Audio Engineering Society

Convention e-Brief 56

Presented at the 148th Convention
2020 May 25 – 28, Vienna, Austria

Tutorial on Scaling of the Discrete Fourier Transform and the Implied Physical Units of the Spectra of Time-Discrete Signals

Jens Ahrens, Carl Andersson, Patrik Höstmad, and Wolfgang Kropp

Division of Applied Acoustics, Chalmers University of Technology, SE - 412 96 Gothenburg

Correspondence should be addressed to Jens Ahrens (`jens.ahrens@chalmers.se`)

ABSTRACT

The combination of the time-discrete property of digital signals together with the commonly employed definition of the discrete Fourier transform (DFT) can cause ambiguity when interpreting magnitude spectra with respect to the physical unit of the signal under consideration. Standardized scaling of spectra increases the comparability of frequency-domain data that are published in scientific articles or data sheets of commercial products. We present and discuss in this tutorial a collection of the most relevant scaling options for DFT spectra to yield amplitude spectra, power spectra, and power density spectra, and we illustrate how an implied physical unit of the underlying signal is reflected by the magnitude of the spectrum. The tutorial is accompanied by Matlab/Octave scripts that demonstrate the different cases.

1 Introduction

The Fourier transform (FT) is one of the most common mathematical operations in acoustics and audio. Definitions for continuous signals (CFT) as well as for discrete signals (DFT) exist. When speaking of a discrete signal, we refer to a signal that exhibits a dependency on a discrete variable such as time or space. We assume the instantaneous amplitude of the signal to be known with infinite accuracy.

Discrete signals require a somewhat different treatment than continuous ones because discrete signals are not physical. One aspect with which this becomes evident is interpreting the magnitude of the spectrum of a discrete signal, which is greatly facilitated when the spectrum is scaled. Although the scaling methods that we present here are widely known, we are not aware of a compact resource that summarizes the important information in an educational manner. This tutorial aims at filling this gap. Our treatment will be simplifying in a latent manner in the sense that we leave out certain less tangible details of the matter and focus on the fundamental concepts. We refer the reader to the references based on which we compiled this tutorial [1] [2, Sec. 4.2] [3, Ch. 4] [4, Ch. 6] and to the supplementary materials that we provide [5]. For ease of compactness, we omit stating proofs.

After a brief excursion to the CFT in Sec. 2, scaling options for the DFT are presented in Sec. 3.1 and 3.2. Sec. 3.3 introduces the concept single-sided spectra, which the scaling approach requires to produce a representation of the spectrum that can be interpreted conveniently. Examples are presented. Further aspects are discussed in Sec. 4 and 5.

2   The Continuous Fourier Transform

We assume the following definition of the CFT:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\,e^{-i\omega t}dt \tag{1}$$

$x(t)$ is a purely real signal that is dependent on time, $\omega = 2\pi f$ is the angular frequency, f is the frequency in Hz $= s^{-1}$, and t is time in s. $X(\omega)$ is termed the spectrum of $x(t)$. Note that above example represents a CFT over time. CFT over space exists, too [6].

Let us assume that $x(t)$ in (1) represents a microphone signal in V. In terms of the units, the complex exponential is dimensionless (i.e., it has unit 1, or no unit in other words), and it is easy to show that integration over t results in multiplying with s, the unit of t. The spectrum $X(\omega)$ is therefore given in the unit of

$$[X(\omega)] = V \cdot s = \frac{V}{Hz}. \tag{2}$$

$X(\omega)$ therefore represents an amplitude density, i.e. amplitude per frequency interval.

3   The Discrete Fourier Transform

Interpretation of the spectrum of the DFT in terms of the units is not as straightforward as with the CFT. We assume the following definition of the DFT in this tutorial:

$$X(k) = \sum_{n=0}^{N-1} x(n)\,e^{-i\frac{2\pi k}{N}n} \tag{3}$$

$x(n)$ is a signal that is dependent on the integer time index n. One also speaks of $x(n)$ as the signal at *tap* n. k is the integer frequency index also known as *bin*. Other definitions exist that differ only with respect to a normalization constant and/or the sign of the exponent. Eq. (3) appears to be the most widely used definition of the DFT and is also used by MATLAB and Python.

3.1   Scaling of DFT Spectra of Discrete Tones

A continuous signal $x(t)$ that is composed of discrete tones is characterized primarily by the amplitude and phase of each of the tones. A first property of $X(k)$ that we notice is that $|X(k)|$ is directly proportional to N, which is undesired in most situations (cf. Fig. 1). It is

preferable to scale the spectrum such that the amplitudes of said tones are apparent. Compensating for the total number of samples as

$$\overline{X}(k) = \frac{1}{N}X(k) \tag{4}$$

yields an *amplitude spectrum*, i.e. a spectrum $X(\omega)$ whose implied unit[1] is the implied unit of $x(n)$. Refer to Fig. 2 and 3 in Sec. 3.3 for examples of amplitude spectra.
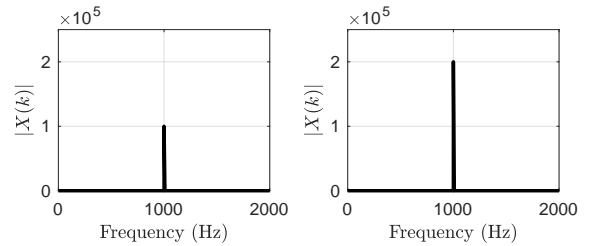

Fig. 1: $|X(k)|$ of a sine wave of frequency $f = 1000$ Hz and with amplitude $A = 1$ for $N = 1 \cdot 10^5$ (left) and $N = 2 \cdot 10^5$ (right)[2]. The horizontal axis was converted from bin index k to frequency in Hz as explained in Appendix A.

The square of this spectrum constitutes a *power spectrum*,

$$\overline{\overline{X}}(k) = |\overline{X}(k)|^2 = \frac{1}{N^2}|X(k)|^2 \tag{5}$$

the values of which are directly proportional to the power in each bin k [3]. This means that the power of each of the discrete tones in a signal is represented by $\overline{\overline{X}}(k)$ (if the energy of each discrete tone is confined to one bin). Assuming the implied unit of $x(t)$ is V, then the implied unit of $\overline{\overline{X}}(k)$ is V^2.

---

[1]We speak of an implied unit as, strictly speaking, discrete signals do not have a physical unit. If $x(n)$ is the discrete representation of a physical signal $x(t)$, then we consider $[x(t)]$, the unit of $x(t)$, the *implied unit* of $x(n)$.

[2]Note that magnitude spectra are typically plotted on a logarithmic scale, i.e. $20\log_{10}|X(k)|$ when representing amplitude or $10\log_{10}|X(k)|$ when representing power or energy. We chose a linear scale here for ease of demonstration.

[3]With electrical signals, the power P is obtained as $P = U^2/R$, with U being the effective (RMS) voltage and R being a resistance. The RMS spectrum $\overline{X}_{RMS}(k)$ is defined in (10), and its square is directly proportional to $\overline{\overline{X}}(k)$ via (8).

3.2 Scaling of DFT Spectra of Broadband Signals

As we will demonstrate in Sec. 3.3, for a continuous signal $x(t)$ that can be described as a broadband stochastic signal of some frequency distribution and power rather than an amplitude, it can be favorable to adapt the scaling.

The most important alternative to amplitude and power spectra is the *power spectral density* (PSD) or *power density spectrum*

$$\overline{\overline{\underline{X}}}(k) = \frac{N}{f_s}\overline{\overline{X}}(k) = \frac{N}{f_s}|\overline{X}(k)|^2 = \frac{1}{f_s N}|X(k)|^2 \quad (6)$$

with the implied unit V^2/Hz. The PSD is the power spectrum (5) divided by the frequency resolution f_s/N. Consider a continuous broadband signal that is sampled at different sample rates but with the same number of samples. At a lower sample rate, each bin represents a narrower frequency band so that the magnitude of the power spectrum will be lower. The PSD compensates for this.

It is also possible to create an amplitude density spectrum like (1). However, the usefulness of this with discrete signals is not obvious.

3.3 Single-Sided and Double-Sided Spectra

When computing $X(k)$ in (3) for a sufficient range of k, one finds that $X(k)$ is periodic with a period equal to the length N of the signal $x(n)$. When $x(n)$ is purely real – as it is the case with most audio-related scenarios – then $X(k)$ exhibits certain symmetries with respect to k that allow for concluding that some of the values of $X(k)$ inside one period are redundant and may be removed from the representation.

It is helpful in many situations – such as the context of this paper – to convert the (symmetric) double-sided spectrum, i.e., $X(k)$ for a given range of length N of k to a single-sided spectrum $X_{SS}(k)$ that contains only non-redundant values. To account for the change of representation through omitting non-redundant values, we multiply all values of $X(k)$ that represent a pair of a value of $X(k)$ as well as a redundant value by 2.

Let us assume in the following a spectrum $X(k)$ of a purely-real signal $x(n)$ of even length N. The symmetry is summarized as

$$X(k) = X(-k)^* \quad \forall\, 0 < k < \frac{N}{2}. \quad (7)$$

The asterisk $*$ denotes complex conjugation. The direct current (DC) bin ($k = 0$) and the bin at $k = N/2$, i.e., the bin that corresponds to the Nyquist frequency (cf. Appendix A), are purely real and unique.

We therefore define the single-sided spectrum $X_{SS}(k)$ as

$$X_{SS}(k) = \begin{cases} X(k) & \forall k = 0 \\ 2 \cdot X(k) & \forall\, 0 < k < \frac{N}{2} \\ X(k) & \forall k = \frac{N}{2} \end{cases} \quad (8)$$

Note that for odd N, all bins other than $k = 0$ have to be multiplied with 2. The single-sided representations of $\overline{X}(k)$, $\underline{X}(k)$, $\overline{\overline{X}}(k)$, and $\overline{\overline{\underline{X}}}(k)$ are obtained equivalently to (8).

To highlight the usefulness of the single-sided representation, Fig. 2 depicts the single-sided amplitude spectrum $|\overline{X}_{SS}(k)|$ (cf. (4)) of a signal $x(n)$ that represents a pure sine wave with amplitude 1 V and an amplitude offset (DC) of 1 V. The amplitude of the sine wave as well as and the DC can be directly deduced from $|\overline{X}_{SS}(k)|$.
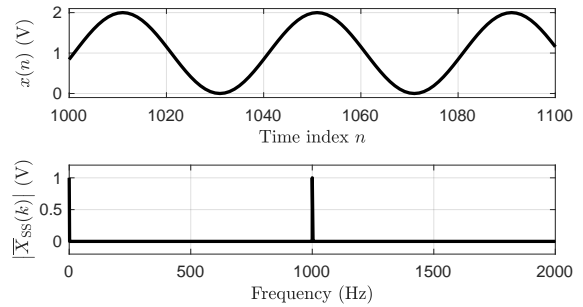

Fig. 2: Illustration of single-sided amplitude spectra. The upper plot depicts $x(n)$, a sine wave of frequency $f = 1000$ Hz, with amplitude $A = 1$ V, and a DC of 1 V. The lower plot depicts $|\overline{X}_{SS}(k)|$.

However, the picture is very different when interpreting amplitude spectra of broadband signals as illustrated in Fig. 3. The spectrum of a sine wave with additive noise is depicted for two different lengths N of the DFT. The magnitude of the sine wave is independent of N while the magnitude of the noise changes with N.

When the analysis of the noise in the signal from Fig. 3 is of interest, a power density scaling is more favorable. Fig. 4 presents the single-sided power density spectrum
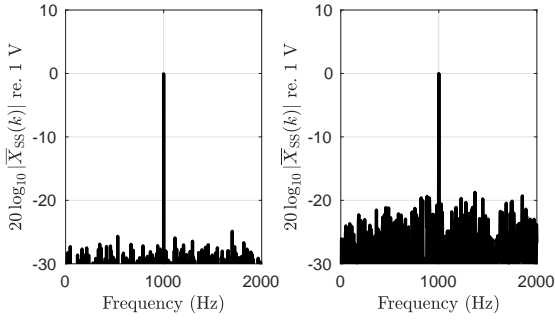
Fig. 3: Single-sided amplitude spectra $|\overline{X}_{\text{SS}}(k)|$ on a logarithmic scale of a sine of amplitude 1 and implied unit V with additive white noise. $f_{\text{s}} = 100$ kHz. Left: $N = 2 \cdot 10^5$. Right: $N = 2/8 \cdot 10^5$

$\overline{\overline{X}}_{\text{SS}}(k)$ of the signal from Fig. 3 for different sampling frequencies f_{s} and different lengths N. Contrary to the amplitude spectrum in Fig. 3, the magnitude of the noise is independent of f_{s} and N. However, the magnitude of the discrete tone changes. This is a fundamental dilemma that cannot be avoided. The reason for this is that the concept of a spectral density is not meaningful for discrete tones (and the concept of a spectral amplitude is most of the times not meaningful for broadband signals).
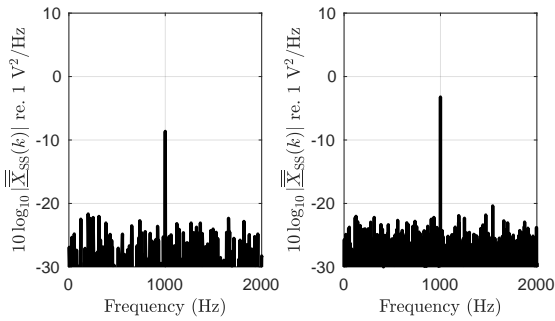

Fig. 4: Single-sided power spectral density $|\overline{\overline{X}}_{\text{SS}}(k)|$ of the signal from Fig. 3 for different sampling frequencies f_{s} and lengths N. Left: $f_{\text{s}} = 100$ kHz, $N = 25000$. Right: $f_{\text{s}} = 12.5$ kHz, $N = 12500$.

Note that while $\overline{\overline{X}}(k) = |\overline{X}(k)|^2$ (Eq. (5)) holds for double-sided power spectra, this relation reads

$$\overline{\overline{X}}_{\text{SS}}(k) = \begin{cases} \frac{1}{2}|\overline{X}_{\text{SS}}(k)|^2 & \forall 0 < k < \frac{N}{2} \\ |\overline{X}_{\text{SS}}(k)|^2 & \forall k = 0, k = \frac{N}{2} \end{cases} \quad (9)$$

for single-sided spectra (and similarly for the PSD). This can be interpreted in terms of the crest factor of sine waves of $\sqrt{2}$, i.e., the ratio of maximum value (in the case of sine waves the amplitude) to root-mean-square (RMS, cf. Sec. B). For $0 < k < \frac{N}{2}$, the RMS of the signal at each bin is obtained by dividing the magnitude of the amplitude spectrum by $\sqrt{2}$ as the basis functions of the DFT may be interpreted as sine waves:

$$\overline{X}_{\text{RMS}}(k) = \begin{cases} \frac{1}{\sqrt{2}}\overline{X}_{\text{SS}}(k) & \forall 0 < k < \frac{N}{2} \\ \overline{X}_{\text{SS}}(k) & \forall k = 0, k = \frac{N}{2} \end{cases} \quad . \quad (10)$$

We term $\overline{X}_{\text{RMS}}(k)$ *RMS spectrum*. It represents the RMS amplitudes of the discrete tones in the signal, and $\overline{\overline{X}}_{\text{SS}}(k) = |\overline{X}_{\text{RMS}}(k)|^2$ holds. A similar definition can be established for an RMS density spectrum $\underline{\overline{X}}_{\text{RMS}}(k)$. Note that RMS-spectra are inherently single sided and have no equivalent double-sided representation.

4 Windowing

It is common in spectral analysis of discrete signals to apply a window $w(n)$ to the signal before the DFT in (3). This needs to be taken into account in the scaling of the spectra, which has be performed differently for tonal signals as in Sec. 3.1 and for broadband signals as in Sec. 3.2:

$$\overline{X}_w(k) = \frac{1}{\sum_n w(n)} X_w(k) \quad (11)$$

$$\overline{\overline{X}}_w(k) = \frac{1}{(\sum_n w(n))^2} |X_w(k)|^2 \quad (12)$$

$$\underline{\overline{X}}_w(k) = \frac{1}{f_{\text{s}} \sum_n w^2(n)} |X_w(k)|^2 \quad (13)$$

where $X_w(k)$ is the DFT of $x(n) \cdot w(n)$. Simply put, the factors N in (4) and (5) are replaced by the sum $\sum_n w(n)$ of the window samples, and the factor N in the last equality in (6) is replaced by the sum $\sum_n w^2(n)$ of the squared window samples. Eq. (11) simplifies to (4) for the case of $w(n) = 1$, (12) to (5), and (13) to (6).

5 Other Remarks

A complex number z may be represented either as real and imaginary parts $\Re\{z\}$ and $\Im\{z\}$ or as magnitude $|z|$ and phase $\angle z$ as

$$z = \Re\{z\} + i\Im\{z\} = |z| \cdot e^{i\angle z} \quad . \quad (14)$$

Since the complex exponential is dimensionless, $\Re\{z\}$, $\Im\{z\}$, and $|z|$ carry the same physical unit that is determined from the type of spectrum that z is part of.

The transfer function $H(k)$ of a system can be represented as the ratio of output and input spectra as $H(k) = Y(k)/X(k)$. Any scaling of $Y(k)$ and $X(k)$ will cancel out. Regarding the units, $[H(k)] = [Y(k)]/[X(k)]$ holds so that $H(k)$ is dimensionless if $[Y(k)] = [X(k)]$. Similar considerations hold of the impulse response. Note that the magnitude of the unscaled spectrum of an impulse with amplitude 1 is always 1 independent of N.

6 Conclusions

The combination of scaling and the single-sided representation of the DFT spectra of purely real signals allow for spectral representations like amplitude spectra, power spectra, and power density spectra of discrete signals that can be conveniently interpreted in terms of the implied physical units. Different scaling is required for discrete tones and for stochastic signals to make the scaled magnitude spectrum of one of the two signal types independent of the length of the Fourier transform and the sampling frequency. This causes a fundamental dilemma when analysing signals that contain both types of elementary signals. We provide supplementary material to this tutorial online [5], where we illustrate, amongst other things, how the signal-to-noise ratio can be deduced from scaled spectra.

APPENDICES

A The Relation Between Bin Index and Frequency

When taking the into account with what frequency f_s a previously time-continuous signal was sampled, then it is possible to relate the bin index k in (3) to a frequency $f(k)$ in Hz that represents the center frequency of the bin as

$$f(k) = \frac{k}{N} \cdot f_\text{s} \quad \forall\, 0 \le k \le N/2.$$ (15)

Eq. (15) was used in all figures in this paper that depict spectra to express the horizontal axis in Hz.

B Root Mean Square

The RMS x_RMS of a discrete time domain signal $x(n)$ is given by

$$x_\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2}\,.$$ (16)

Inserting Parseval's theorem given by

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2$$ (17)

into (16) allows for computing the RMS from the spectrum $X(k)$ as

$$x_\text{RMS} = \sqrt{\frac{1}{N^2} \sum_{n=0}^{N-1} |X(k)|^2}\,.$$ (18)

References

[1] G. Heinzel et al. Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows. Tech. Rep. http://hdl.handle.net/11858/00-001M-0000-0013-557A-5, 2002.

[2] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles Algorithms and Applications*. Prentice-Hall, Upper Saddle River, New Jersey, USA, 3rd edition, 1996.

[3] L. Tan and J. Jiang. *Digital Signal Processing: Fundamentals and Applications*. Academic Press, Amsterdam, The Netherlands, 2nd edition, 2013.

[4] D. C. Swanson. *Signal Processing for Intelligent Sensor Systems with MATLAB*. CRC Press, London, UK, 2017.

[5] J. Ahrens, C. Andersson, P. Höstmad, and W. Kropp. Scaling of the Discrete Fourier Transform. https://github.com/AppliedAcousticsChalmers/scaling-of-the-dft, 2020.

[6] E. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, New York, 1999.