The Fast Fourier Transform

Derek L. Smith
SIAM Seminar on Algorithms - Fall 2014
University of California, Santa Barbara

October 15, 2014

Table of Contents

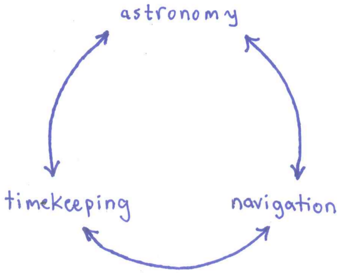
Table of Contents

Navigating the Origins of the FFT

The Royal Observatory, Greenwich, in London has a stainless steel strip on the ground marking the original location of the prime meridian. There's also a plaque stating that the GPS reference meridian is now 100m to the east.

This photo is the culmination of hundreds of years of mathematical tricks which answer the question: *How to construct a more accurate clock? Or map? Or star chart?*

Time, Location and the Stars



The answer involves a naturally occurring reference system. Throughout history, humans have measured their location on earth, in order to more accurately describe the position of astronomical bodies, in order to build better time-keeping devices, to more successfully navigate the earth, to more accurately record the stars... and so on... and so on...

Time, Location and the Stars

Transoceanic exploration previously required a vessel stocked with maps, star charts and a highly accurate clock. Institutions such as the Royal Observatory primarily existed to improve a nations' navigation capabilities. The current state-of-the-art includes atomic clocks, GPS and computerized maps, as well as a whole constellation of government organizations.

This cycle of continual improvement has fueled many mathematical discoveries and it is within this story that we find the origins of the Fast Fourier Transform. In particular, the FFT grew out of mathematical techniques for determining the orbit of planetary bodies.

Orbit as a Periodic Curve

The orbit of a body like a planet can be modeled by a curve

$$\mathbf{c} : [0, T] \to \mathbb{R}^3,$$

For a given time $0 \leq t \leq T$, the function produces a location

$$\mathbf{c}(t) = (x(t), y(t), z(t))$$

in some unspecified coordinate system. We assume the motion is *periodic* with period T to extend the function to any time:

$$\mathbf{c}(t + T) = \mathbf{c}(t) \qquad (t \in \mathbb{R}).$$

Example. The Earth's orbit is approximately circular (eccentricity 0.01671123) with period 365.256 days.

Orbit Determination

Imagine an Earth-bound observatory as it tracks a planet. The astronomers record a sequence of observations derived from the (unknown) orbital curve $\mathbf{c} : [0, T] \to \mathbb{R}^3$, with the intention of predicting the future location of the body. The problem of recovering the function $\mathbf{c}$ from observations is called orbit determination.

Because the underlying function $\mathbf{c}$ is periodic, the derived observations are also periodic. Due to work in orbital mechanics, numerous techniques were developed to approximate periodic functions using linear combinations of trigonometric functions.

Spotting a Missing 'Planet'

The planets Mecury, Venus, Mars, Jupiter and Saturn were known to the Babylonians. The 18th century brought the discovery of Uranus, whose nearly circular orbit was relatively easy to determine. The location of Uranus fit the Titus-Bode law, which also predicted a planet between Mars and Jupiter.

At the beginning of the 19th century, Giuseppe Piazza observed an object in this predicted location. It was smaller than any object known to date. At 950km in diameter, Ceres is the largest object in the asteroid belt between Mars and Jupiter. Discovered on 1 January 1801, Piazzi made 21 observations over 42 days before losing the object in the sun. A race ensued to recover the orbit of Ceres and capture new observations of the presumed planet.

Enter Gauss

After working in number theory and completing a dissertation on the fundamental theorem of algebra, the twenty four year old Carl Friedrich Gauss turned to the problem of determining the orbit of Ceres from Piazzi's limited number of observations.

Gauss developed a technique based on Kepler's laws of motion and the theory of conic sections. An overview of the method and further resources can be found in this presentation. Based on his predictions, Ceres was again observed in December of 1801.

Here are a few amazing facts about Gauss' method.

1. Requires only three Earth-based measurements.
2. One of the first demonstrations of the method of least squares.
3. Developed the FFT to quickly interpolate periodic functions.

Table of Contents

The Discrete Fourier Transform

Problem Given N equally spaced measurements of a periodic function, write down an interpolating trigonometric polynomial.

The answer to this problem is prescribed by the discrete Fourier transform (DFT). Given the sequence $x_0, x_1, \ldots, x_{N-1}$ of N complex-valued measurements, define

$$\hat{x}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \exp(-2\pi i k n / N) \qquad (k = 0, 1, \ldots, N-1).$$

This expression defines a linear transformation $\mathscr{F} : \mathbb{C}^N \to \mathbb{C}^N$ called the discrete Fourier tranform.

Interpolating Polynomial

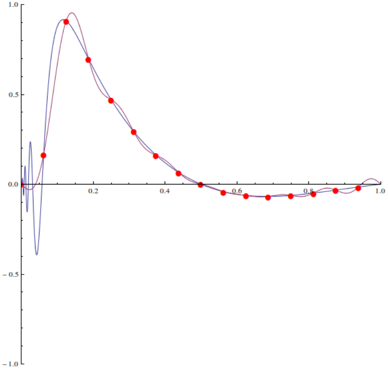


Figure: Original function in blue, 16 samples and interpolant in red.

Given the DFT of N equally spaced samples of a periodic function

$$g : [0, 1] \to \mathbb{C},$$

we can write an interpolating trigonometric polynomial as

$$\tilde{g}(t) = \sum_{|n| \leq N/2} \hat{x}_n \exp(2\pi i n t)$$

using that the DFT is periodic.

DFT as Coordinate Transformation

Let $N = 8$ and $\omega_8 = \exp(-2\pi i/8)$. Consider the vectors

$$
\begin{aligned}
\mathbf{u}_0 &= \frac{1}{\sqrt{8}}\langle 1, 1, \ldots, 1\rangle \\
\mathbf{u}_1 &= \frac{1}{\sqrt{8}}\langle 1, \omega_8, \omega_8^2, \ldots, \omega_8^7\rangle \\
&\vdots \\
\mathbf{u}_k &= \frac{1}{\sqrt{8}}\langle 1, \omega_8^k, \omega_8^{2k}, \ldots, \omega_8^{7k}\rangle \\
&\vdots \\
\mathbf{u}_7 &= \frac{1}{\sqrt{8}}\langle 1, \omega_8^7, \omega_8^{14}, \ldots, \omega_8^{49}\rangle
\end{aligned}
$$

which are sampled versions of the complex exponentials. These eight vectors form an orthonormal basis for $\mathbb{C}^8$.

Orthogonality

Given the definition

$$\mathbf{u}_k = \frac{1}{\sqrt{8}} \langle 1, \omega_8^k, \omega_8^{2k}, \ldots, \omega_8^{7k} \rangle$$

consider the inner product

$$\begin{aligned} \mathbf{u}_k^T \mathbf{u}_{k'}^* &= \frac{1}{8} \sum_{n=0}^{7} \exp(-2\pi i k n / N) \exp(2\pi i k' n / 8) \\ &= \frac{1}{8} \sum_{n=0}^{7} \omega_8^{(k-k')n} = \delta_{kk'}. \end{aligned}$$

This follows from the sum of the geometric progression, or by the visual proof to the right.

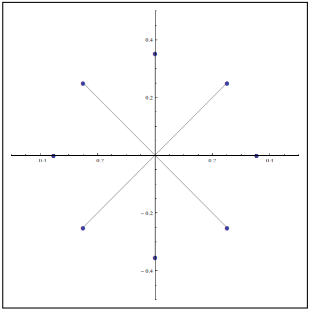


Figure: A visualization of the roots of unity W_8^{3n}, corresponding to $k = 6$ and $k' = 3$.

Basis as Sampled Complex Exponential

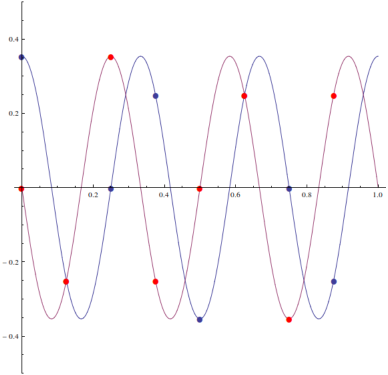


Figure: This graph depicts $\mathbf{u}_3$ overlayed on f_3; the real part in blue and the imaginary part in red.

We have sampled the function

$$f_3(t) = \frac{1}{\sqrt{8}} \exp(-2\pi i(3t))$$

uniformly at values

$$t \in \left\{0, \frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}\right\}$$

resulting in

$$\mathbf{u}_3 = \frac{1}{\sqrt{8}} \langle 1, \omega_8^3, \omega_8^6, \ldots, \omega_8^{21} \rangle.$$

DFT in Matrix Form

Focusing again on $N = 8$, we can write the DFT in matrix form

$$\begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \\ \hat{x}_5 \\ \hat{x}_6 \\ \hat{x}_7 \end{bmatrix} = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega_8 & \omega_8^2 & \omega_8^3 & \omega_8^4 & \omega_8^5 & \omega_8^6 & \omega_8^7 \\ 1 & \omega_8^2 & \omega_8^4 & \omega_8^6 & \omega_8^8 & \omega_8^{10} & \omega_8^{12} & \omega_8^{14} \\ 1 & \omega_8^3 & \omega_8^6 & \omega_8^9 & \omega_8^{12} & \omega_8^{15} & \omega_8^{18} & \omega_8^{21} \\ 1 & \omega_8^4 & \omega_8^8 & \omega_8^{12} & \omega_8^{16} & \omega_8^{20} & \omega_8^{24} & \omega_8^{28} \\ 1 & \omega_8^5 & \omega_8^{10} & \omega_8^{15} & \omega_8^{20} & \omega_8^{25} & \omega_8^{30} & \omega_8^{35} \\ 1 & \omega_8^6 & \omega_8^{12} & \omega_8^{18} & \omega_8^{24} & \omega_8^{30} & \omega_8^{36} & \omega_8^{42} \\ 1 & \omega_8^7 & \omega_8^{14} & \omega_8^{21} & \omega_8^{28} & \omega_8^{35} & \omega_8^{42} & \omega_8^{49} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix}.$$

Or, more compactly, $\hat{\mathbf{x}} = W\mathbf{x}$, where W is the Vandermode matrix.
Observe that the k-th row and column is simply the vector $\mathbf{u}_k$.
Hence the matrix W is unitary, $W^*W = I$.

This formulation yields a cost of $O(N^2)$ multiplication operations.

Table of Contents

The Discrete Fourier Transform

Recall that given a vector $\mathbf{x} = \langle x_0, x_1, \ldots, x_{N-1} \rangle \in \mathbb{C}^N$ its discrete Fourier transform is another vector $\hat{\mathbf{x}} \in \mathbb{C}^N$ given by

$$\hat{x}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \omega_N^{nk} \qquad (k = 0, 1, \ldots, N-1)$$

where $\omega_N = \exp(-2\pi i/N)$.

Problem. How to compute the DFT more efficiently?

The Trick. Focus on the case $N = 2m$. Then we can partition $\mathbf{x}$ into even- and odd-indexed components and write the DFT as two DFTs of length m.

The Fast Fourier Transform

Details for $N = 2m$:

$$\begin{aligned}
\hat{x}_k &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \omega_N^{nk} \\
&= \frac{1}{\sqrt{N}} \sum_{n=0}^{m-1} x_{2n} \omega_N^{2nk} + \frac{1}{\sqrt{N}} \sum_{n=0}^{m-1} x_{2n+1} \omega_N^{(2n+1)k} \\
&= \frac{1}{\sqrt{N}} \sum_{n=0}^{m-1} x_{2n} \omega_m^{nk} + \frac{\omega_N}{\sqrt{N}} \sum_{n=0}^{m-1} x_{2n+1} \omega_m^{nk}
\end{aligned}$$

When $N = 2^m$, a divide and conquer scheme leads to the conceptually simple, recursive radix-2 formulation.

WARNING!

DO NOT IMPLEMENT YOURSELF!

While it seems easy, refrain from reinventing the wheel (except for educational purposes). Even the creators of Matlab have not attempted to implement the FFT. They chose to use the Fastest Fourier Transform in the West (FFTW).

You may also investigate hardware vendor-provided libraries.

Computational Cost of FFT

Let's count the number of multiplications for the case $N = 2^4$.

16-point DFT	$16^2 = 256$ multiplies
Two 8-point DFTs	$2 * (8^2) = 128$ multiplies
Four 4-point DFTs	$2 * 2 * (4^2) = 64$ multiplies
Eight 2-point DFTs	$2 * 2 * 2 * (2^2) = 64$ multiplies

The recursive formulation of the radix-2 algorithm has a complexity of $O(N \log_2 N)$ multiplication operations. In fact, this is the complexity for a general value of N.

The efficiency of the FFT can also be viewed as a special factorization of the Vandermonde matrix.

Variations of the FFT

There are numerous variations of the FFT algorithm.

1. Cooley-Tukey (arbitrary length)
2. Rader (prime length)
3. Bluestein (arbitrary length)

The recursive nature of the FFT depends on the factorization of the length

$$N = N_1 N_2 \dots N_m.$$

FFTW is fast partly because it cleverly combines the above algorithms based on N and the machine architecture.

Table of Contents

Recording Sound

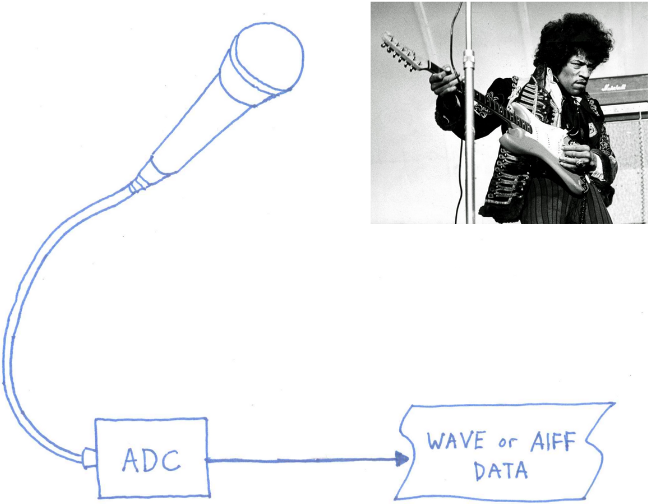


Figure: Hendrix image source.

Recording Sound

A description of the previous schematic.

1. Sound is the propagation of mechnical waves through the air.
2. A microphone converts pressure variations into voltage fluctuations.
3. An analog-to-digital converter creates a sequence of samples from electrical signal. For CD-quality sound, 44100 samples occur each second and each is measured with 16-bit accuracy.
4. The samples are commonly encoded in a PCM format like WAV or AIFF.

We model a (single-channel) sound file as an element of $\mathbb{R}^N$ where each component has value in the range $[-1, 1]$. Note that N is the product of the length of the recording and the sample rate.

Compression of Data and Sound Perception

The MP3 format is a lossy compression scheme; information is discarded to acheive smaller file size. But how to decide what data to through away? In the time-domain, it's not clear how to proceed without significantly affecting sound quality.

The key insight for the MP3 format relies on perception of sound in the human brain. The first step is to understand that the brain essentially performs a Fourier transform on sound. If one played a pure sinusoidal sound through a speaker, a listener would recognize the sound as consisting of a single frequency.

Frequency Masking

This signal exhibits two clusters of energy, close together in frequency but one of much greater amplitude. Due to a phenomenon called auditory masking, the average human would have difficulty perceiving the frequencies highlighted in red.

Using spectrum analysis and limitations of human hearing, we can decide what data in frequency space to discard!

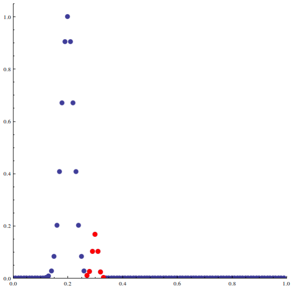


Figure: Magnitude of DFT of signal of length $N = 100$.

The Basic MP3 Algorithm

Input: sampled sound as a vector in $\mathbb{R}^N$.

1. Subdivide sound file into small snippets called 'frames'.
2. For each frame:
   2.1 Compute DFT vector.
   2.2 Apply frequency masking to approximate spectrum.
   2.3 Losslessly compress (zip) modified DFT vector.

To decode, unzip and perform an inverse DFT on each frame.
Both encoding and decoding require the FFT.

The JPEG algorithm is structurally similar, using knowledge of human vision to decide which Fourier coefficients to discard.

Table of Contents

The Fourier Transform

The DFT is one of many version of a very general tool called the Fourier transform, which is used in many branches of modern mathematics. We considered the DFT as a linear operator on the vector space $\mathbb{C}^N$, but we could also consider the vector $\mathbf{x}$ as a discrete complex-valued function

$$\mathbf{x} : \mathbb{Z}_N \to \mathbb{C}.$$

That is, $\mathbf{x}$ is a map from the cyclic group $\mathbb{Z}_N$ into the complex numbers $\mathbb{C}$. Here $\mathbb{Z}_N$ is the set $\{0, 1, 2, \ldots, N-1\}$ with modular addition as the operation. We were able to decompose the function $\mathbf{x}$ into a linear combination of special functions

$$u_k : \mathbb{Z}_N \to \mathbb{C} \qquad (k = 0, 1, 2, \ldots, N-1)$$

called characters.

Harmonic Analysis

In general, we may begin with any locally compact abelian group G and consider complex-valued functions

$$f : G \to \mathbb{C}.$$

The field of harmonic analysis studies the characters of G and attempts to decompose functions like f in terms of the characters (which also forms a locally compact abelian group).

The most familiar choices of G are

1. $\mathbb{Z}_N$
2. $\mathbb{Z}$
3. $\mathbb{T} = [-\pi, \pi]$
4. $\mathbb{R}$.

Theory of Fourier Series

For a periodic function

$$f : \mathbb{T} \to \mathbb{C}$$

the Fourier transform is given by

$$c(n) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) \exp(-inx)\, dx \qquad (n \in \mathbb{Z}).$$

This expression defines a function $c : \mathbb{Z} \to \mathbb{C}$, so we say that $\mathbb{T}$ is the dual group to $\mathbb{Z}$. We can invert the transform by the formula

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} c(n) \exp(inx).$$

The Fourier transform on $\mathbb{T}$ and $\mathbb{R}$ is an essential tool in the theory of partial differential equations, as discovered by Joseph Fourier in his work on the heat equation.

Linear Partial Differential Equations

An evolution equation is a partial differential equation which describes the time evolution of a physical system starting from a given initial configuration.

Example. The solution to the initial value problem

$$\begin{cases} \partial_t v + \partial_x^3 v = 0 \\ v(x,0) = v_0(x) \end{cases}$$

with $x, t \in \mathbb{R}$ is provided by Fourier analysis

$$v(x,t) = [e^{-it\xi^3}\hat{v}_0]^\vee(x) = V(t)v_0.$$

That is, we take Fourier transform of initial data, multiply the spectrum by $e^{-it\xi^3}$ and then take inverse transform. It can then be proved using this formula that many properties of the solution are inherited from the initial data.

Nonlinear Partial Differential Equations

What about the nonlinear KdV equation?

$$\begin{cases} \partial_t u + \partial_x^3 u + u\partial_x u = 0 \\ u(x,0) = u_0(x) \end{cases}$$

Using Duhamel's principle for inhomogeneous differential equations, we can represent the solution (implicitly) as a combination of a linear and inhomogeneous term

$$u(x,t) = V(t)u_0 + \int_0^t V(t-t')(u\partial_x u)(x,t')\,dt'.$$

Using the linear theory and the contraction mapping principle, one can prove existence of solutions to this integral equation. This representation also allows one to show that many properties of the solution are inherited from the initial data.

Fourier Transform in Applied Mathematics

Signal Processing

Much of the history of the FFT is tied to the field of signal processing, from Gauss' work to that of Cooley and Tukey. The later were concerned with analyzing seismographic data to detect Soviet nuclear tests. The DFT and related Z-transform are essential in designing digital filters. Computer scientists study the discrete Fourier analysis of boolean functions.

Analysis

In functional analysis, the Gelfand representation is a generalization of the Fourier transform. The Hilbert transform provides a connection between complex and Fourier analysis. The Kakeya problem has connections to harmonic analysis and geometric measure theory.

Fourier Transform in Pure Mathematics

Algebra and Topology
The Fourier transform may be viewed as a specific case of representation theory. The theory can be easily extended to arbitrary finite groups and, with more work, to compact topological groups via the Peter-Weyl Theorem.

Number Theory
Fourier analysis can be used to understand the Hardy-Littlewood Circle Method, which has been used to attack Goldbach's conjecture. Shor's algorithm is a technique for factoring integers using the quantum Fourier tranform. The discrete Fourier transform may also be generalized for functions taking values in arbitrary fields. This so-called "number theoretic transform" finds application in efficiently multiplying large integers using a version of the FFT.

[1] James W. Cooley and John W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, Math. Comp. **19** (1965), 297–301. MR0178586 (31 #2843)

[2] Michael T. Heideman, Don H. Johnson, and C. Sidney Burrus, *Gauss and the history of the fast Fourier transform*, Arch. Hist. Exact Sci. **34** (1985), no. 3, 265–277, DOI 10.1007/BF00348431. MR815154 (87f:01018)