A practical introduction to Fourier Analysis of speech signals

Part I: Background

Let’s assume we have used supersine.m to synthesize with just 3 sine components a rough approximation to a square wave, using the following specification:

Frequency = [100 300 500]
Amplitude = [1 1/3 1/5]
Phase = [0 0 0]

If we don’t already know,
how can we find what components are in the signal?

Basic idea:

Compare the given signal with the set of possible component signals.

What’s a plausible way of doing this comparison?

For each possible component, multiply the signal sample by sample with the component.
Form the cumulative sum of these products.

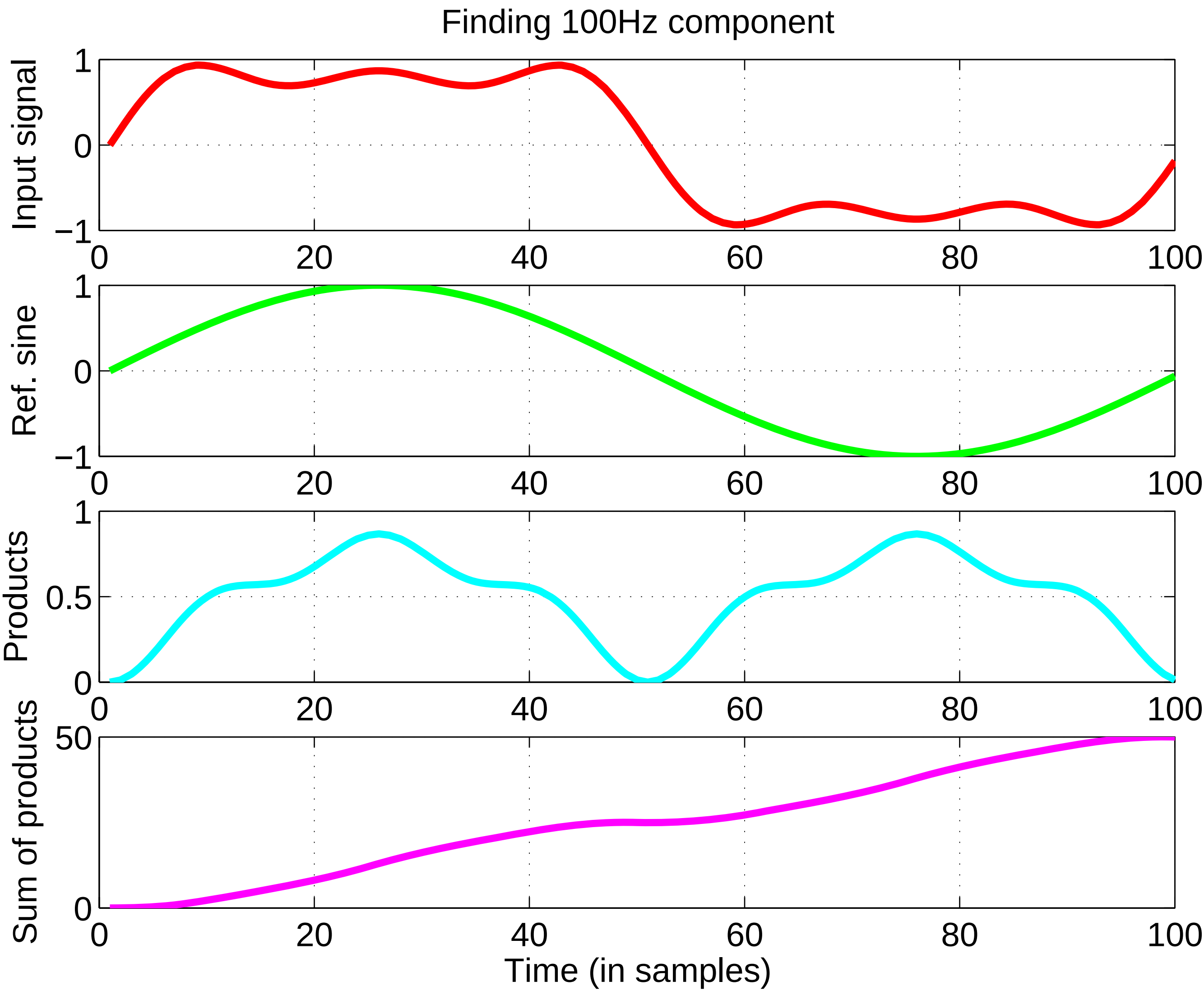
This is quite similar to calculating a correlation coefficient

For the moment we will assume that we already know that the signal is periodic at 100Hz.

So we will only look at a 10ms segment of the signal just synthesized

Sine waves that fit precisely into a 10ms window are, of course, 100Hz, 200Hz, 300Hz, 400Hz etc.

First of all, we compare the square−wave approximation with a 100Hz sine


Finding 100Hz component
Input signal
Ref. sine
Products
Sum of products
1
0
−1
0.5
50
0
20
40
60
80
100
Time (in samples)

Maybe no surprise, but when we multiply
the input signal (top panel) sample by sample
with a 100Hz sine (2nd panel)
then all the products (3rd panel) are positive,
so the cumulative sum of the products (4th panel)
increases continuously from left to right.

The value we are now interested in is the final (right−most) value in the cumulative sum (bottom panel):

Cumulative sum = 50

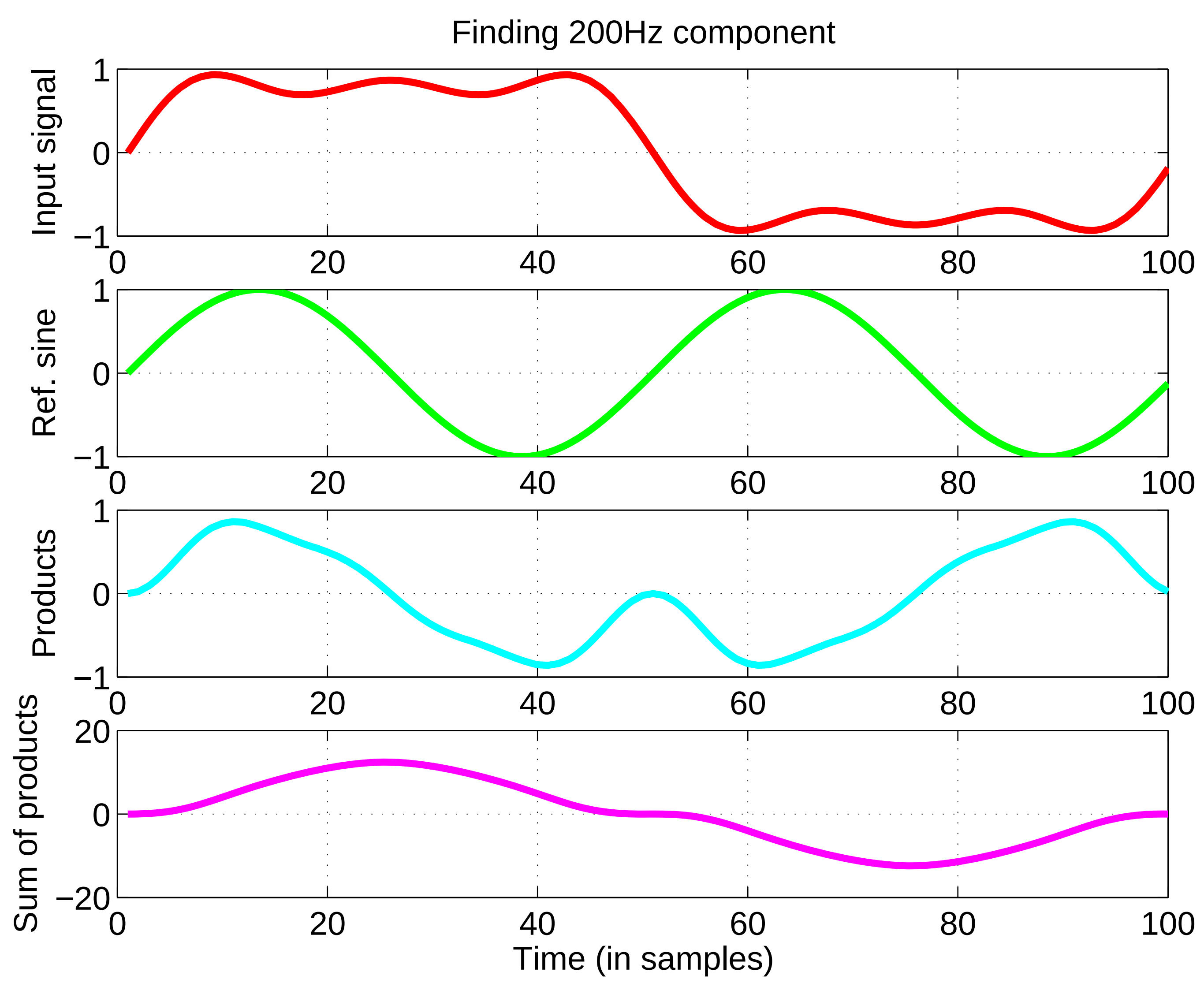
How can we interpret this value of 50?

First lets divide it by the number of samples over which the cumulative sum was derived (the number of samples is 100 because the synthesis by supersine.m used a samplerate of 10000Hz):

(Cumulative Sum)/(Number of Samples) = 0.5

When synthesizing the signal we used an amplitude value of 1 for the 100Hz component.
This refers to the **peak** amplitude.
As the amplitudes of the sine waves used for comparison are also 1, then what we have in effect calculated here is the **mean square amplitude** of a sine with a peak amplitude of 1.
We see here that this mean square amplitude is 0.5.
(**Root Mean Square** would thus be 0.7071.)
It is easy to show that in general the result of our multiplication and summation procedure must be **half the peak amplitude** of the component in the input signal

To be sure, we check that the amplitudes of the other components come out as expected.
So the next step is to compare the signal with a 200Hz sine

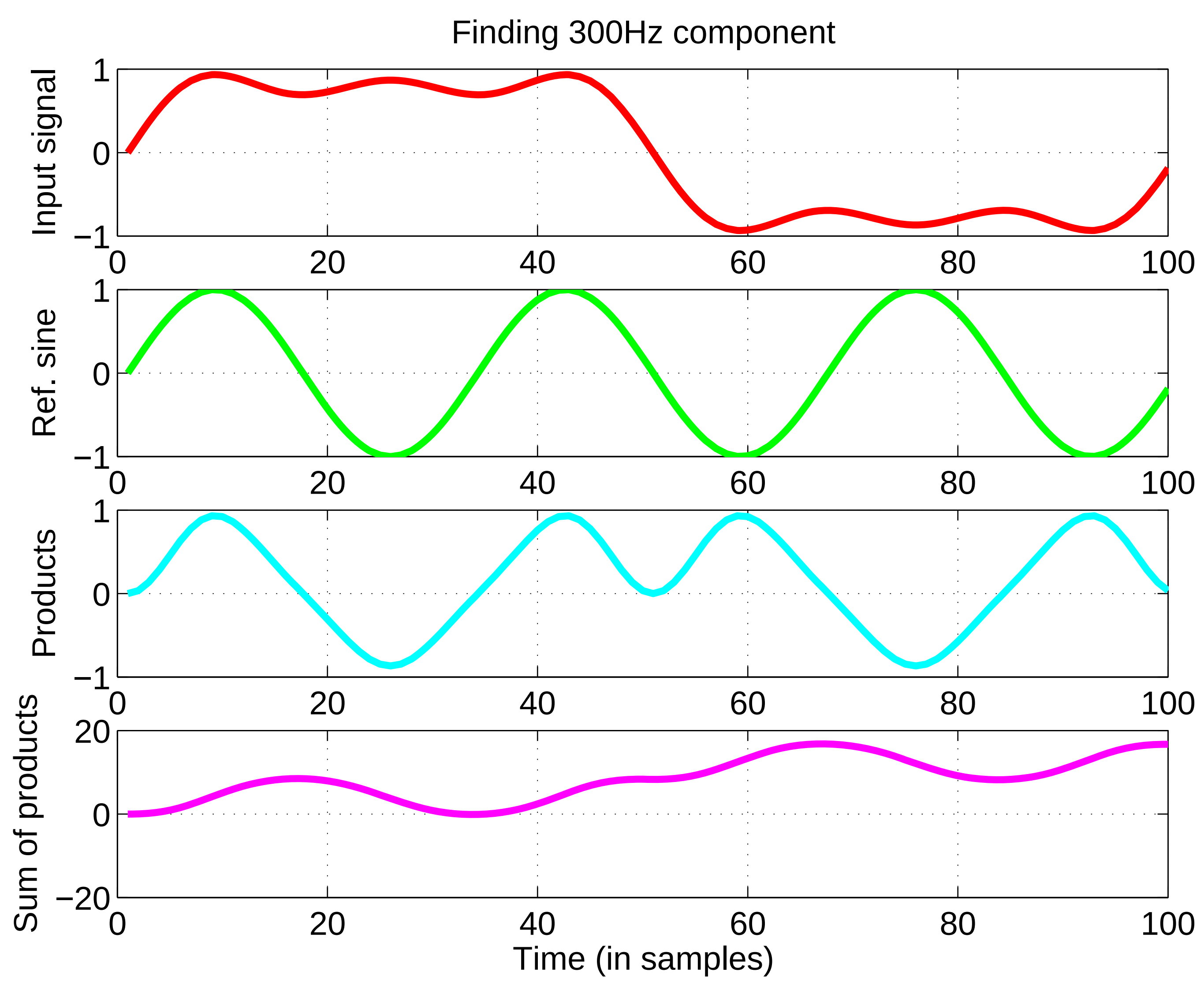

Finding 200Hz component
Input signal
Ref. sine
Products
Sum of products
1
0
−1
20
0
−20
0
20
40
60
80
100
Time (in samples)

This is very close to zero.
(As expected, since even−numbered harmonics are not present in a square−wave signal)

Amplitude of 200Hz component = 4.6942e−017

Now for the third component (300Hz).
This was synthesized with peak amplitude of 1/3.

Thus, expected value (0.5 * 1/3) =   0.16667



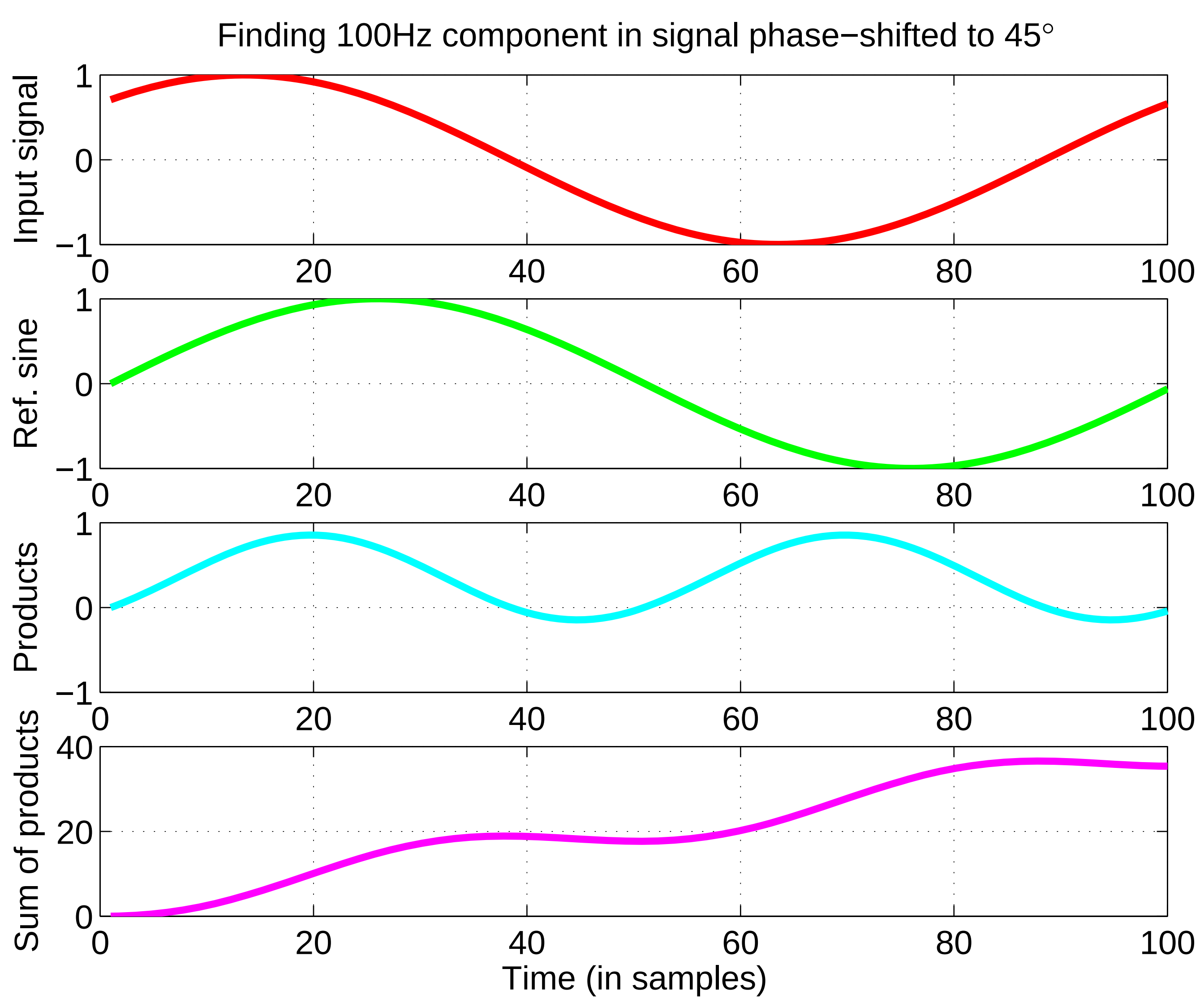
Finding 300Hz component
Input signal
Ref. sine
Products
Sum of products
Time (in samples)
1
0
−1
20
−20
0
20
40
60
80
100

Amplitude of 300Hz component = 0.16667

So far, so good.
We appear to have a way of recovering the amplitude of the individual components of a signal.

But in this signal the PHASE of all components was zero.
What happens when this is not the case?

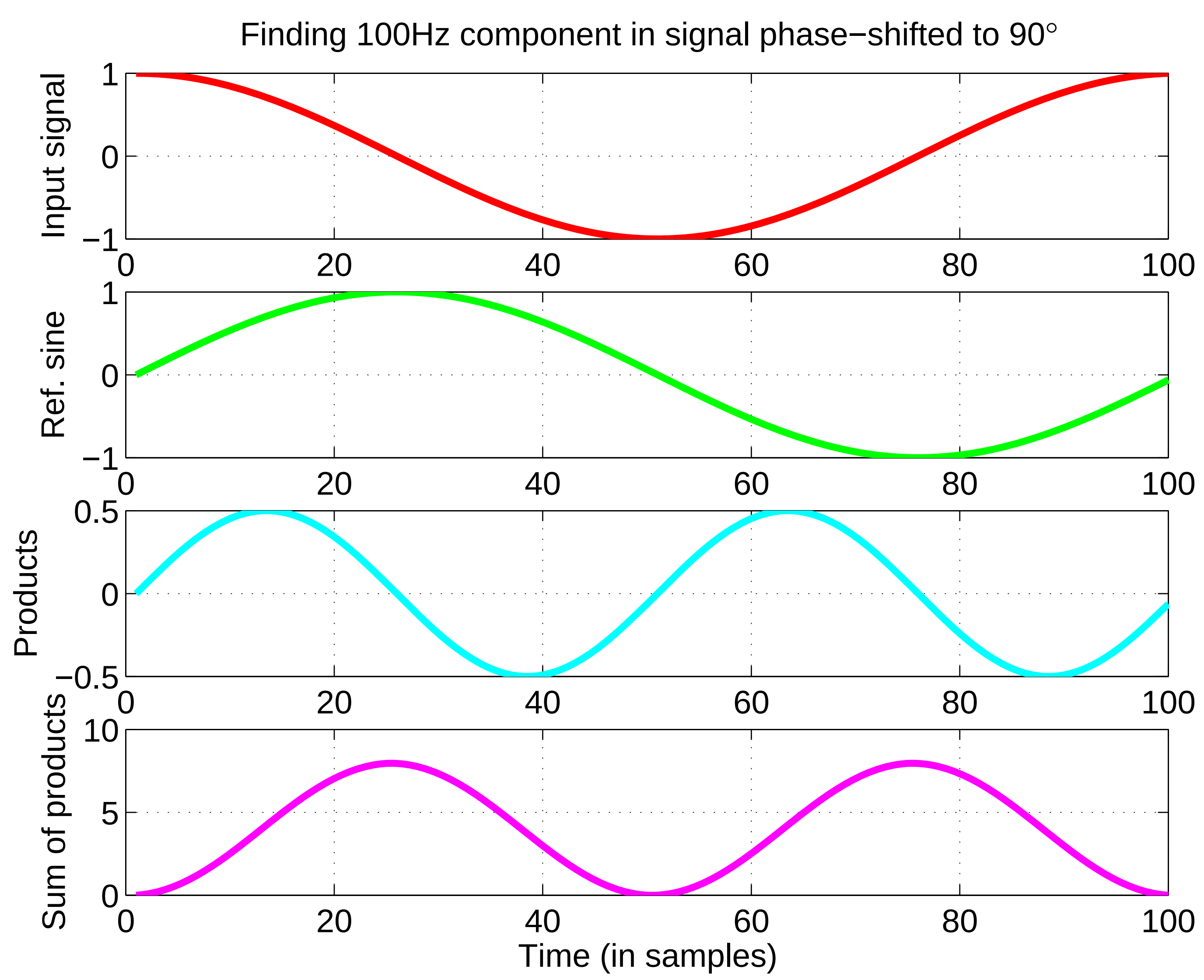
We will examine this by using a signal consisting of a single component (100Hz), but phase shifted.



Finding 100Hz component in signal phase−shifted to 45°
Input signal
Ref. sine
Products
Sum of products
Time (in samples)
1
0
−1
40
20
0
0
20
40
60
80
100

100Hz component, phase shifted to 45°.
Amplitude = 0.35355
This value is clearly below the figure of 0.5 obtained with no phase shift

It’s probably not too difficult to guess what phase value in the input signal is going to give an amplitude of zero



Finding 100Hz component in signal phase−shifted to 90°
Input signal
Ref. sine
Products
Sum of products
1
0
−1
0.5
−0.5
10
5
0
20
40
60
80
100
Time (in samples)

100Hz component, phase shifted to 90°.
Amplitude = 1.4849e−017
A very small value!

This should give us a clue as to how to handle signals regardless of their phase:

Perform the multiplication−summation with TWO comparison signals and combine the result.

Using both a zero−phase sine and a 90°−phase sine would give us correct results for the extreme cases of a zero−phase and 90°−phase input signal, and we can hope that they can be combined to give sensible results for phase angles in between, e.g 45°

What’s another name for a sine shifted 90°?

COSINE

Using comparison signals 90° apart is intuitively appealing:

It's like setting up a two−dimensional coordinate system
(X and Y axes are at 90° to each other)

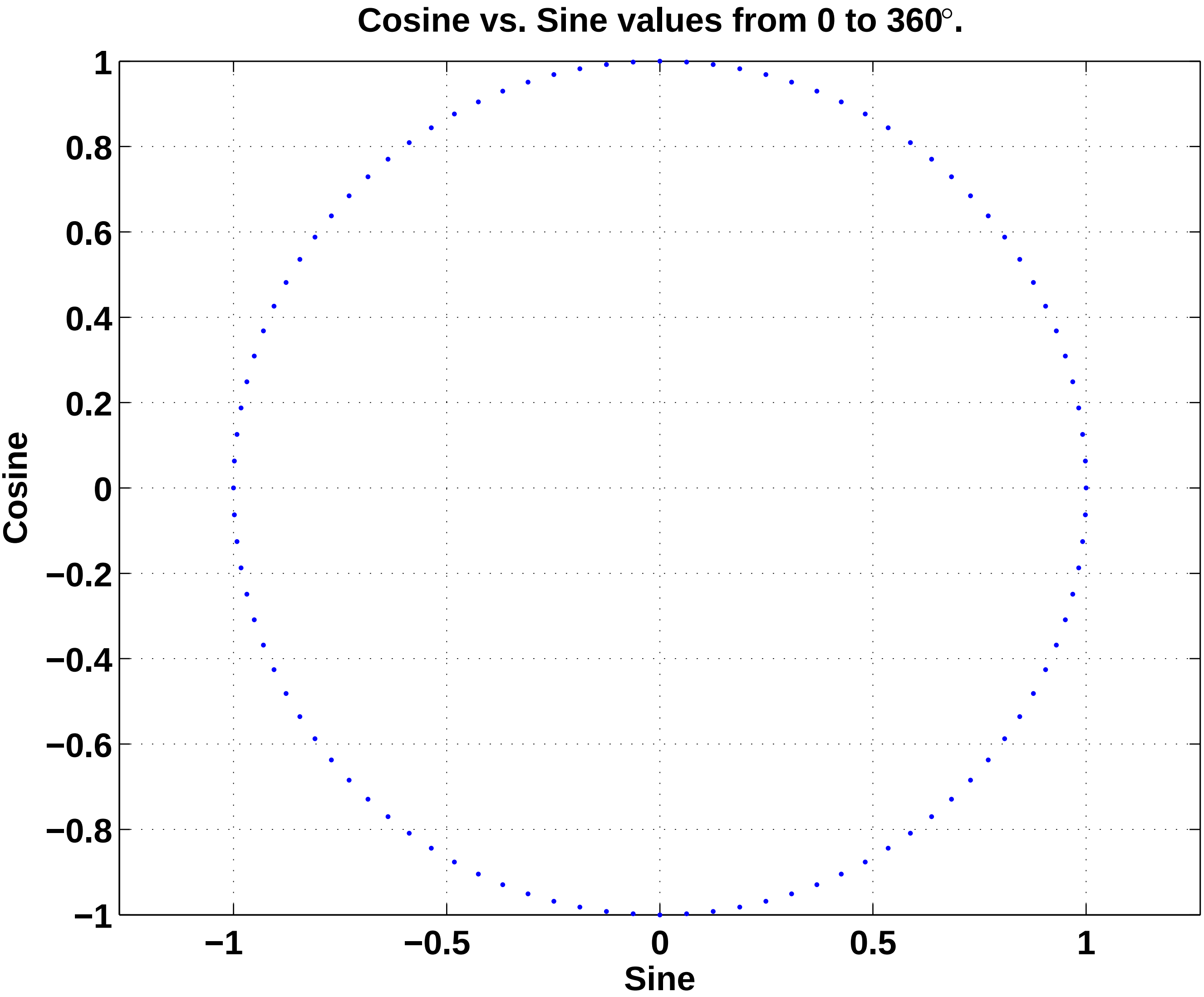
Up to now we have defined our signals relative to a sine with a phase of 0°, and are now thinking of a cosine as a sine rotated 90°.

So we can imagine plotting our multiplication−summation results such that the sine−based result gives the x value and the cosine result the y value

(We will see shortly that − for a good reason − standard Fourier analysis procedures swap this axis arrangement.)

Just to reinforce the point that sine and cosine together form the appropriate bases for analyzing a signal:

What does one get when one plots one cycle of (say) a 100Hz cosine against a 100Hz sine?


Cosine vs. Sine values from 0 to 360°.
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
Cosine
−1
−0.5
0
0.5
1
Sine

A circle, of course.

This indicates that the sine and cosine component are independent of each other

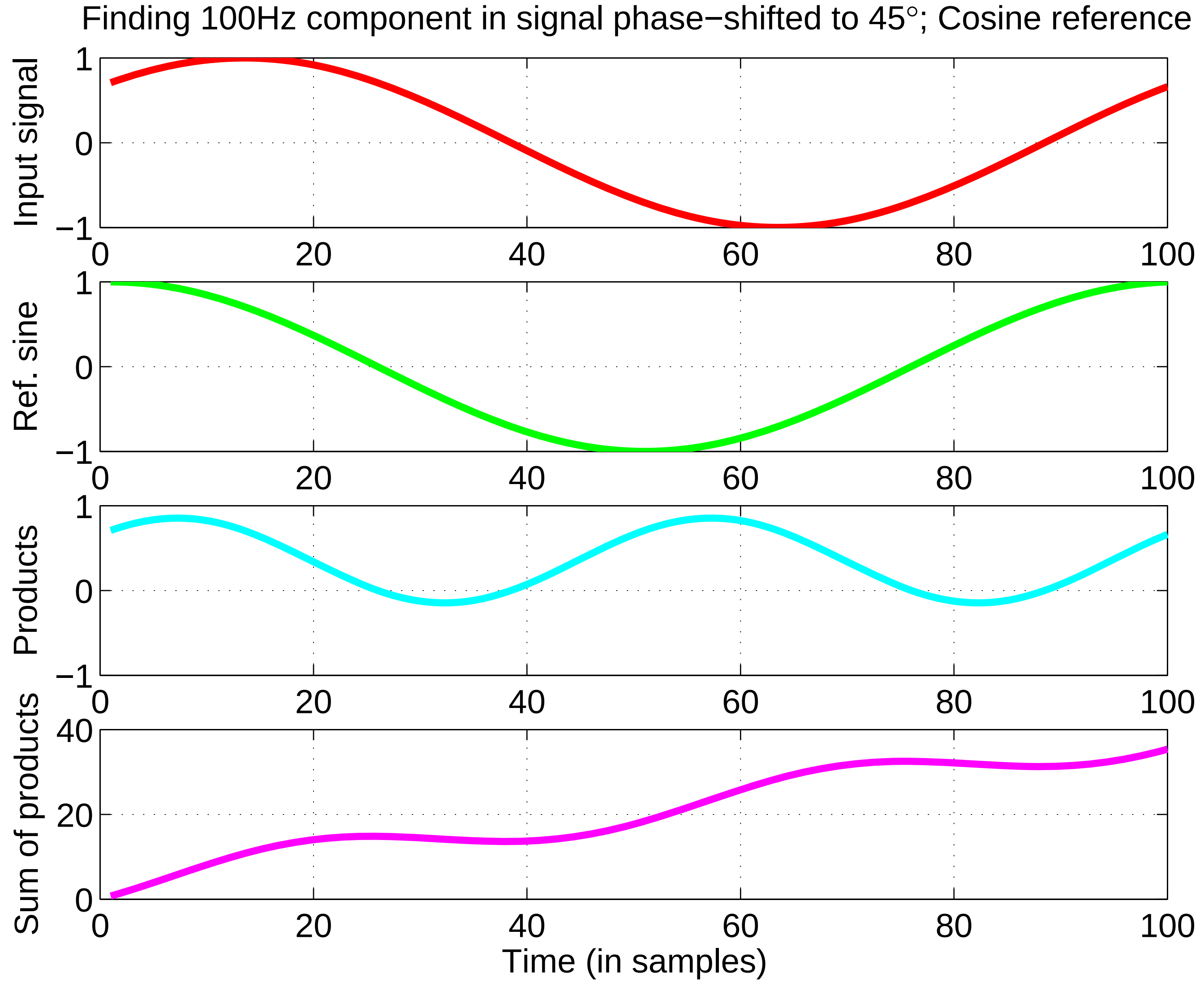
(The correlation coefficient over this set of points would be zero.)

So this provides another indication that they could be an appropriate way of setting up a coordinate system that captures all the information about a signal.

Now let’s go back to the input signal consisting of a 100Hz sine, phased−shifted 45°.

We will now do the multiplication−summation procedure on it, but using the **COSINE** as the comparison signal.

(Accordingly the left−most point of the green reference signal in the second panel of the following figure has an amplitude of 1.)


Finding 100Hz component in signal phase−shifted to 45°; Cosine reference
Input signal
Ref. sine
Products
Sum of products
1
0
−1
0
20
40
60
80
100
Time (in samples)

Now compare the amplitude of the cosine component with that of the sine component we found earlier:

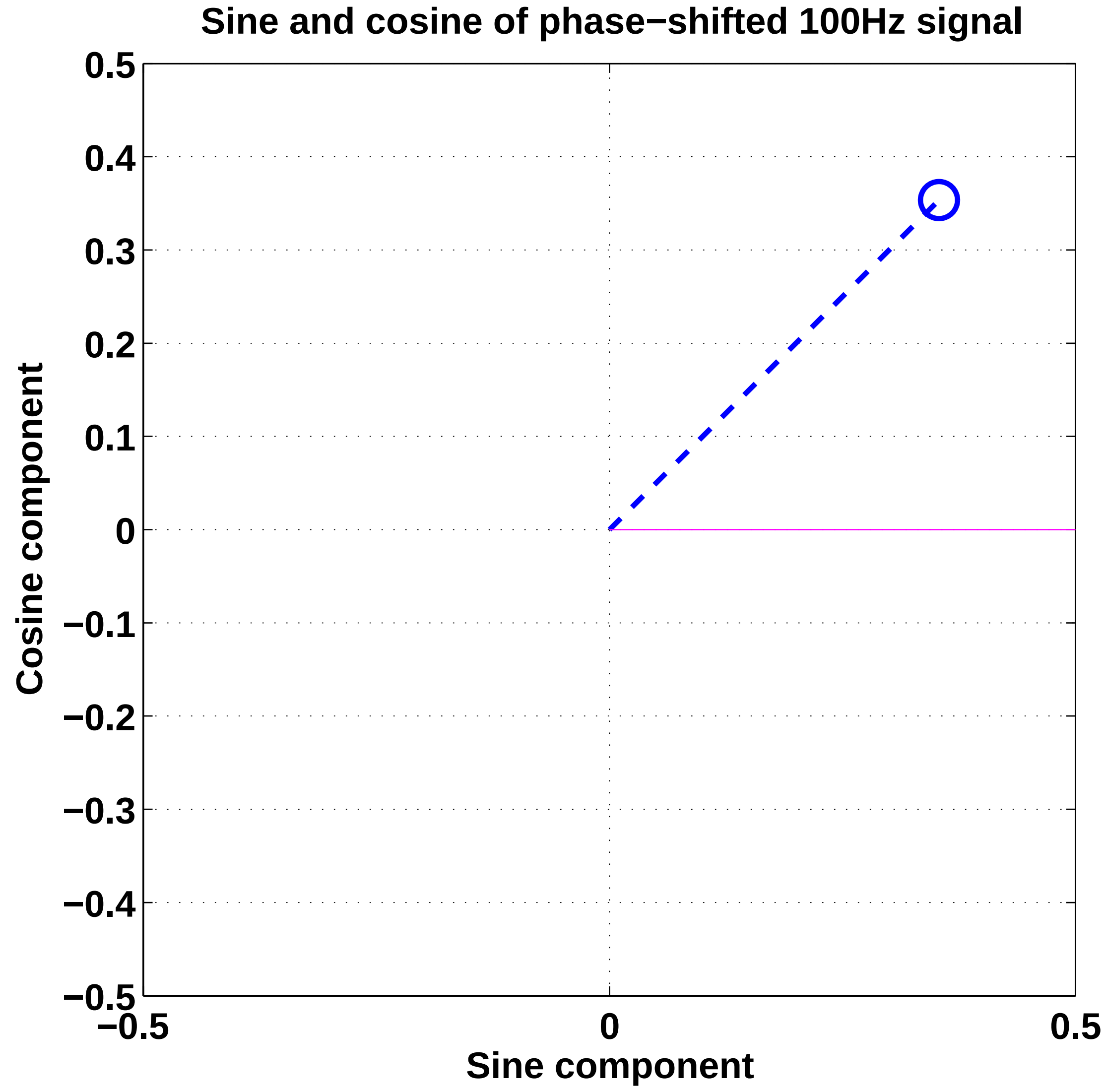
Cosine component = 0.35355
Sine component = 0.35355

Note that the values are the same.

Hardly surprising, as 45° is mid−way between 0 and 90°.

But how can we combine these two values to get the expected value of 0.5?

Lets plot the data in a two−dimensional coordinate system.


Sine and cosine of phase−shifted 100Hz signal
Cosine component
Sine component
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.5
0
0.5

How about the length of the line joining the data−point to the origin?

Using Pythagorus: sqrt(x*x + y*y)

Amplitude of combined sine and cosine = 0.5

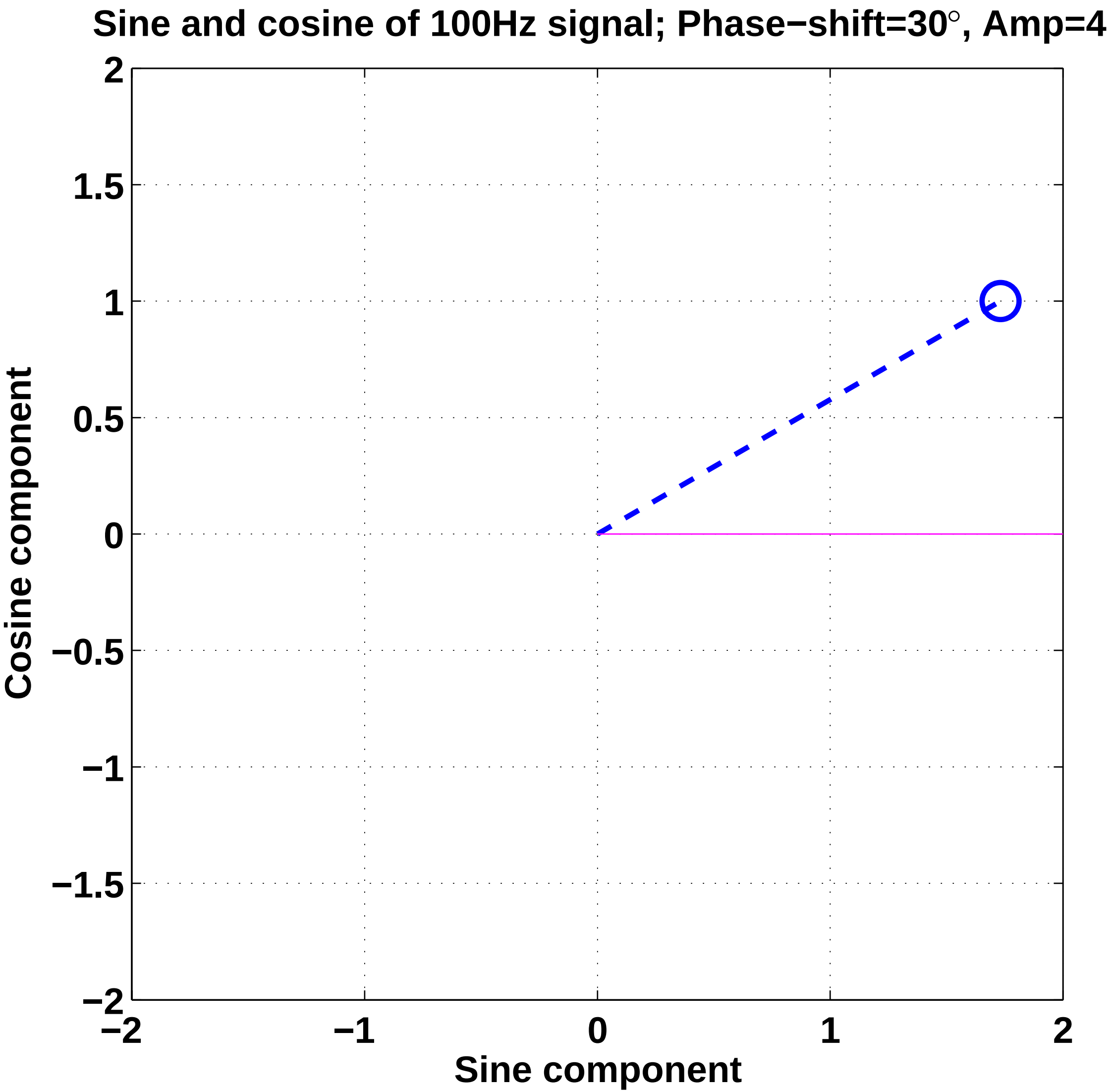
This has given us the value we wanted for the amplitude.

Moreover, note the ANGLE of the line joining
the data point to the origin:

It is obviously 45°, because in this case the
cosine and sine components (i.e the x and y coordinates)
are equal.

So this approach gives us both the amplitude and the
phase for any input signal.

As a check, let's look at the result for an input signal (100Hz), synthesized with an amplitude of 4 and a phase of 30°.



Sine and cosine of 100Hz signal; Phase−shift=30°, Amp=4
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
Cosine component
−2
−1
0
1
2
Sine component

Applying the above procedures (muliplication−summation with sine and cosine, combining them, and measuring the angle in the plane)
we are reassured to to find the following:

Amplitude = 2
Phase angle = 30

Up to now, we have been working out the amplitude and phase just for individual components in the input signal.

We will now start using the standard Matlab function FFT (i.e an implementation of the Fast Fourier Transform) that extracts all components up to samplerate/2 in one go.

To understand the results we need to go back to the remark above that standard procedures usually swap the assignment of sine and cosine components to x and y axes.

Why is this in fact preferable?

Consider how the sine and cosine of an angle is defined when we plot a right−angled triangle inside a circle:


Definition of Sine and Cosine
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
Y
−2
−1
0
1
2
X
a
b
c
ϕ

$\cos(\phi) = a\_b / a\_c$
$\sin(\phi) = b\_c / a\_c$

So assigning the cosine component to the X axis fits in better with the conventional definition of these trigonometric functions.

For consistency we must now regard an input signal that is in phase with a cosine reference signal as having a phase angle of zero.

So then we need to ask what phase a sine has relative
to a cosine:

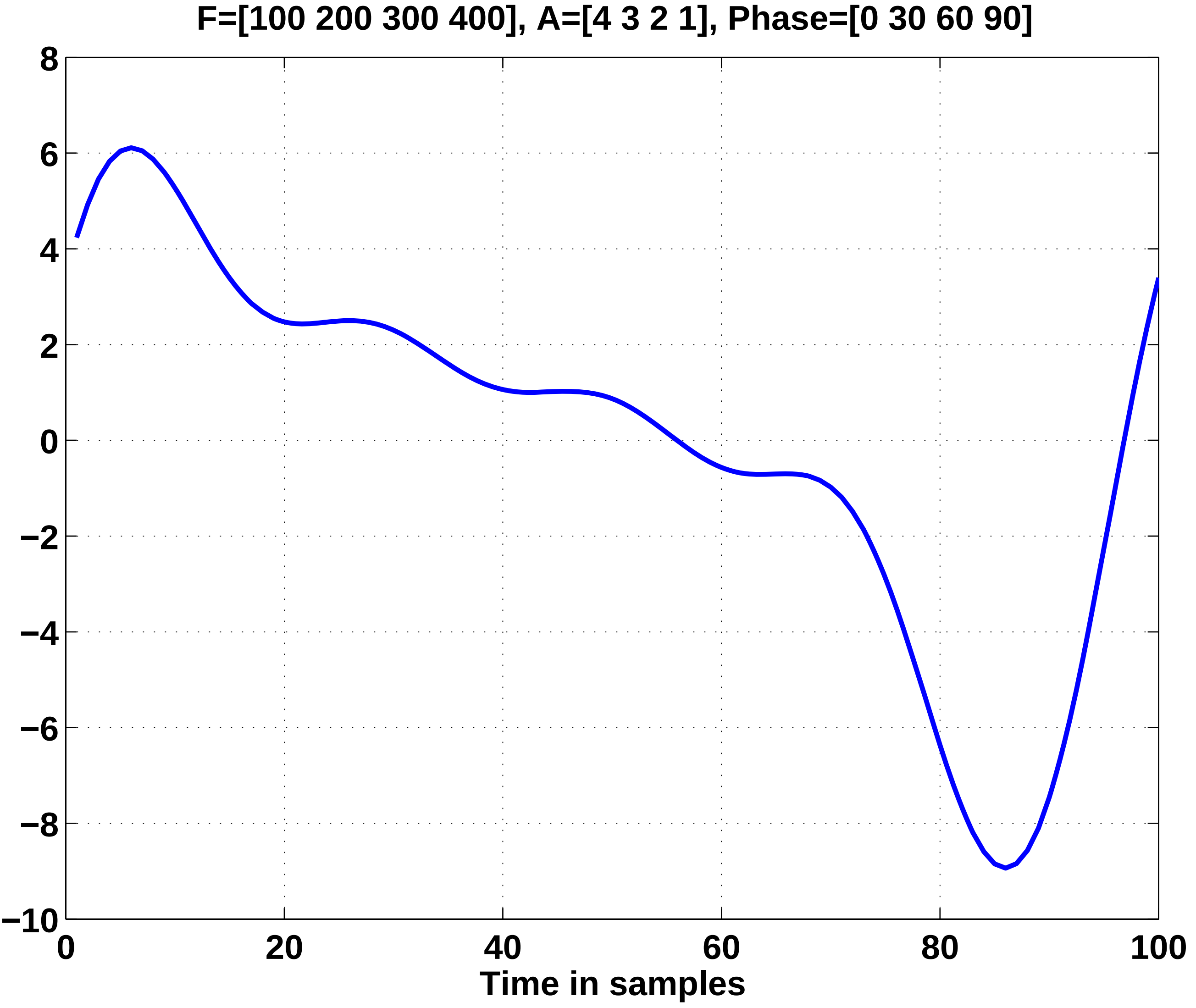
Since cosine is at $+90^\circ$ relative to sine
then sine must be at -90° (or $+270^\circ$)
relative to cosine.

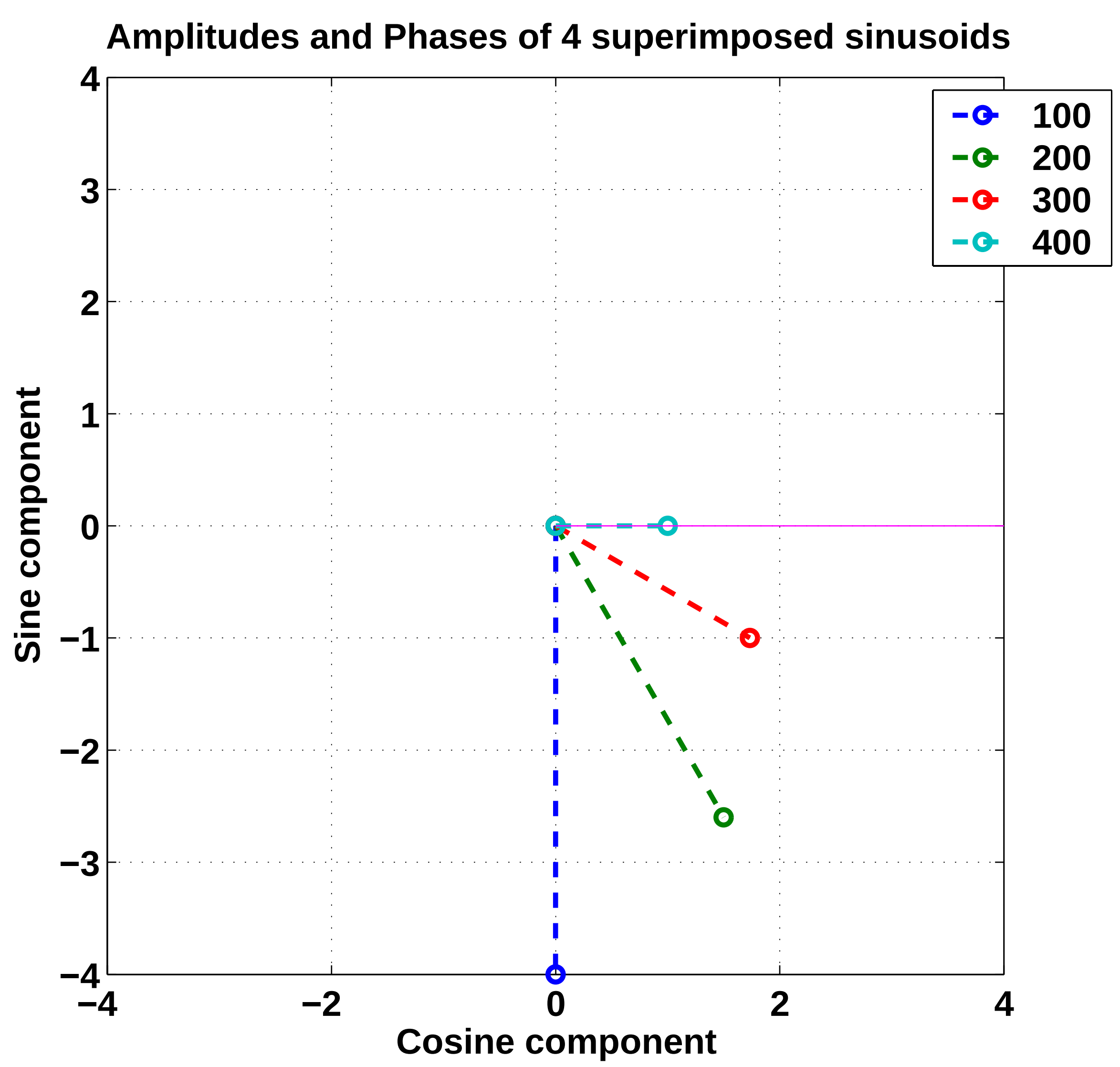
(If this is not clear refer back to the figure with an example
of a cosine reference signal, and consider where a sine wave
starts within the cosine cycle.)

So now lets look how a superposition of 4 different sine waves comes out of the standard fft procedure.

This example uses:
Frequency: 100, 200, 300, 400
Amplitude: 4, 3, 2, 1
Phase: 0, 30, 60, 90

(From now on we will now also automatically scale the amplitude so that an input amplitude of 1
comes out as 1, and not as 0.5)


F=[100 200 300 400], A=[4 3 2 1], Phase=[0 30 60 90]
8
6
4
2
0
−2
−4
−6
−8
−10
0
20
40
60
80
100
Time in samples



Amplitudes and Phases of 4 superimposed sinusoids
100
200
300
400
Sine component
Cosine component
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

Amplitudes of the 4 superimposed sine waves :

4	3	2	1

End of Part I

In Part II we look at some practical problems in the spectral analysis of speech signals