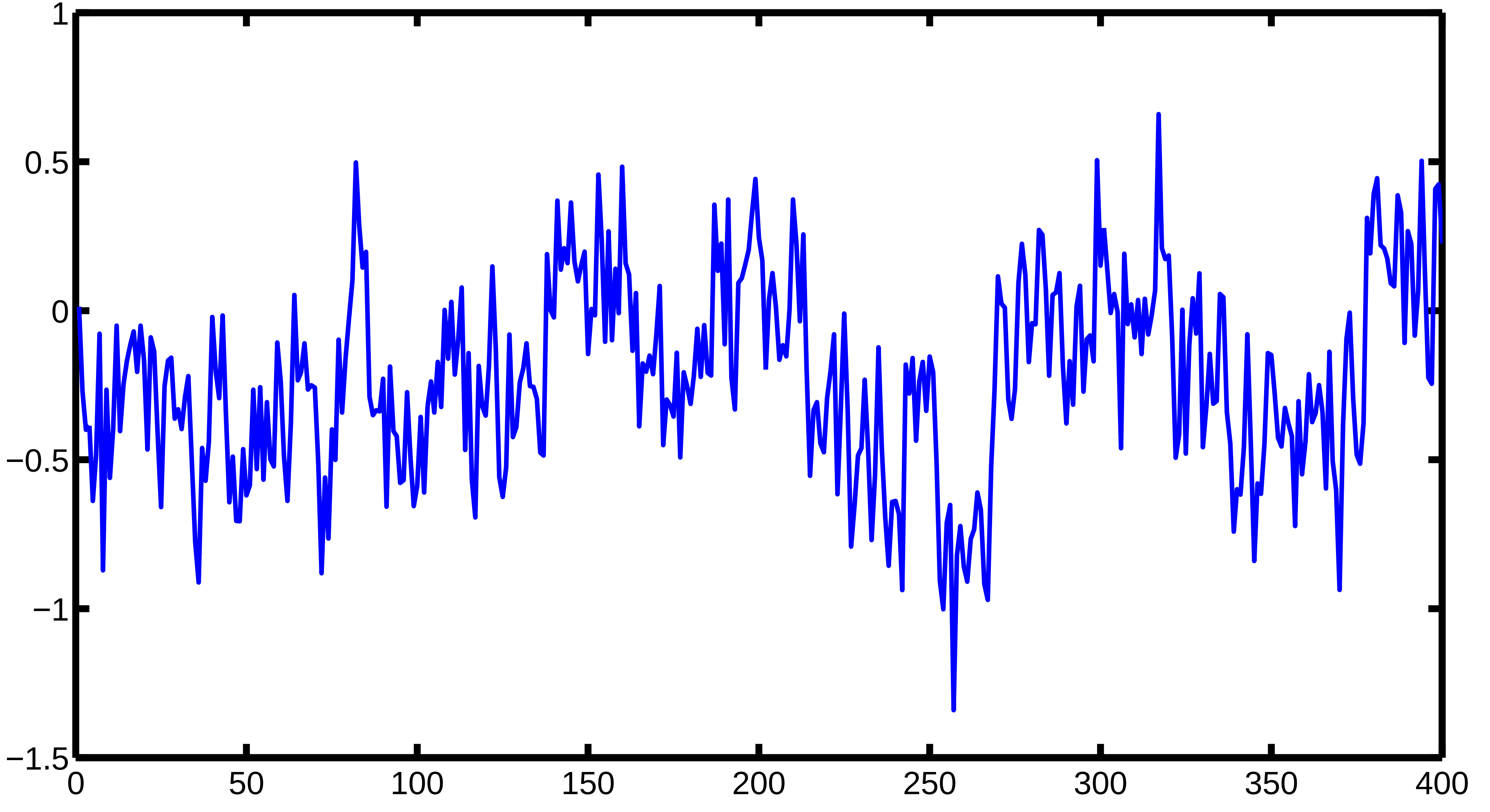
Example fGN: $(H = 0.5)$



Example fGN: ($H = 0.75$)



Example fGN: ($H = 0.99$)



Properties of Self-Similar Process

- Stationary so $\mathbb{E}X_i = 0$, Var$X_i = \sigma^2$ (constant).
- Cov(X_i, X_{i+k}) depends only on the lag k and is given by
$$\gamma(k) = \frac{1}{2}\sigma^2\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right).$$
- Cov$(X_i^{(m)}, X_{i+k}^{(m)})$ is given by
$$\gamma(k) = \frac{1}{2}m^{2(H-1)}\sigma^2\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right).$$
- Asymptotic behavior of the autocorrelation
$$\lim_{k\to\infty}\frac{\rho_k}{k^{2(H-1)}} = H(2H-1).$$
- The variance varies with the aggregation level as
$$\text{Var}X^{(m)} = m^{2(H-1)}\sigma^2,$$

Long-range dependence

Long-range dependence (LRD) for stationary process

- LRD = slow (power-law) decay in the autocovariance

$$\gamma_X(k) \sim c_\gamma |k|^{-(1-\alpha)}$$

  as $k \to \infty$, for some $\alpha \in (0,1)$

- implies for all N

$$\sum_{k=N}^{\infty} \gamma_X(k) \to \infty$$

  this is sometimes used as an alternative definition

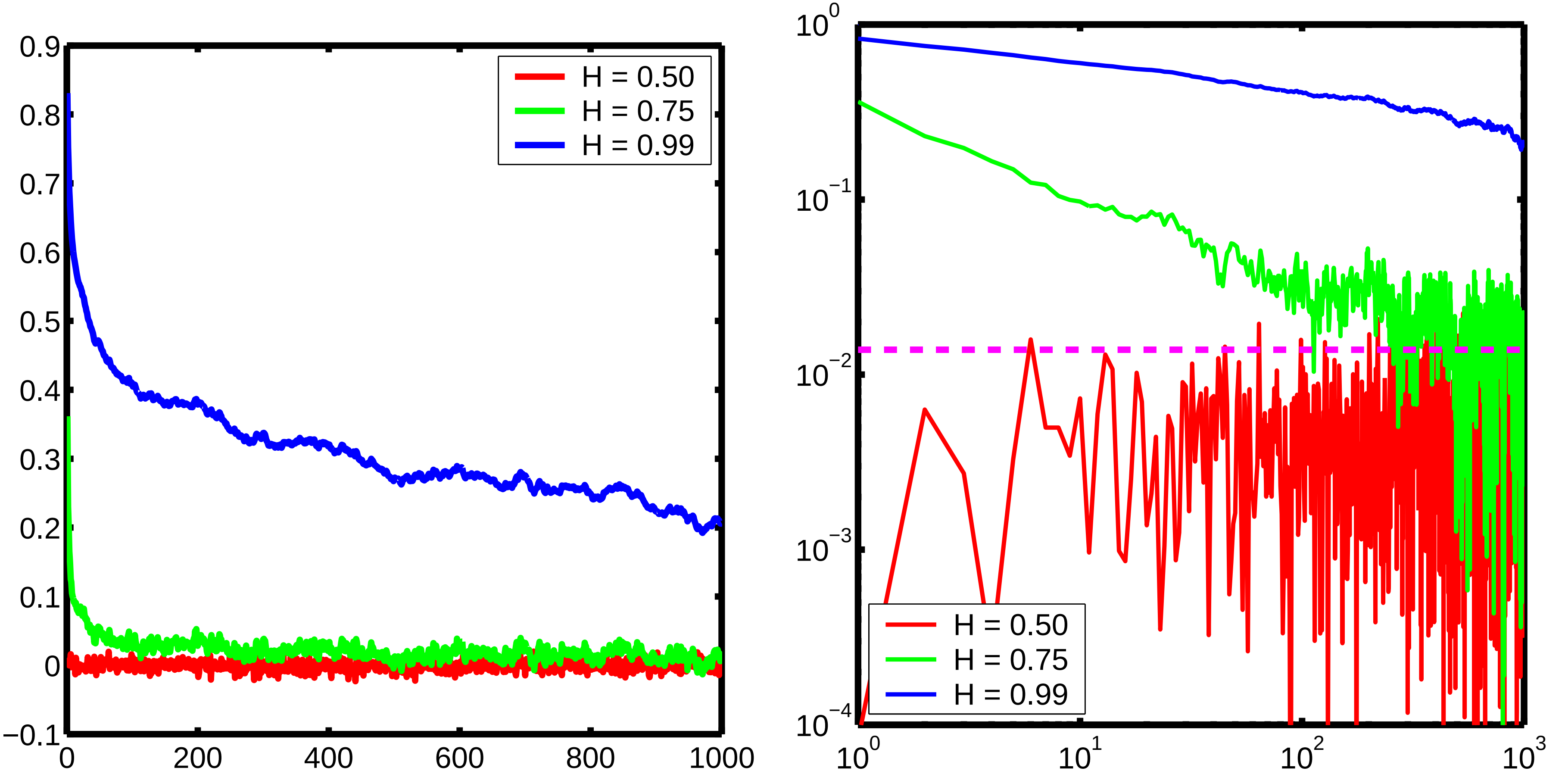
- also called **long-memory process**

LRD and SS

Notice that self-similarity implies LRD with

$$\alpha = 2H - 1$$

for $0.5 \leq H < 1$, and $0 \leq \alpha < 1$



LRD in the frequency domain

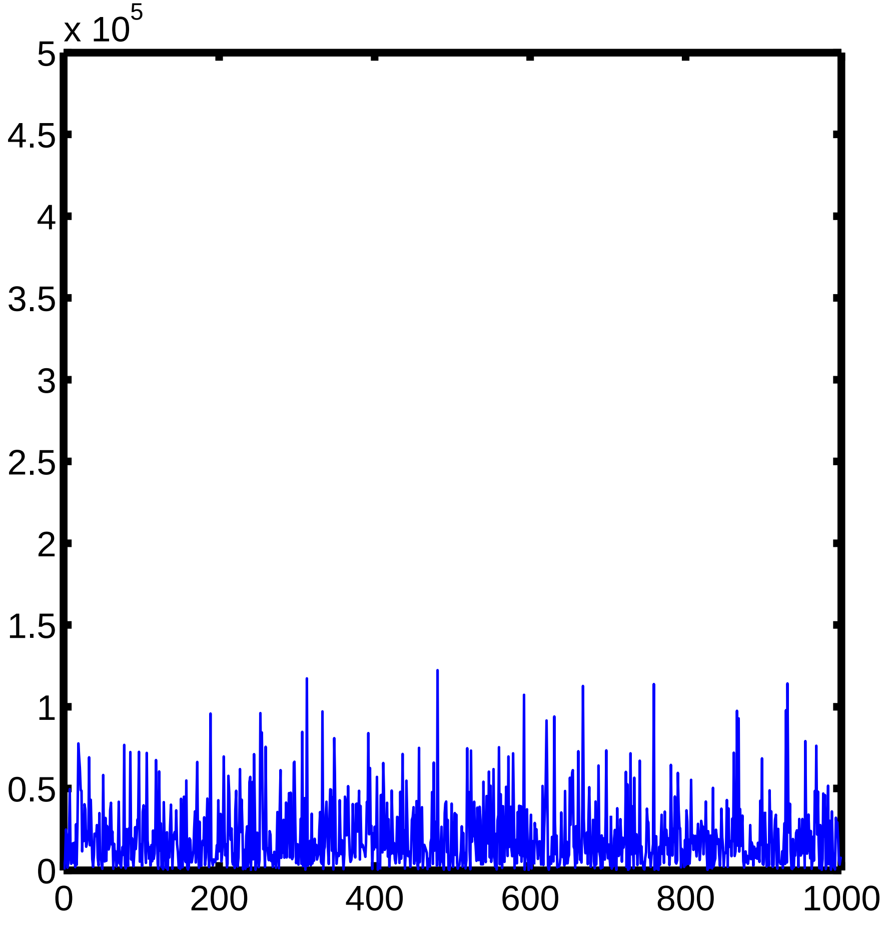
Long-range dependence (LRD) can also be defined in the frequency domain using the Fourier transform of the autocovariance

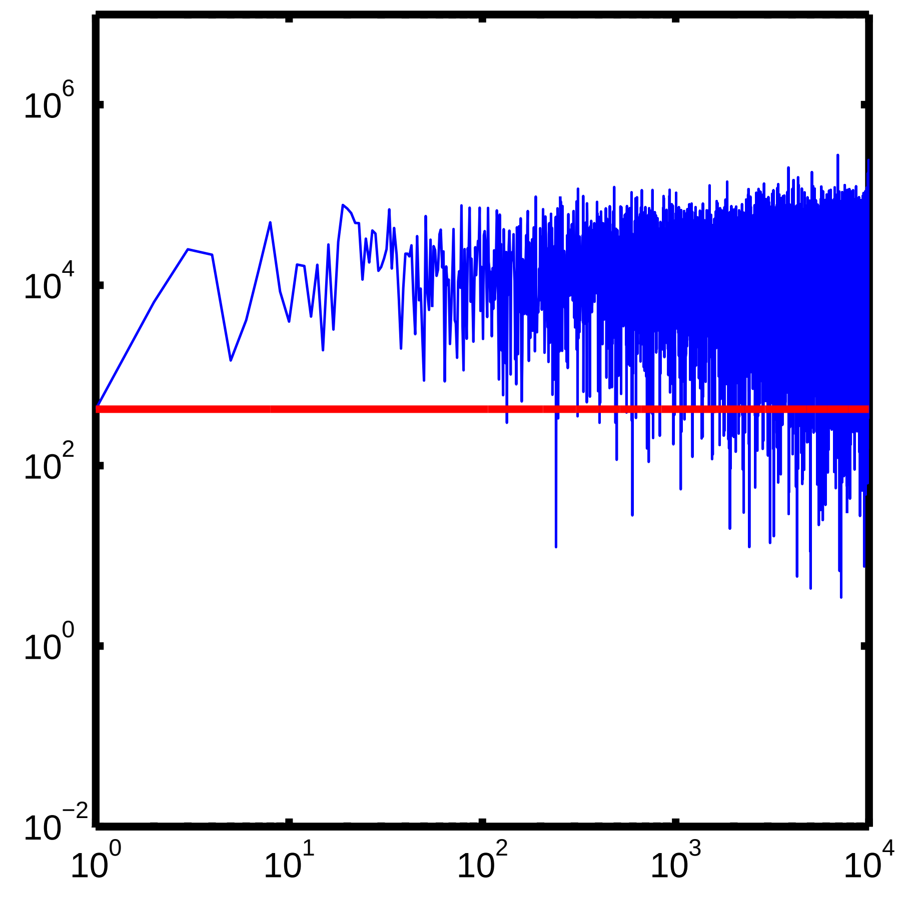
$$f_x(s) \sim c_f |s|^{-\alpha}, |s| \to 0$$

When $\alpha = 1$ we get **1/f noise**, but the term is often applied to the range of values of $\alpha = 2H - 1$.

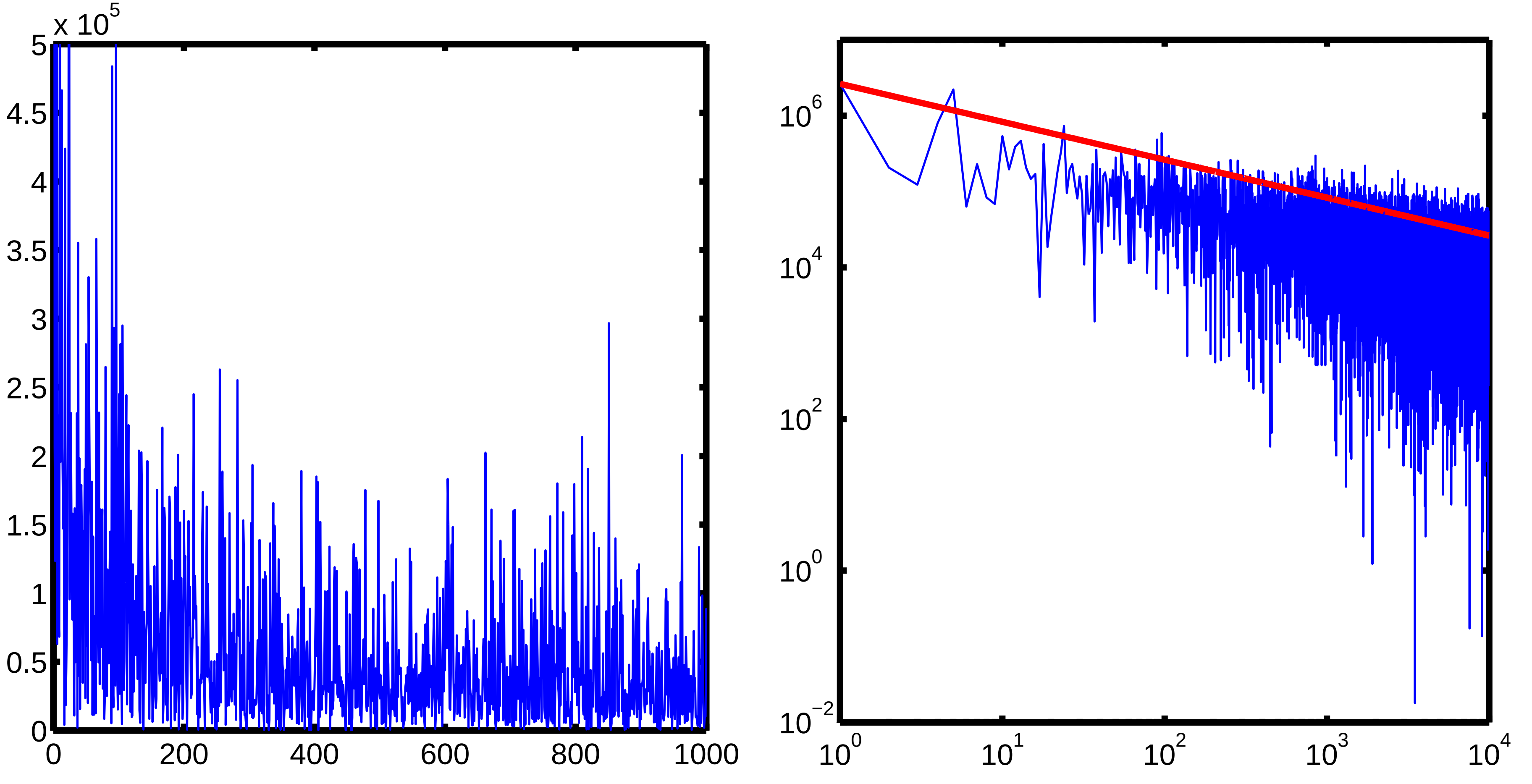
- frequency spectrum of white noise is flat
- frequency spectrum of Brownian motion is $1/f^2$
- frequency spectrum of "pink" noise is $1/f$

Example fGN spectrum ($H = 0.5$)

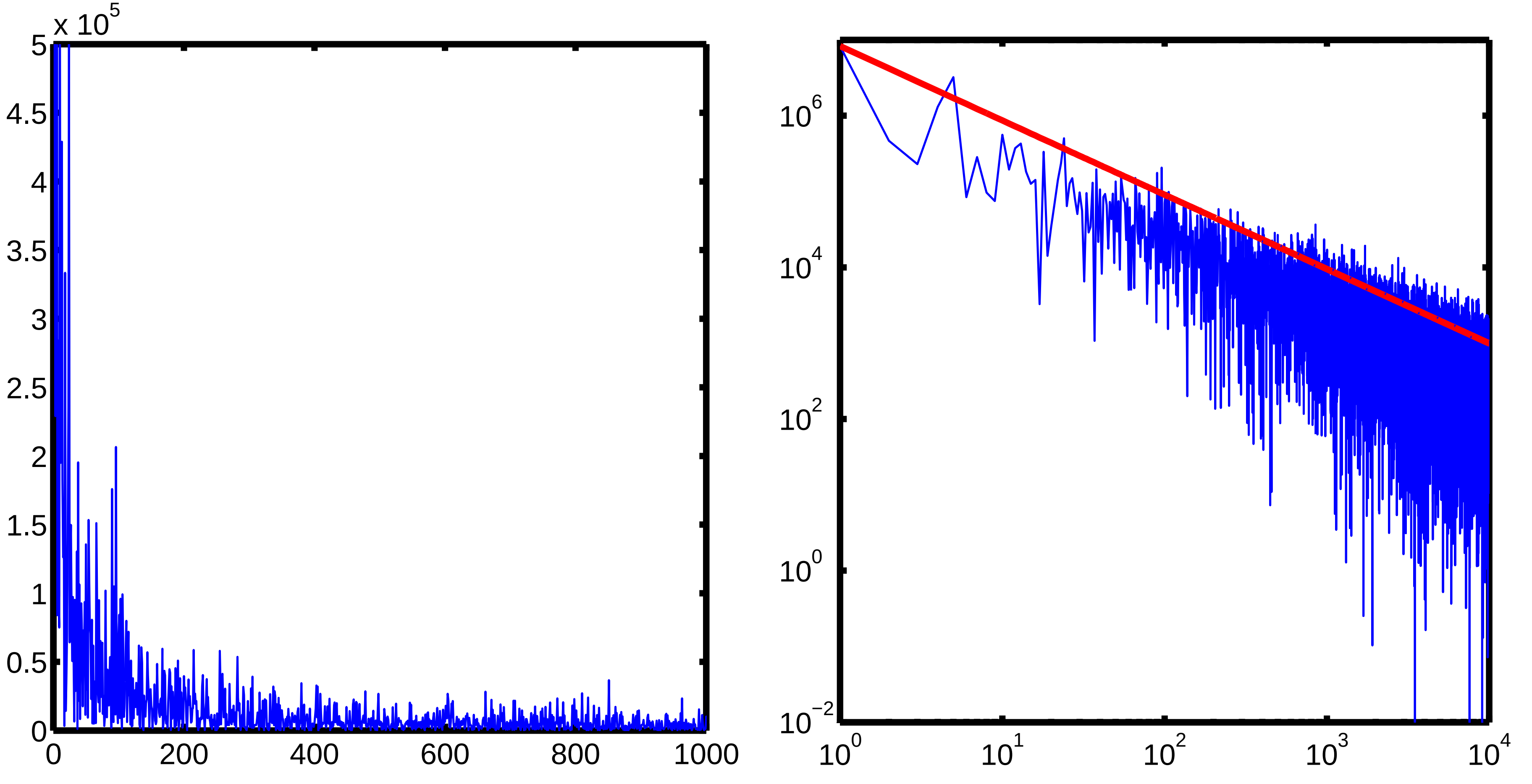




Example fGN spectrum ($H = 0.75$)



Example fGN spectrum ($H = 0.99$)



1/f noise

LRD and SS are also seen elsewhere

- cardiac rhythms (in healthy hearts)
- hydrological data (rainfall, and river flow)
  - Hurst's early work was actually in Nile river data
- music seems to have similar characteristics
- turbulence
- chaotic processes in general
- financial modelling

Connection to Fractals

Fractals more concerned with scaling laws at small scales and high-frequencies

$$f_x(s) \sim c_f |s|^{-\alpha}, |s| \to \infty$$

Hölder exponent $h = (\alpha - 1)/2$

- If $0 < h < 1$ the Hausdorff dimension $D = 5 - \alpha/2$
- If $h < 0$ sample paths are everywhere discontinuous

fractional Gaussian Noise

fGN (fractional Gaussian Noise) is stationary Gaussian process X_t with mean μ, variance σ^2 and autocorrelation function

$$\rho(k) = \frac{1}{2}\left(|k+1|^{2H} - |k|^{2H} + |k-1|^{2H}\right)$$

which asymptotically goes like

$$\rho(k) \sim H(2H-1)|k|^{2H-2}, \quad k \to \infty$$

so $c_\gamma = H(2H-1)$. In the frequency domain,

$$f_x(s) \sim c_f |s|^{1-2H}, \quad |s| \to 0$$

where now

$$c_f = \sigma_Z^2 \cdot 2(2\pi)^{1-2H} H(2H-1)\Gamma(2H-1)\sin(\pi(1-H)),$$

where $\Gamma(x)$ is the gamma function.

fractional Gaussian Noise

Synthesis of fGN:

- Durbin-Levinson: generate white noise, and then impose exact correlation structure. Slow $O(N^2)$ algorithm
- Spectral synthesis:
  - generate white noise
  - take FFT
  - multiply by desired spectrum
  - inverse FFT, to get back to time domain

Note that discrete version of continuous process is no longer exactly self-similar.

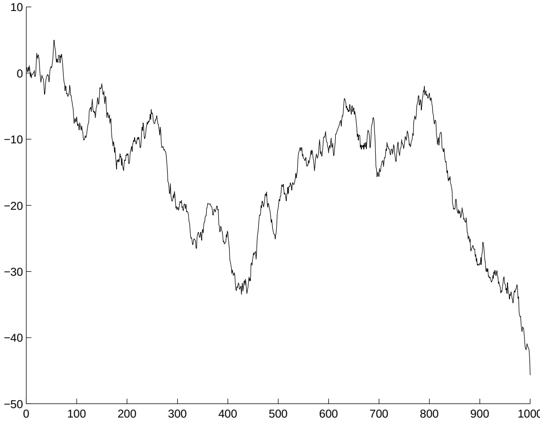
fractional Brownian Motion

The (non-stationary) Gaussian process with covariance function given by

$$\Gamma(s,t) = \frac{1}{2}\sigma^2 \left(s^{2H} - (t-s)^{2H} + t^{2H}\right),$$

variance σ^2 and expectation 0 is called fractional Brownian motion (fBM).

Note the increment process of fBM is fGN, just as the increments of BM are white noise.



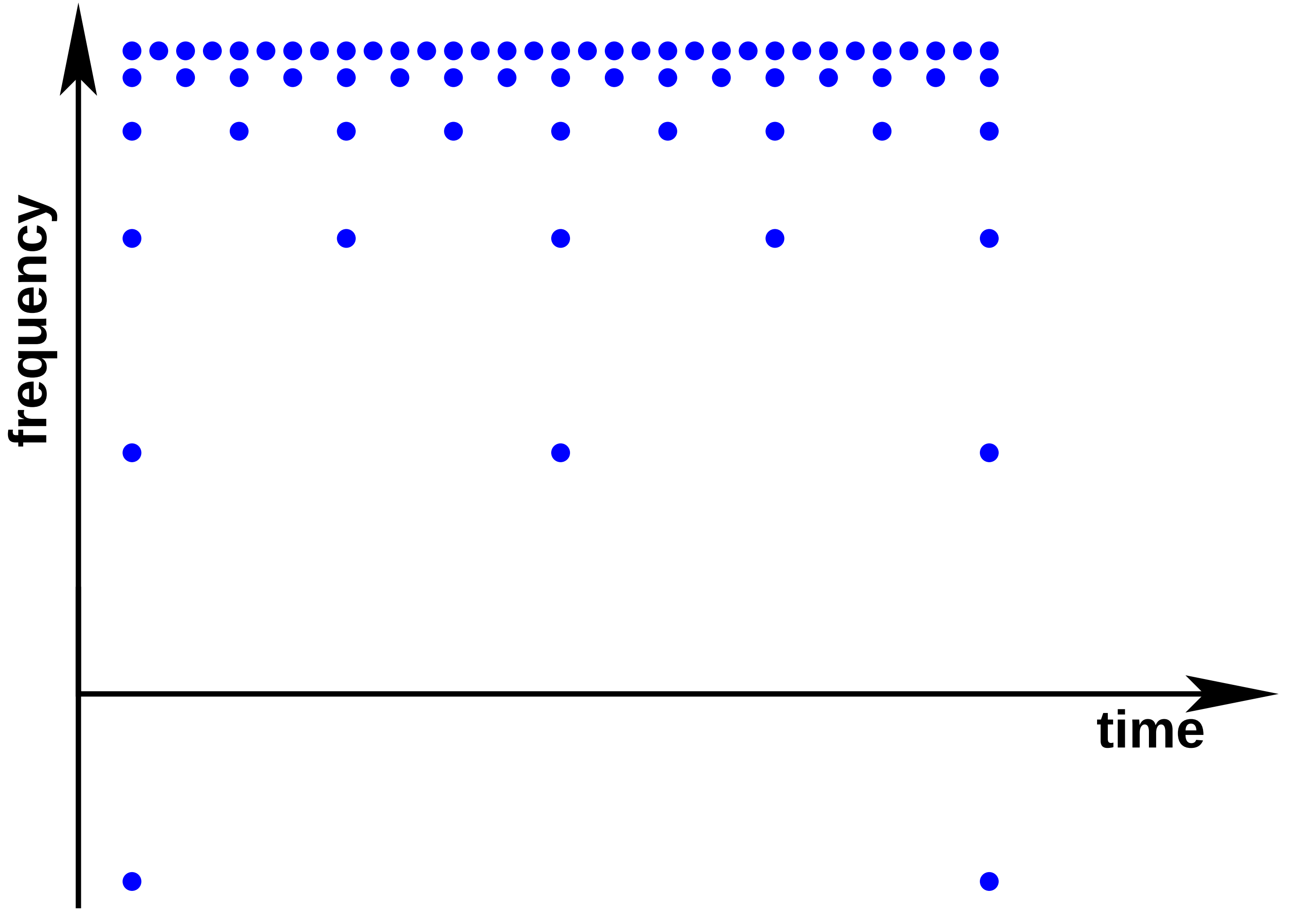
FBM with $H = 0.7$ and $\sigma^2 = 1$.

Wavelets: interpretation

- Multi-Resolution Approximation (MRA)
  - aggregation at different scales is like approximating the data at different scales
  - data stats have known scaling properties
  - a more general way of doing multi-scale approximation is wavelets
- sub-band filters (logarithmically placed)
  - logarithmically placed, so natural log scale arises in frequency domain.
  - sub-bands sampled at frequency appropriate to the bandwidth
  - has the advantage of **de-correlation** of wavelet coefficients

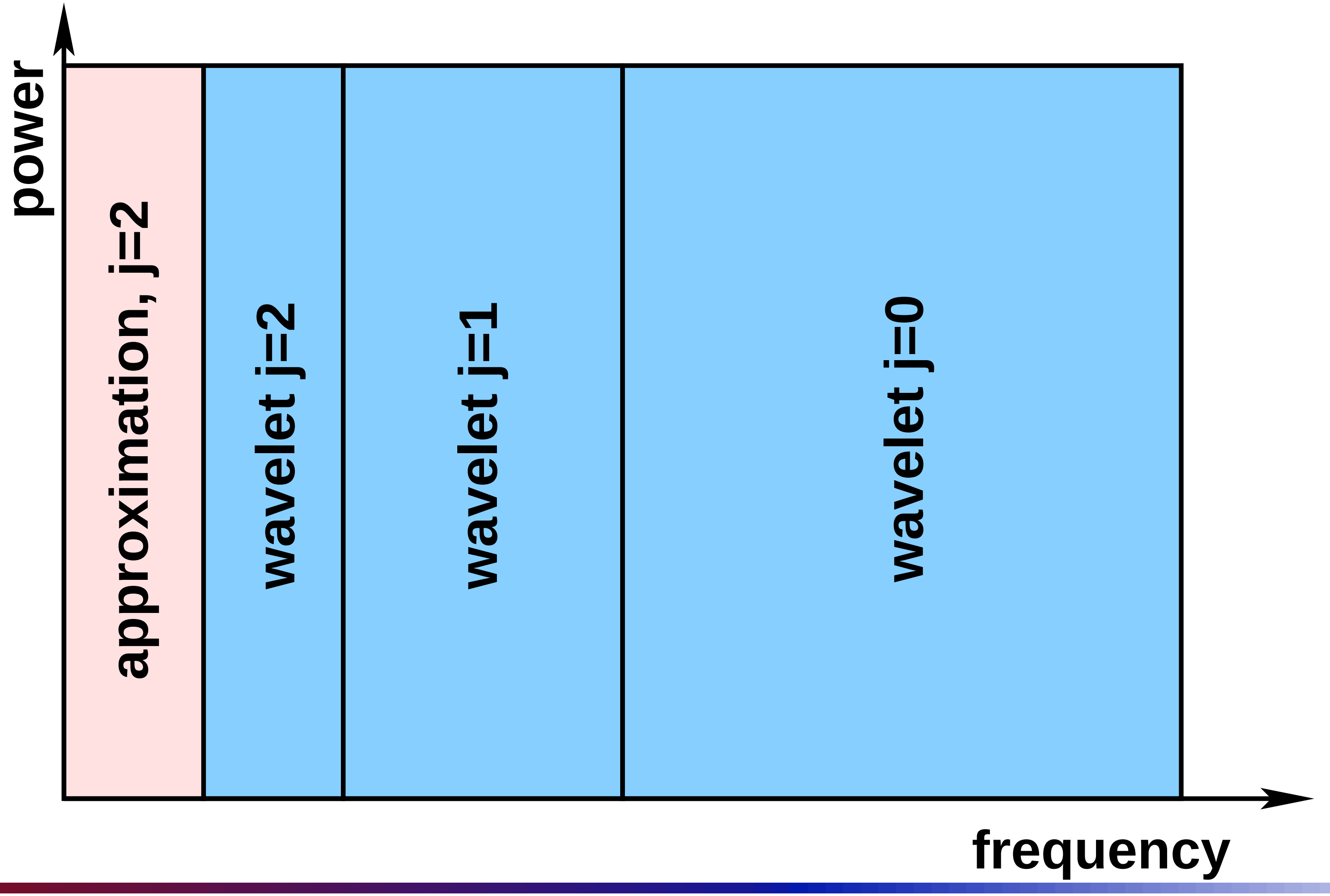
Dyadic grid

Dyadic grid has self-similar scaling behavior!



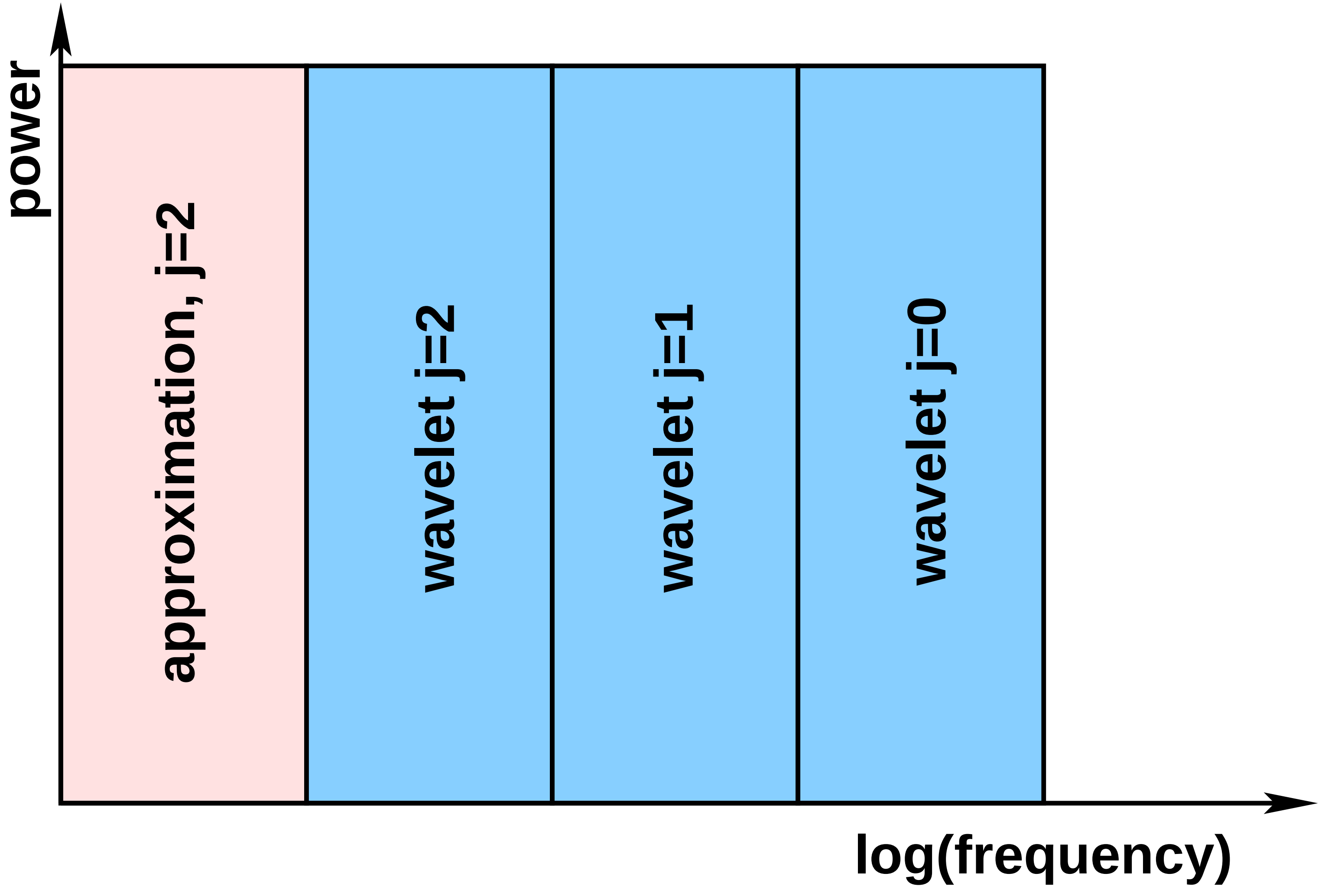
Wavelet's as sub-band filters

The idea (looking across frequencies or scales) is that the transform breaks frequency spectrum into bands.



Wavelet's as sub-band filters

Each band equal size on **log(frequency)** graph



Wavelets and scaling

- the wavelet transform de-correlates details, so can think of each series of $\{d_{j,k}\}_{k\in\mathbb{Z}}$ for each j as a time series, with short-range correlations.
- wavelet conditions ensure

$$E\left[d_{j,k}\right] = 0$$

- we know the distribution of energy in each sub-band
- this translates to energy in each scale of wavelet coefficients $d_{j,k}$, e.g.

$$\text{Var}\left[d_{j,k}\right] = E\left[d_{j,k}^2\right] = \mu_j$$

- we form an estimator of μ_j by

$$\hat{\mu}_j = \frac{1}{N_j}\sum_{k=1}^{N_j} |d_{j,k}|^2$$

Wavelets and scaling

$$f_x(s) \sim c_f |s|^{-\alpha}$$

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{2^j}} \psi^* \left(\frac{t}{2^j} - k \right) dt$$

$$E\left[d_{j,k}^2\right] = 2^{j\alpha} c_f C$$

where

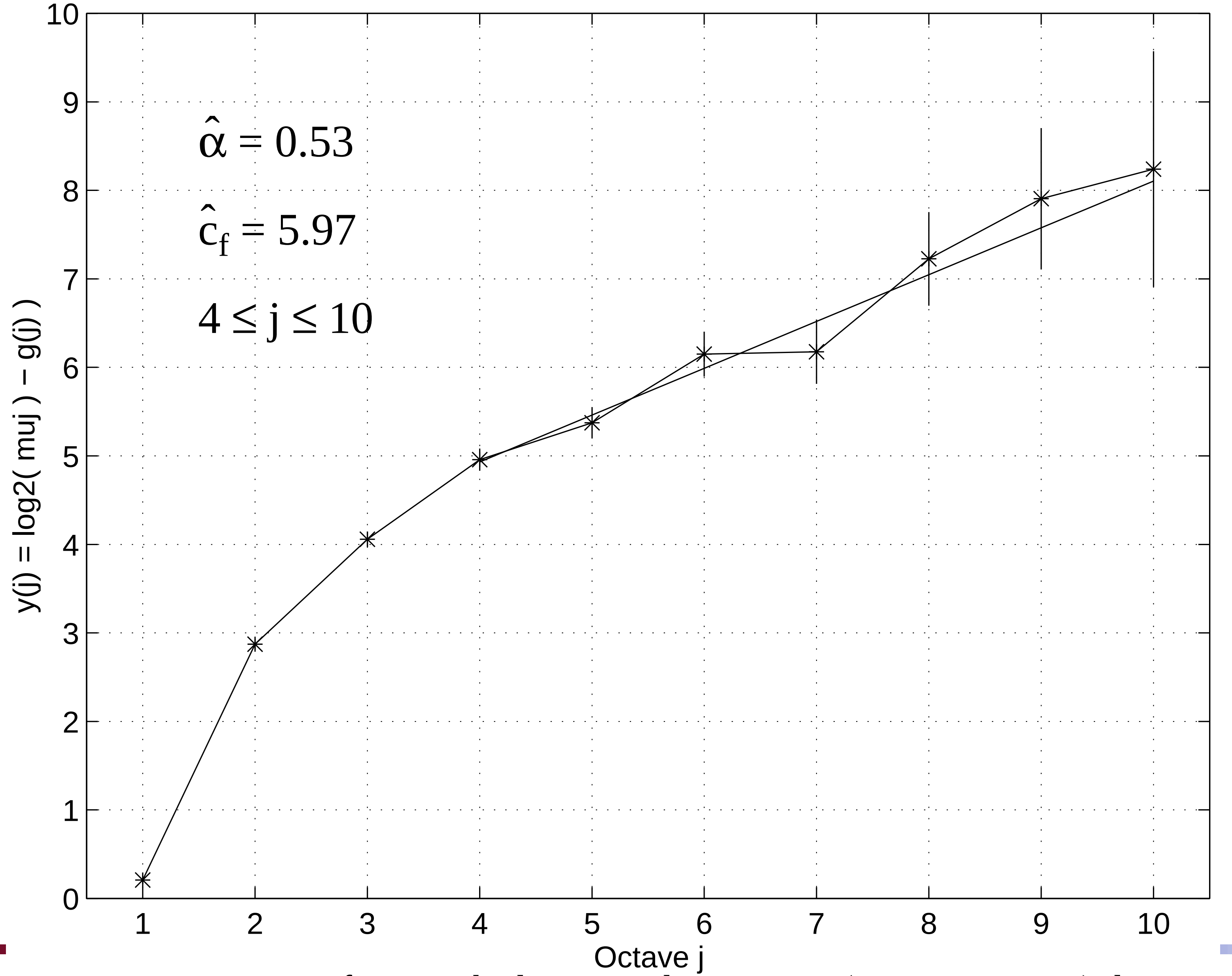
$$C = \int_{-\infty}^{\infty} |s|^{-\alpha} |\Psi^*(s)|^2 ds$$

so

$$\log_2 E\left[d_{j,k}^2\right] = j\alpha + \log_2 c_f C$$

Perform regression on $\log_2 \hat{\mu}_j$ vs the octave j.

Logscale diagram



Logscale diagram

In fact, we can approximate

$$\log_2 \hat{\mu}_j \sim N\left(j\alpha + \log_2 c_f C, \frac{2^{j+1}}{n \ln^2 2} \right)$$

So we can

- estimate confidence intervals for $\log_2 \hat{\mu}_j$ on the Logscale diagram
- perform a weighted regression
- estimate covariance of estimates of α and c_f
- actually worth adding a small correction to get $y_j = \log_2 \mu_j - g_j$ (because log and expectation don't commute)

Wavelet estimator properties

- asymptotically efficient and unbiased
  - almost as accurate as Whittle (MLE)
- joint estimator of H and c_γ
- known variance of estimates
- robustness
  - non-Gaussianity
  - trends in the data
  - short-range correlative structure
  - much better than Whittle in these cases

http://www.cubinlab.ee.mu.oz.au/~darryl/secondorder_code.html/

Non-standard sampling

Shannon theorem

"If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $(1/2W)$ s apart."

Claude Shannon, "Communications in the presence of noise", Proc.IRE, 37, pp.10–21, 1949.

- uniform sampling
  - samples spaced a uniform distance apart
- Nyquist limit

  H.Nyquist, "Certain topics in telegraph transmission theory", AIEE Trans., 47, pp.617-644, 1928.
- Implicitly, we can reconstruct $f(t)$ from its samples
  - if the signal is bandlimited

Shannon theorem

Proof sketch: Assume function is bandlimited so $F(s) = 0$ for $|s| > W$, then the IFT is

$$f(t) = \int_{-\infty}^{\infty} F(s) e^{i2\pi st} \, ds = \int_{-W}^{W} F(s) e^{i2\pi st} \, ds$$

If instead, we make, F periodic, with period $2W$ then we can find a Fourier series for it, e.g.

$$F(s) = \sum_{i=-\infty}^{\infty} A_n e^{i\pi ns/W}$$

where,

$$A_n = \frac{1}{2W} \int_{-W}^{W} F(s) e^{-i\pi ns/W} \, ds = \frac{1}{2W} f\left(\frac{n}{2W}\right)$$

Shannon theorem

Proof sketch:
We can represent $F(s)$ perfectly with the Fourier series coefficients A_n, but these are just proportional to the function sampled at uniform intervals, e.g. $A_n \propto f\left(\frac{n}{2W}\right)$.

Hence, the samples completely define the FT F, and hence the function f. $\square$

Shannon interpolation

Reconstruction of original signal from IFT

$$
\begin{aligned}
f(t) &= \int_{-W}^{W} F(s) e^{-i2\pi st}\, ds \\
&= \int_{-W}^{W} \sum_{i=-\infty}^{\infty} A_n e^{i\pi ns/W} e^{i2\pi st}\, ds \\
&= \sum_{i=-\infty}^{\infty} A_n \int_{-\infty}^{\infty} r(s/2W) e^{i2\pi s(-t+n/2W)}\, ds \\
&= \sum_{i=-\infty}^{\infty} 2WA_n \int_{-\infty}^{\infty} r(-s) e^{i2\pi s(2Wt-n)}\, ds \\
&= \sum_{i=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \operatorname{sinc}(2Wt-n)
\end{aligned}
$$

Shannon interpolation

Assume we sampled at the Nyquist rate, i.e. $f_s = 2W$, or $t_s = 1/2W$, then the sample points would be

$$f\left(\frac{n}{2W}\right)$$

The summation

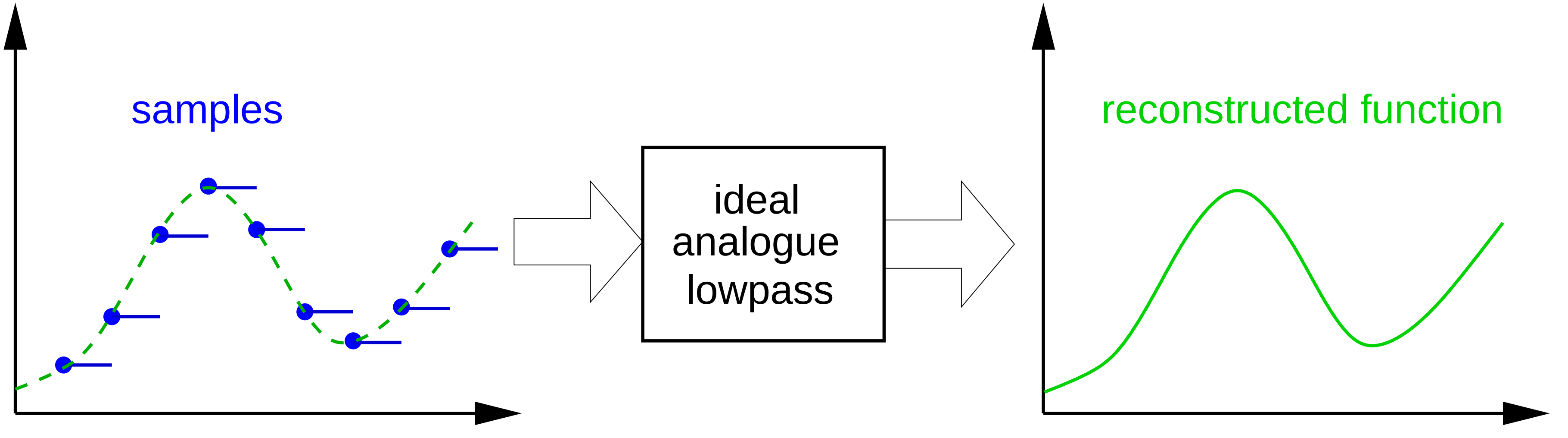
$$f(t) \quad = \quad \sum_{i=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \operatorname{sinc}(2Wt - n)$$

represents a convolution of the sampled signal with a sinc function. Now we know the sinc has a simple rectangular transfer function, and so it acts as a perfect low-pass filter.

Shannon interpolation

Interpretation

- convolution with sinc
- equivalent to ideal (rectangular) bandpass filter



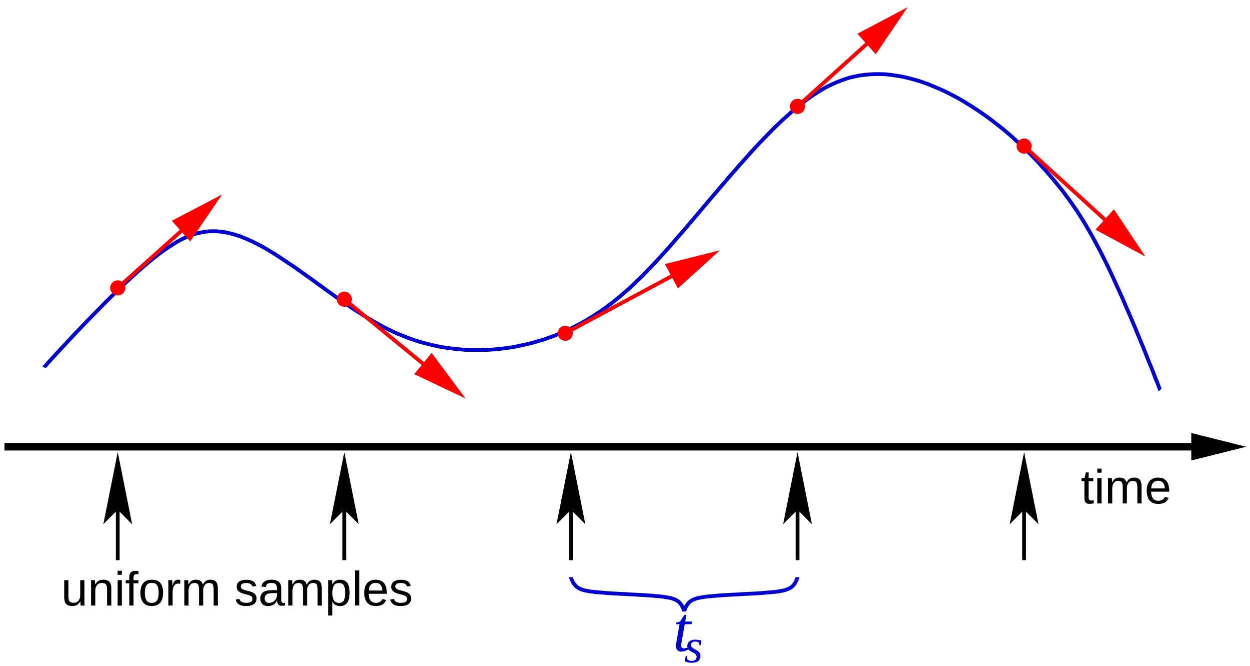
- this is essentially what a Digital to Analogue converter tries to do
- have to build analogue filter — hard to make it ideal

Other sampling schemes

- dyadic grid (wavelets)
- ordinate and slope sampling
- interlaced sampling
- implicit sampling
- irregular sampling
- hexagonal sampling
- many others ...

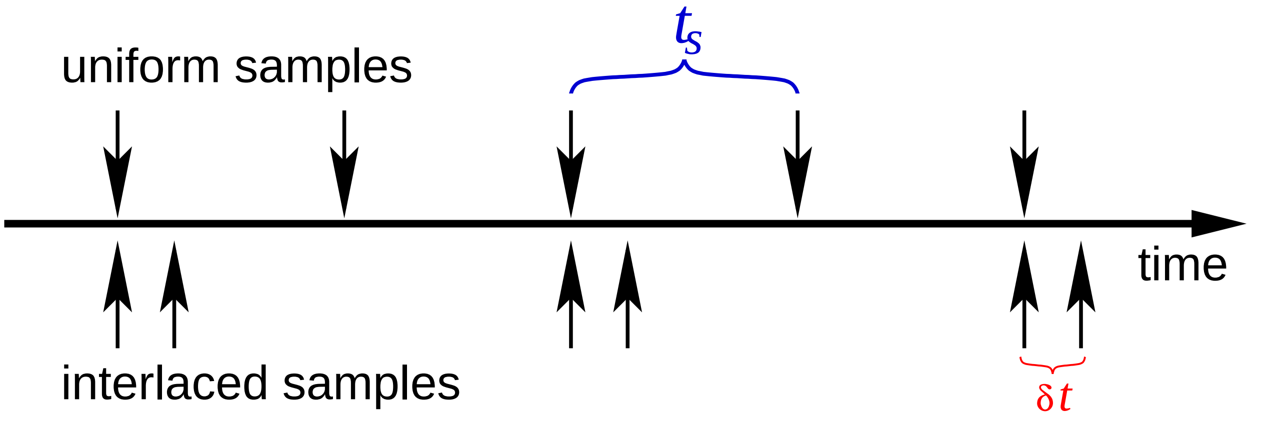
Ordinate and Slope Sampling

- sample the value, and derivative at a point



- Shannon theorem for ordinate/slope sampling
  We can reconstruct a function from knowledge of its ordinate and slope at every other sample point.
- e.g. half the Nyquist sampling rate

Interlaced sampling



- signal is uniquely determined given a series of samples at **recurrent** sample points

$$t_{pm} = t_p + \frac{mN}{2W}$$

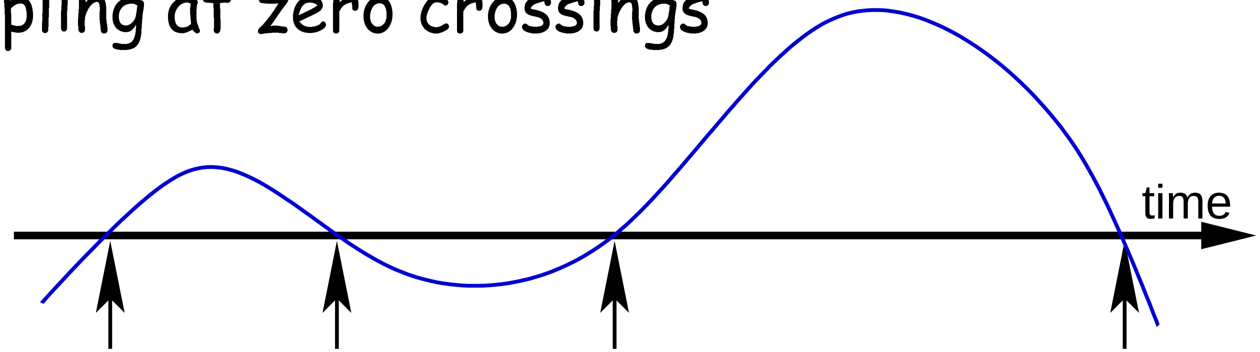
for $p = 1, 2, \ldots, N$ and $m \in \mathbb{Z}$

  - interlaced sampling example above has $N = 2$

- limit $\delta t \to 0$, is equivalent to ordinate/slope sampling

Implicit sampling

- e.g. sampling at zero crossings



- Applications:
  - specify filter by zero crossings
  - reconstruct an image

$\Rightarrow$

Implicit sampling theory

- "Information in the Zero Crossings of Bandpass Signals", B.F. Logan, Bell System Tech. Journal, 56, pp. 487-510, April 1977.
  - a signal is uniquely reconstructible from its zero crossings if
    - The signal $x(t)$ and its Hilbert transform $X_H(t)$ have no zeros in common with each other.
    - The frequency domain representation of the signal is at most 1 octave long, in other words, it is bandpass-limited between some B and 2B.
- "Reconstruction of Two-Dimensional Signals From Threshold Crossings", A. Zakhor and A. V. Oppenheim, Proceedings of the IEEE, January 1990, vol. 78, no. 1, pp. 31-55.

Irregular sampling

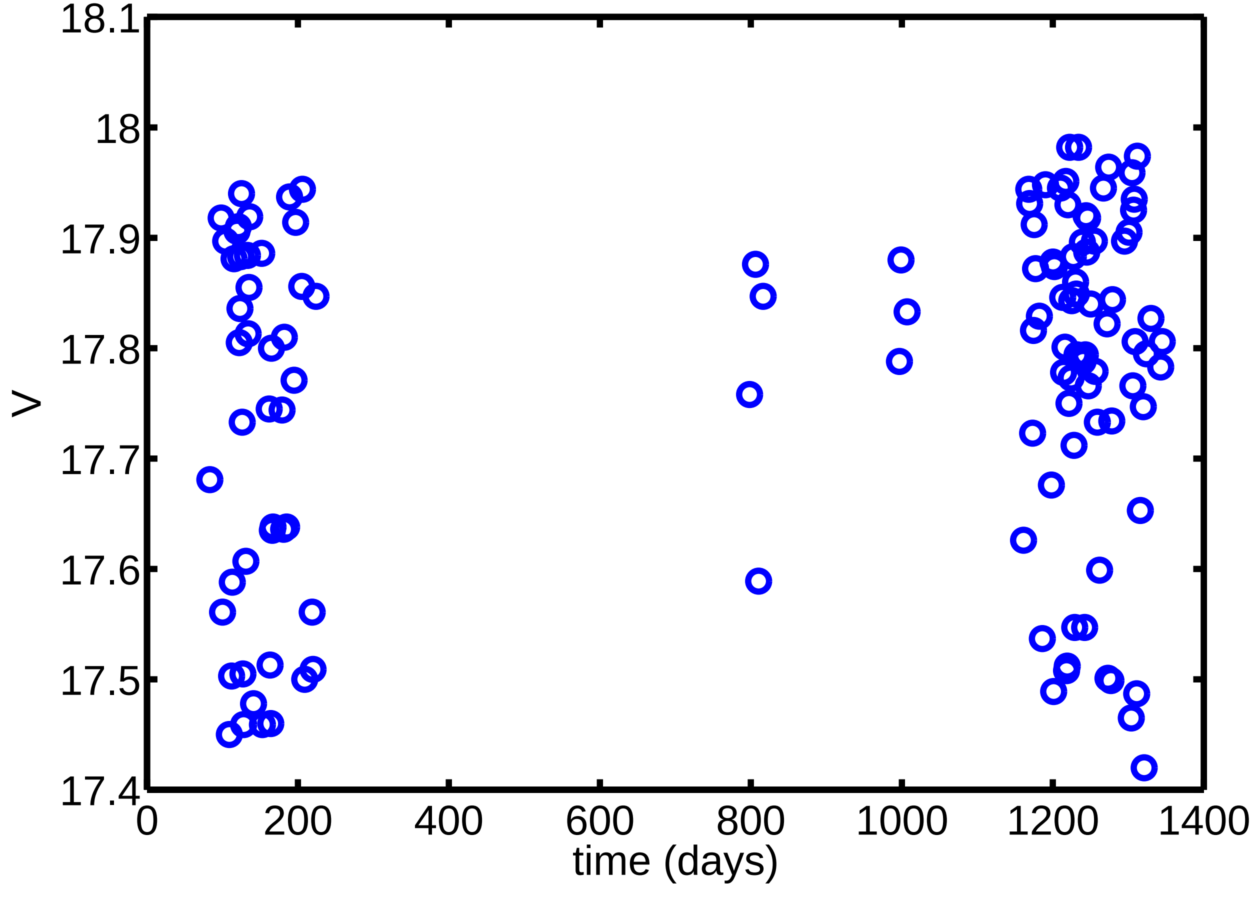
- not all sampling is on a regular grid
  - Astronomical data depends on when you can make observations
    - clouds might get in the way
  - Geophysical data
    - depends on which rock strata you can find
  - Poisson sampling used in Internet performance measurements
  - even regular samples have jitter
- all previous work assumed regular sampling
  - how can we deal with irregularity?

Non-bandlimited signals

- we can't always pre-filter analogue signal with a band-pass before sampling
  - Astronomical data can't be obtained between samples (e.g. clouds)
  - Internet performance measurements are made with probe packets
  - Acoustic measurements of position of an object
    - bounce ultrasound pulse off an object every half a second
    - don't see what happens in between
- aliasing is a problem without pre-filtering
  - how can we cope without pre-filtering?

Astronomical data

- apparent magnitude of a variable star



data courtesy of Laurent Eyer, <Laurent.Eyer@obs.unige.ch>
http://obswww.unige.ch/~eyer/

Astronomical data

- we can see
  - data are not uniformly spaced
    - there is no way to “fix” this
  - no obvious period
- no pre-filter has been applied to the samples
- can we still look for periodicities in the data?

Periodogram

- for uniformly sampled data X_n, use the periodogram

$$P_X(k) = \frac{1}{N}|FT_X(k)|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1} X_n e^{-i2\pi kn/N}\right|^2.$$

- rewrite complex exponential in terms of trig.fn.s

$$P_X(k) = \frac{1}{N}\left[\left(\sum_{n=0}^{N-1} X_n \cos(2\pi kn/N)\right)^2 + \left(\sum_{n=0}^{N-1} X_n \sin(2\pi kn/N)\right)^2\right].$$

- write in terms of frequency $f = k/(Nt_s)$ and sample times $T_n = nt_s$

$$P_X(f) = \frac{1}{N}\left[\left(\sum_{n=0}^{N-1} X_n \cos(2\pi f T_n)\right)^2 + \left(\sum_{n=0}^{N-1} X_n \sin(2\pi f T_n)\right)^2\right].$$

Lomb-Scargle Periodogram

- for irregularly sampled data we use the Lomb-Scargle periodogram

$$P_X^{(LS)}(f) = \frac{1}{2}\left[\frac{\left(\sum_{k=0}^{N-1}(X(T_k)-\bar{X})\cos(2\pi f(T_k-\tau))\right)^2}{\sum_{k=0}^{N-1}\cos^2(2\pi f(T_k-\tau))} + \frac{\left(\sum_{k=0}^{N-1}(X(T_k)-\bar{X})\sin(2\pi f(T_k-\tau))\right)^2}{\sum_{k=0}^{N-1}\sin^2(2\pi f(T_k-\tau))}\right],$$

where $\bar{X}$ is the mean value of X_n and τ satisfies

$$\tan(4\pi f\tau) = \frac{\sum_{k=0}^{N-1}\sin(4\pi f T_k)}{\sum_{k=0}^{N-1}\cos(4\pi f T_k)}.$$

Lomb-Scargle Periodogram explained

- think of a periodogram as fitting sine and cosine functions to the data
  - standard periodogram does a least-squares fit
    - assuming uniform samples
  - Lomb-Scargle Periodogram does the same
    - but allowing arbitrary sampling
- τ allows a shift in time to make everything time-shift invariant
- Fast $O(N \log N)$ variants exist (similar to FFT)

Nyquist limits

For uniform sampling, we must obey Nyquist limit

- or we get aliasing

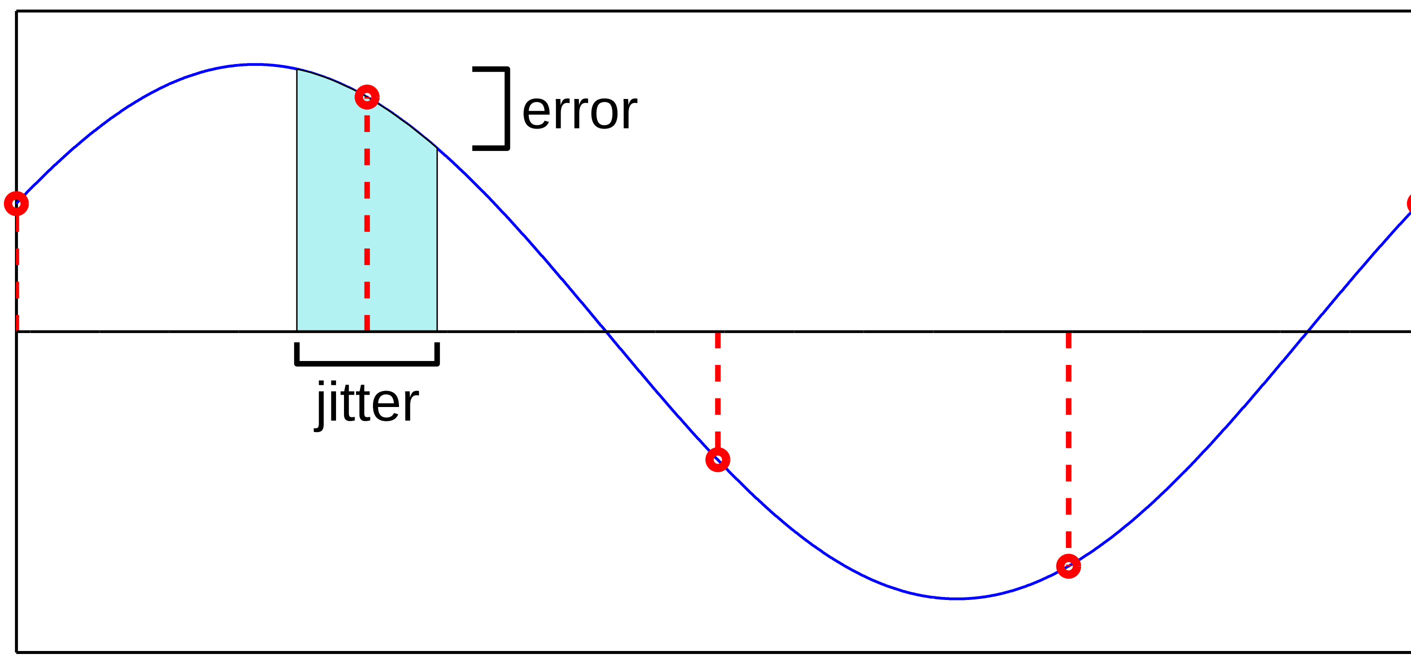
For non-uniform sampling, we don't need to follow the standard (uniform sampling) Nyquist limit

- we don't need to bandpass signal before sampling!

Nyquist limits

Intuition:

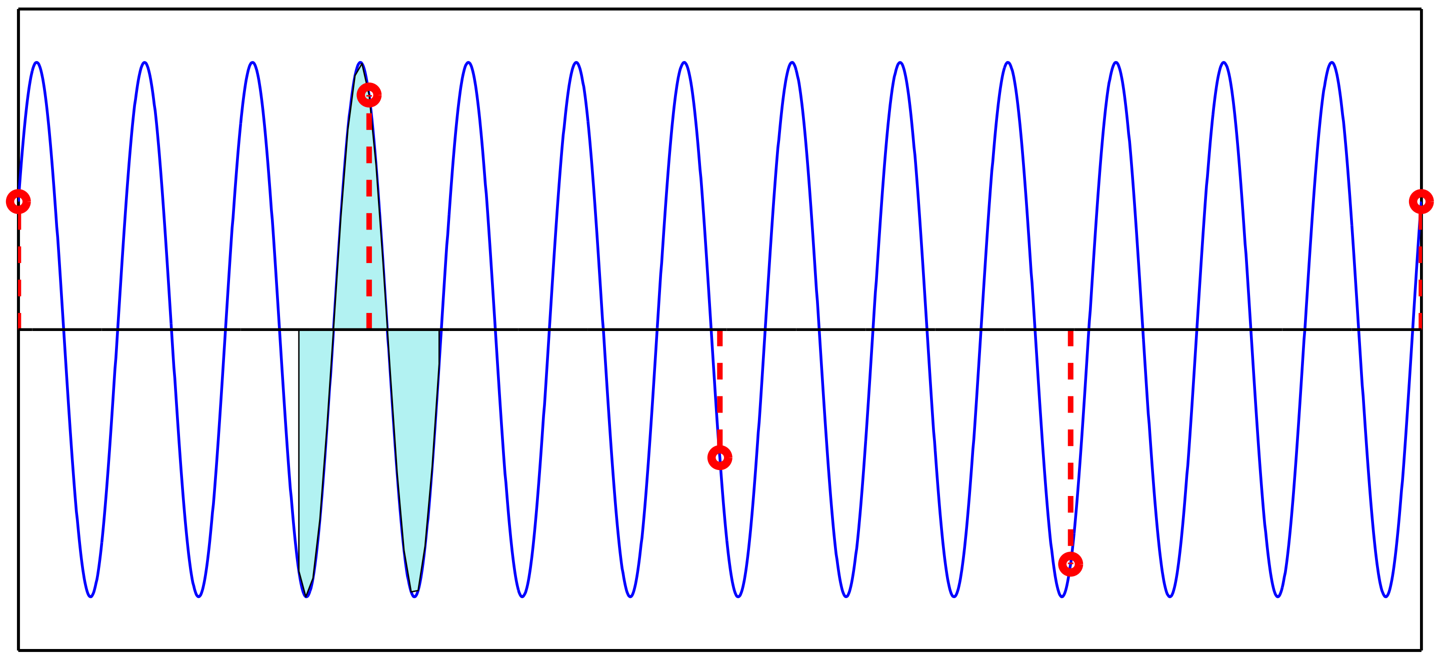
- for low-frequency, jitter in sampling time, is equivalent to error, or similar order of magnitude in sample value



Nyquist limits

Intuition:

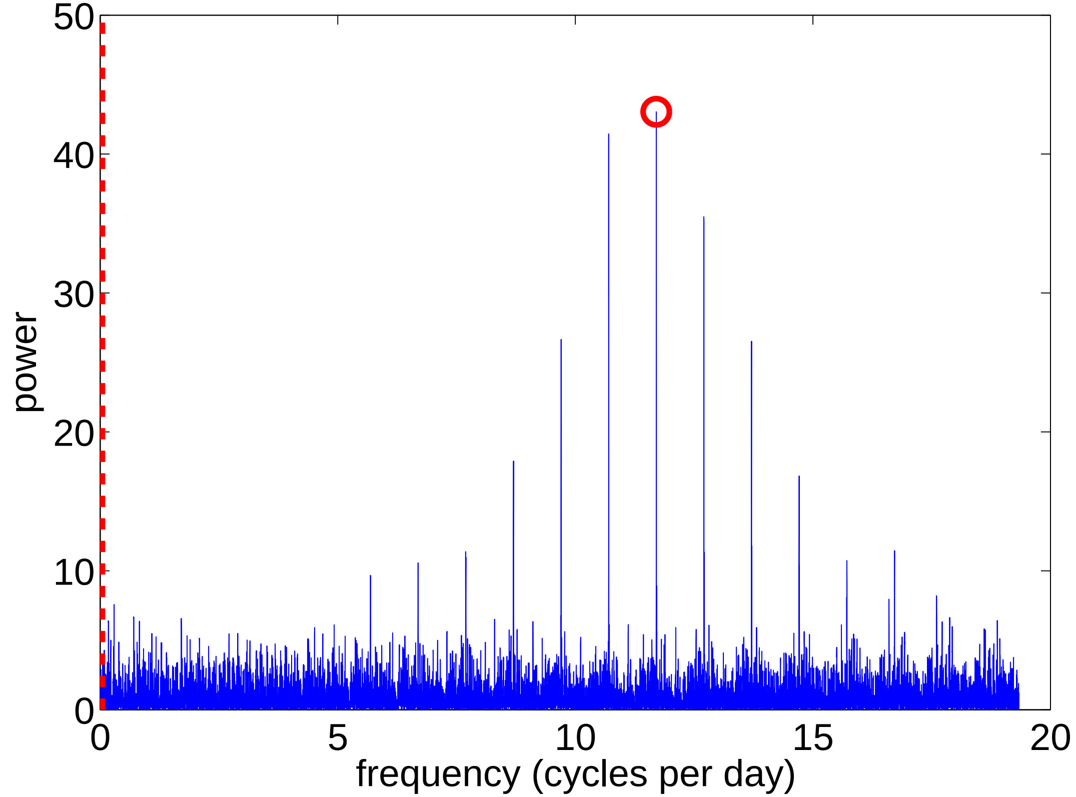
- for high-frequency, jitter in sampling time, introduces errors of similar magnitude to signal



In some sense, there is some filtering going on here.

Lomb-Scargle Periodogram examples

- variable star data from before



Average measurement interval = 10.427 days.

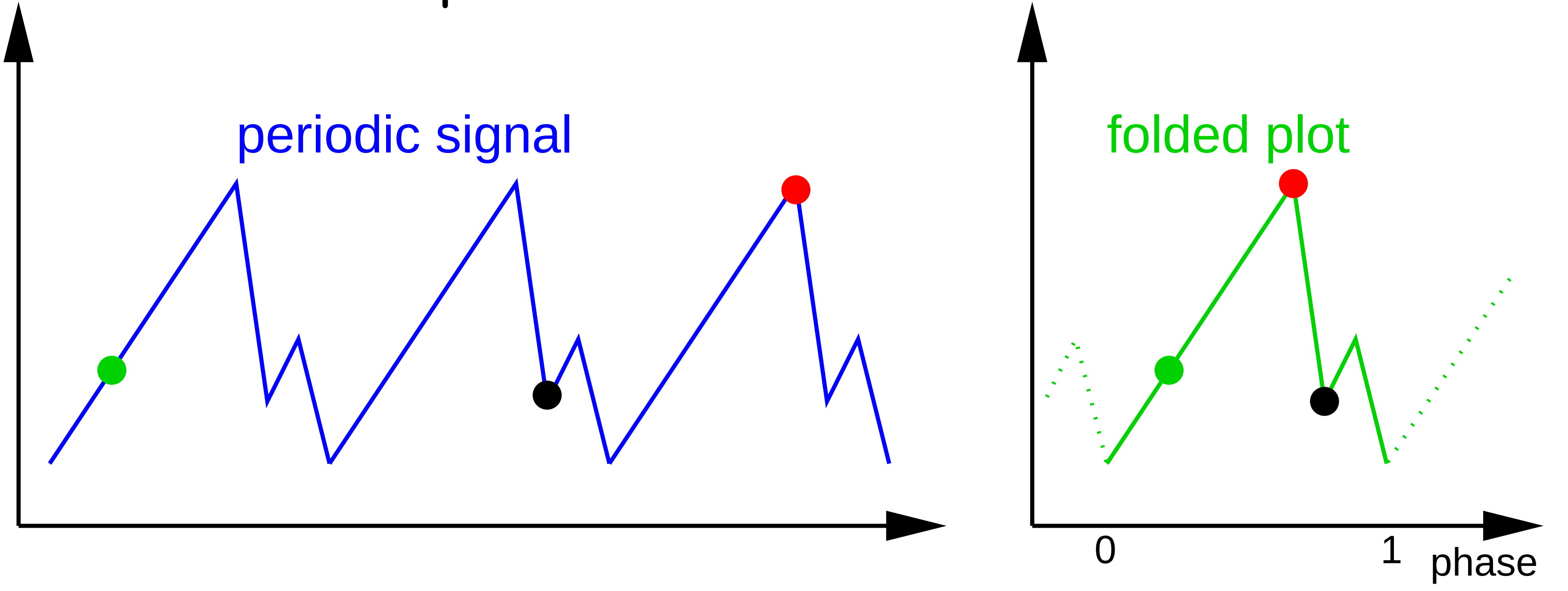
Nyquist frequency $\simeq 1/10$-th cycle per day.

Peak is at 11.7 cycles per day.

Folded Plots

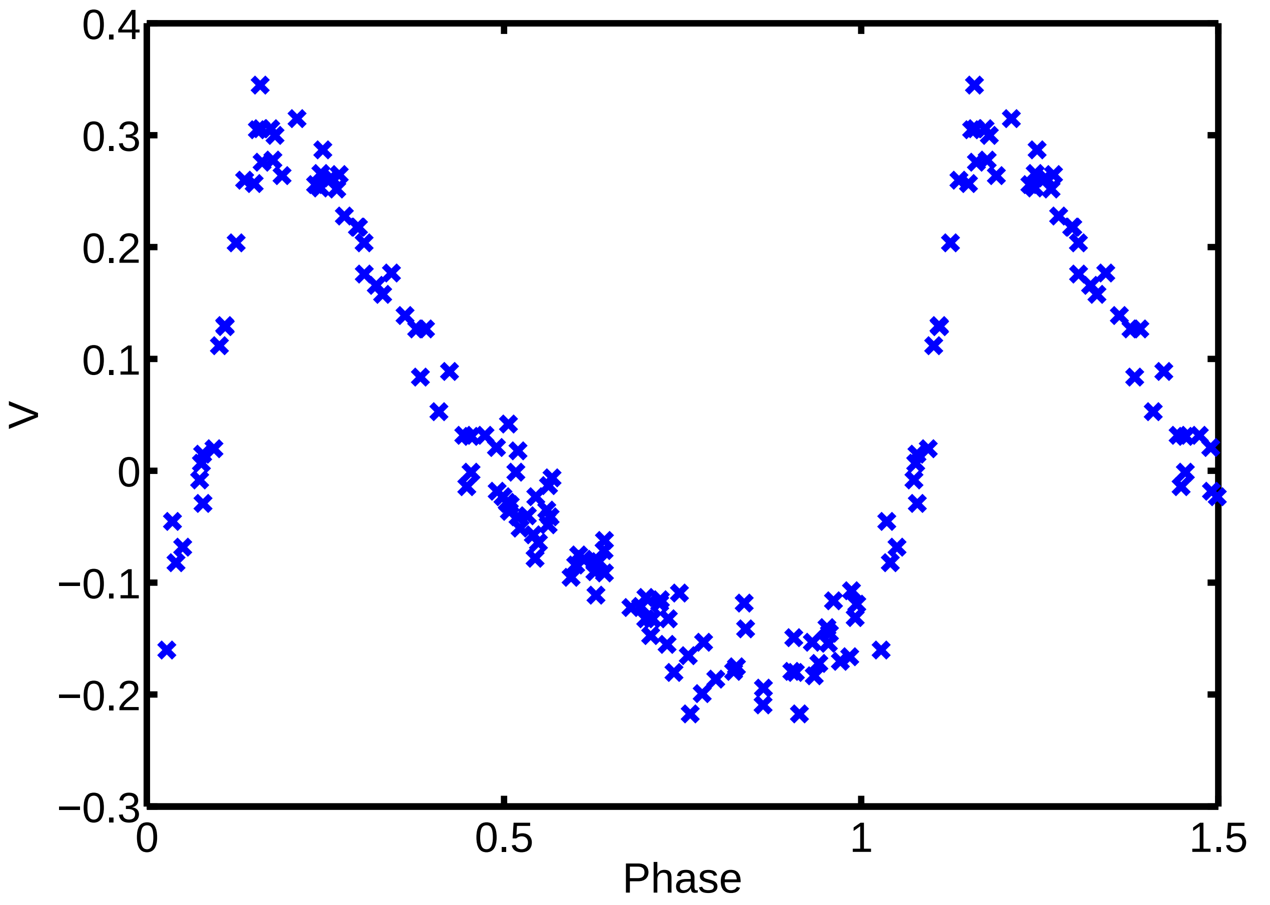
Superimposes a time series upon itself with respect to a specified period.

- if period of fold is correct, then measurements would line up

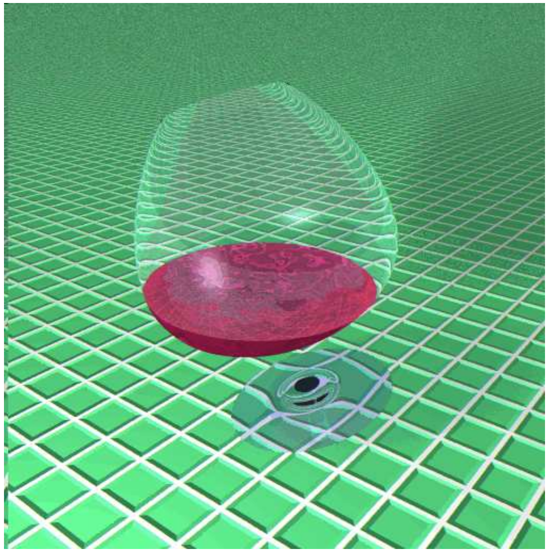


Folded Plot example

- variable star data from before
  - period 11.7 cycles per day



2D irregular sampling: CGI jittering



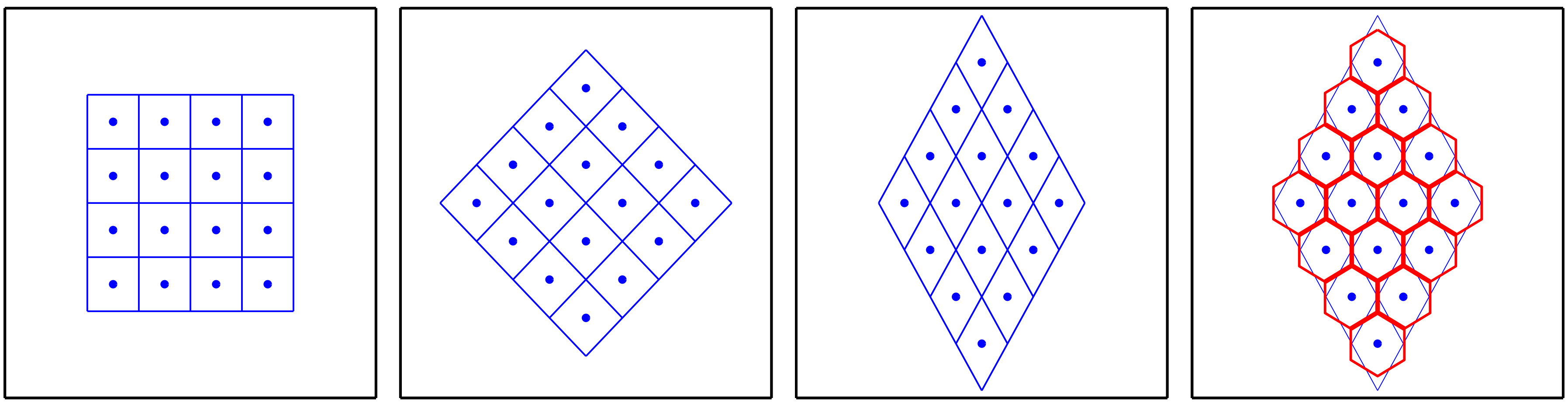
- CGI anti-aliasing by jittering points
  - equivalent to irregular sampling in 1D
  - typically sample irregularly at higher resolution than needed
  - then low-pass (by averaging)
  - don't use this for animations (only stills)

2D possibilities: Hex grids

- sample onto hexagonal grid
  - pixels have nearly circular shape
    - better match to physical systems
      - e.g. printer dots
  - different symmetries
  - better behaved connectivity
    - only one case
      - not edge + corners as for squares
  - Improved Angular Resolution. With more lateral neighbors, curves and edges can be followed more easily and accurately

Hexagonal grids

- we can get a hexagonal sampling grid by
  - start with a rectangular grid
  - rotate by 45 degrees
  - stretch so that adjacent samples are equi-distant



rectangular grid | rotate | stretch vertically | hexagonal grid

Hexagonal Fourier Transform

- transforms above tell us how to take FT
  - rotating an image
    $\Rightarrow$ rotate FT
  - stretch image (in one direction)
    $\Rightarrow$ squeeze the FT in the same direction
- in square grid distance between samples
  - horizontal or vertical distance is 1
  - diagonal, distance is $\sqrt{2}$
  - Nyquist frequency is different for diagonal
- in hex grid distance between samples
  - is always one
  - Nyquist frequency is same in six directions

Sparse signals and compressive sensing

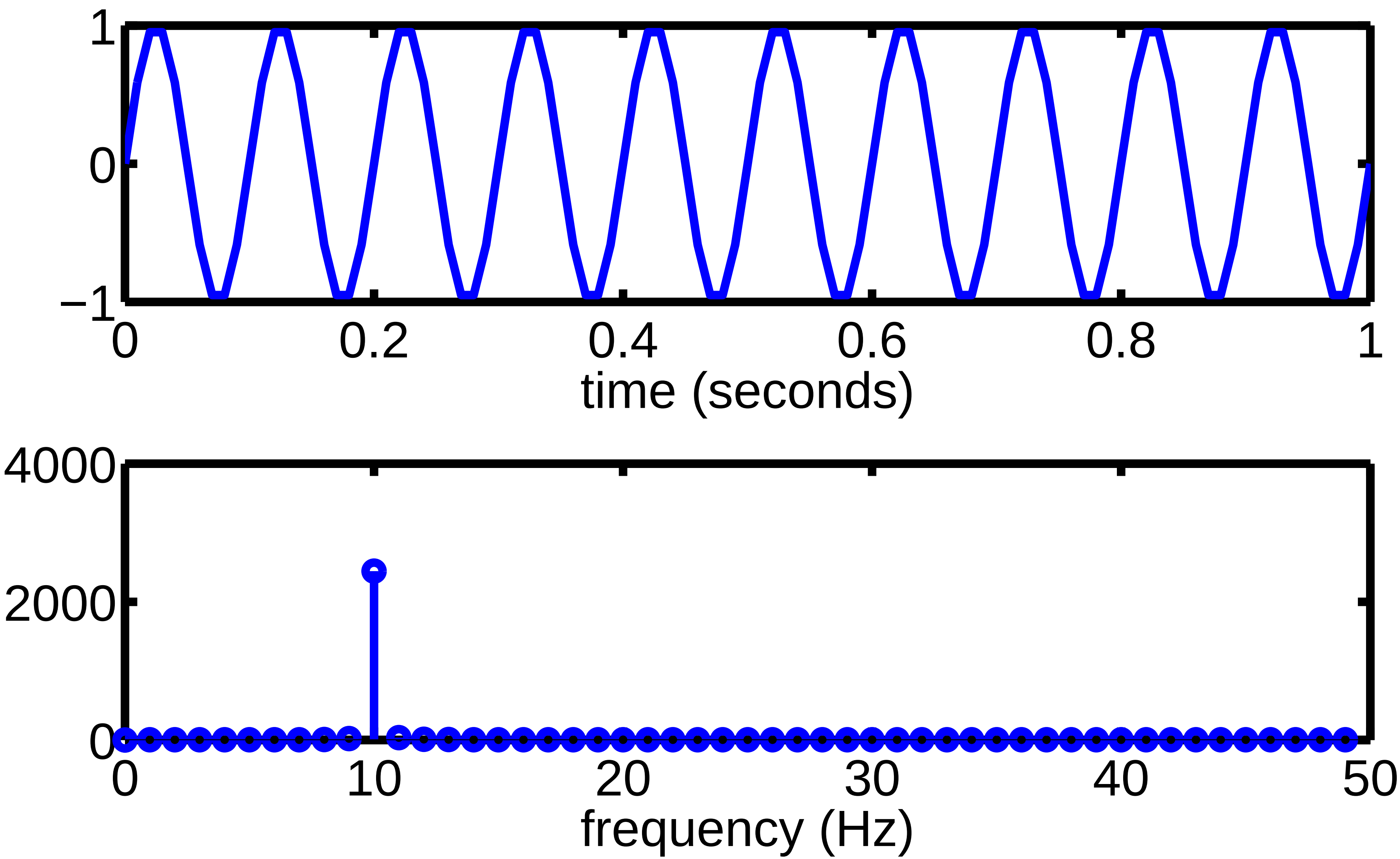
Generalization of L-S periodogram

The L-S periodogram is a special case of a more general set of results.

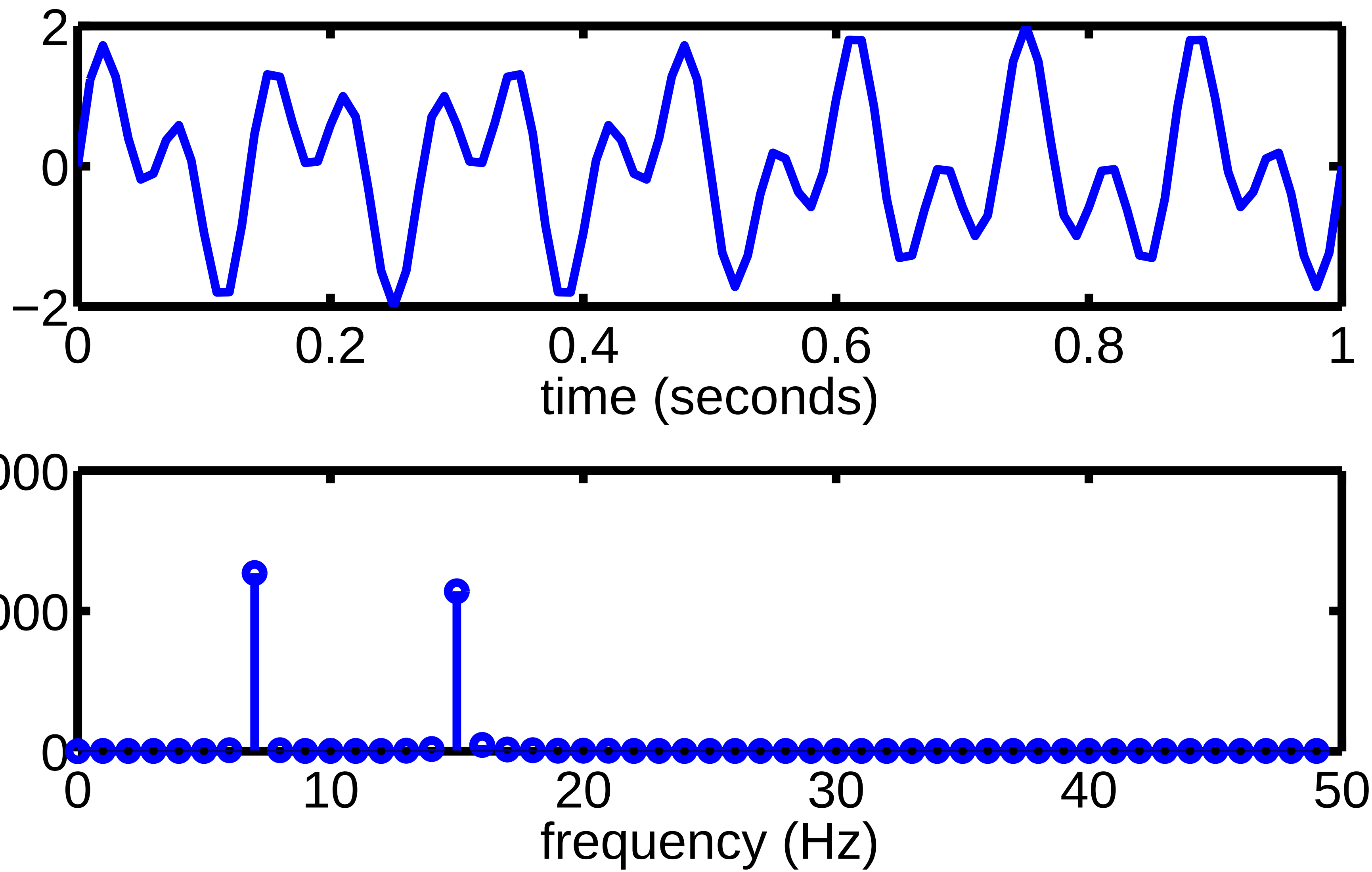
Sparse descriptions

- we should now be familiar with the idea of a basis
  - simple transforms change basis
  - mostly we consider orthogonal bases
  - non-redundant, i.e., efficient
    - but perhaps we get something if we allow redundancy
- Why transform: sparse description of data can be useful
  - this is one reason why the FT can be useful
  - transform into a basis where the description of the signal is sparse
  - if the description is sparse, then we can compress the signal

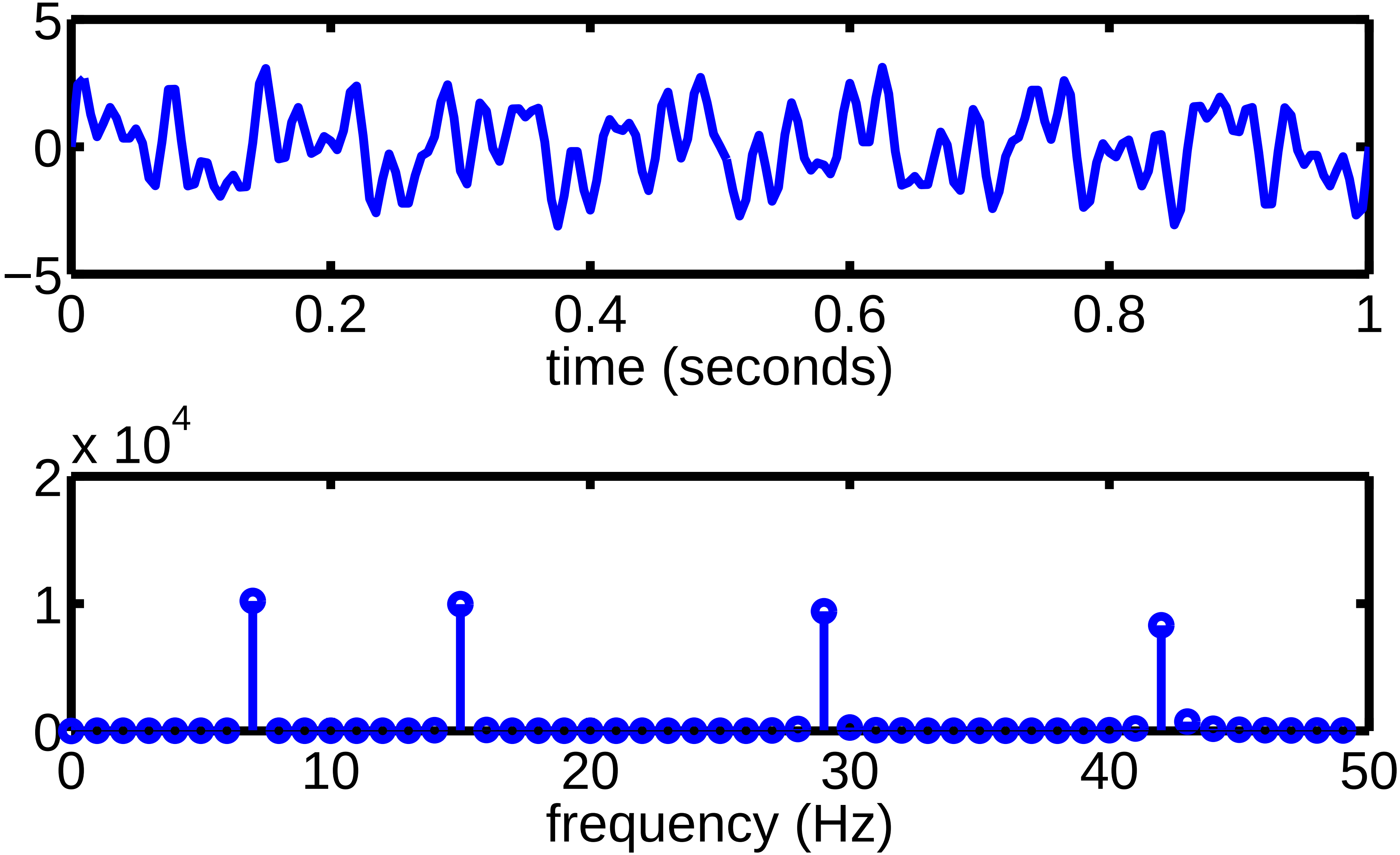
Sparse description example 1



Sparse description example 2

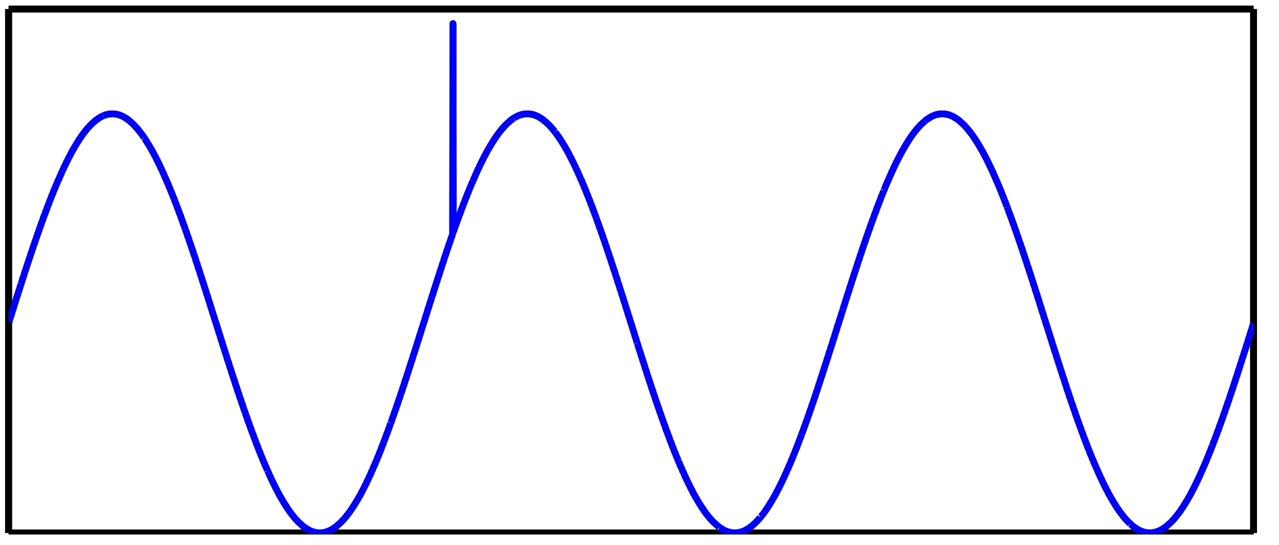


Sparse description example 3



Sparse description example 4

The following is a sine, plus a "spike"



- To represent this in either Fourier or "delta" basis requires all basis terms.
- but with both, we can represent it as

$$x(t) = \sin(t) + \delta(t - t_0)$$

Basis pursuit

- There is no standard orthonormal basis that allows us to represent a spike plus a sine wave.
- We are really picking and choosing the “best bits” of two different bases.
- Allows us to find a sparse description of our data
  - might allow analysis, compression, ...
- So we go in pursuit of a basis

Dictionary

- A dictionary allows us to describe words
- we want a dictionary for our signals
- we want a way to translate into the dictionary
- we want ways to provide translation between different languages

Lets stick to linear combinations, i.e. let us describe our signal by a linear combination

$$x = \sum_i \alpha_i \phi_i$$

for some set of **atoms** ϕ_i from our dictionary $\mathcal{D}$.

Sparse recovery

How can we obtain such a representation?

- we can no longer rely on a simple transform
- the Dictionary could be quite large
  - searches through it for a sparse representation would take too long
  - in fact, NP hard
  - corresponds to minimizing the l^0 norm
  - i.e., we try to solve the optimization problem

$$\text{minimize} \sum_{i:\alpha_i \neq 0} 1 \quad \text{such that } x = \sum_i \alpha_i \phi_i$$

Norms revisited

There are a group of norms on $\mathbb{R}^n$ called the l^p norms defined by

$$\|\mathbf{x}\|_p = \left[\sum_{i=1}^{n} |x_i|^p \right]^{1/p}$$

Simple examples are

- l^2: defined by $\|\mathbf{x}\|_2 = \left[\sum_{i=1}^n |x_i|^2\right]^{1/2}$
  - related to the RMS value
- l^1: defined by $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
  - related to the mean absolute value
- l^0: defined by $\|\mathbf{x}\|_0 = \sum_{i=1}^n I(x_i \neq 0) = \sum_{i:x_i \neq 0} 1$
  - just counts the number of non-zero terms of $\mathbf{x}$

Sparse recovery via l^1 norm

The problem above consists of

$$\text{minimize } \|\mathbf{x}\|_0 \quad \text{such that } x = \sum_i \alpha_i \phi_i$$

However, various papers have shown that for very many cases, one gets a good approximate solution to the above optimization problem by solving

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{such that } x = \sum_i \alpha_i \phi_i$$

Minimization of the l^1 norm

We can rewrite

$$\text{minimize } \|\mathbf{x}\|_1 \quad \text{such that } x(k) = \sum_i \alpha_i \phi_i(k)$$

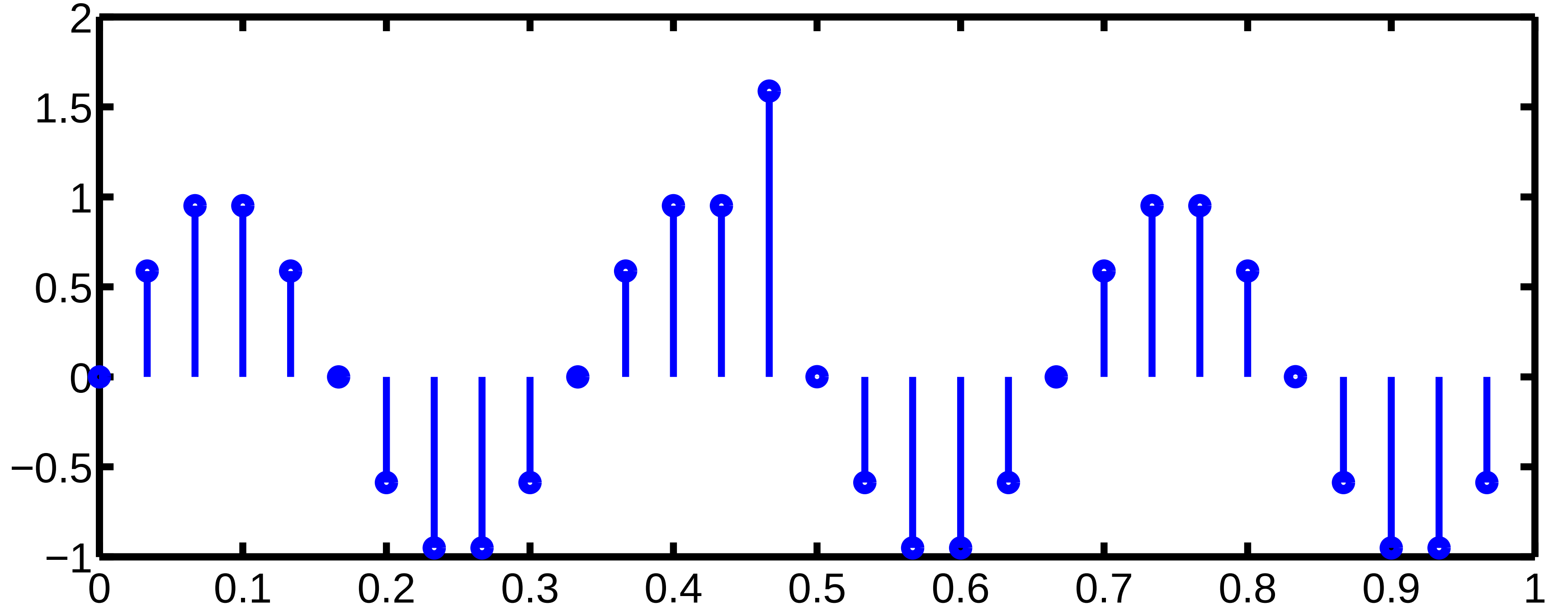
as

$$\begin{aligned} &\text{minimize} && \sum_i \varepsilon_i \\ &\text{such that} && \\ & && x(k) = \sum_i \alpha_i \phi_i(k) \\ & && -\varepsilon_i \leq \alpha_i \leq \varepsilon_i \end{aligned}$$

This is just a linear program, and can be solved by Simplex, or interior point methods for quite large problems.

Example

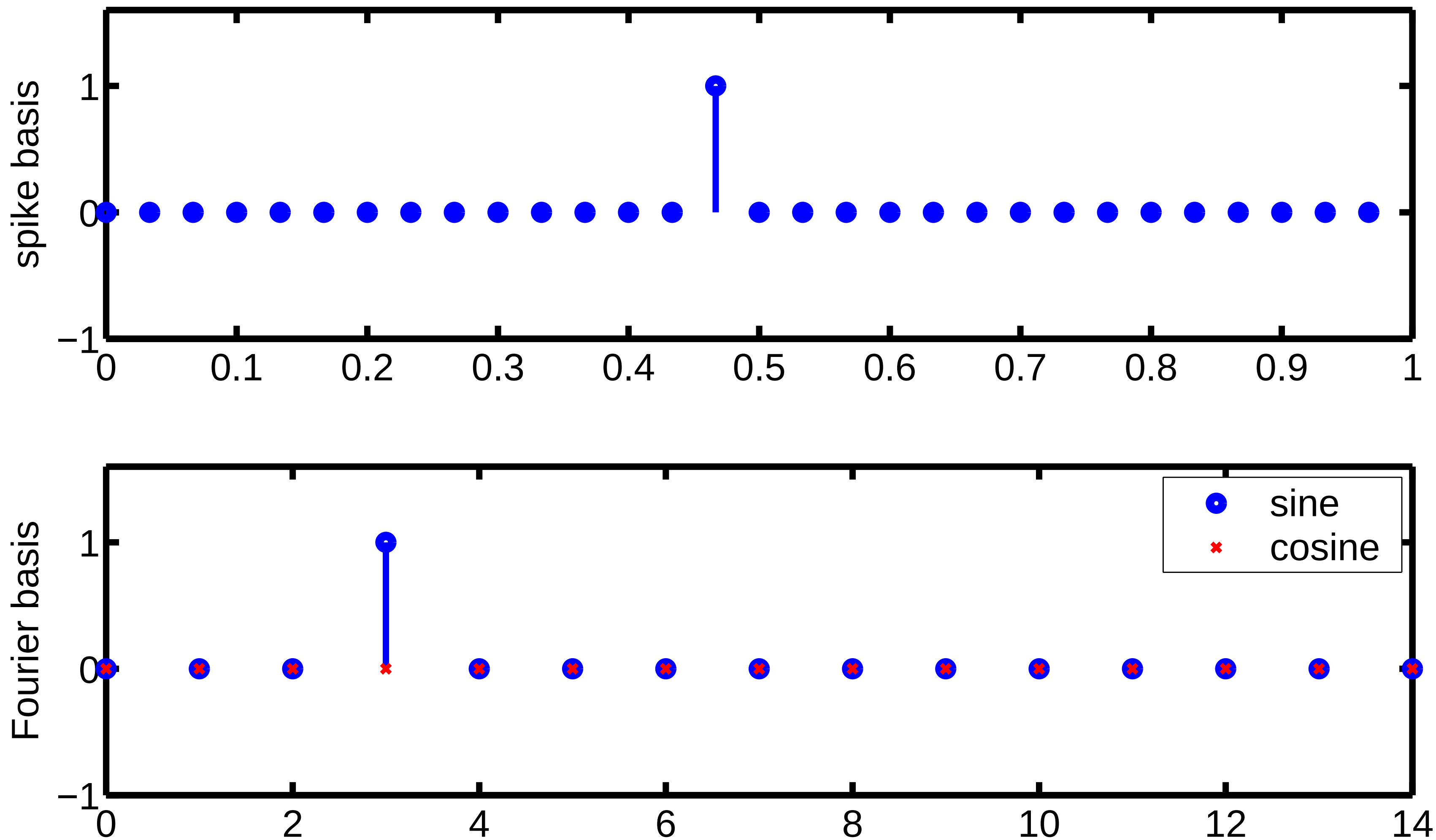
Try to represent the following signal using Fourier and spike basis



Perform the l^1 minimization

Example

Result of the l^1 minimization



Application

One possible application is anomaly detection in traffic data

- traffic data shows periodicities
  - daily (diurnal) cycles (people sleep)
  - weekly cycles (people take the weekend off)
- anomalies (e.g. problems like DoS attacks) often appear as spikes
- if we separate the two, we can find the problems.

Why does it work

Assume sparse representation exists

- then it exists in one of a set of subspaces that are parallel to axes of $\mathbb{R}^n$
- l^0 minimization has to search these
- l^2 looks for solution closest (using Euclidean distance) to a translated subspace (given by constraints).
- l^1 looks for solution closest (using checker distance) to a translated subspace (given by constraints).

Relation to L-S periodogram

- L-S periodogram is implicitly assuming that the signal representation is sparse in the Fourier basis
- do a "least-squares" fit
  - tests each basis function against the signal
- perhaps we can do better using l^1 norm Minimization?