Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Jyrki Kauppinen, Jari Partanen

Fourier Transforms in Spectroscopy

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Jyrki Kauppinen, Jari Partanen

Fourier Transforms in Spectroscopy

WILEY-VCH

Berlin · Weinheim · New York · Chichester
Brisbane · Singapore · Toronto

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen
Copyright © 2001 Wiley-VCH Verlag GmbH
ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Authors:
Prof. Dr. Jyrki Kauppinen
Department of Applied Physics
University of Turku
Finland
e-mail: jyrki.kauppinen@utu.fi

Dr. Jari Partanen
Department of Applied Physics
University of Turku
Finland
e-mail: jari.partanen@utu.fi

1st edition, 2001
with 153 figures

Library of Congress Card No.: applied for

A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek - CIP Cataloguing-in-Publication-Data
A catalogue record for this publication is available from Die Deutsche Bibliothek

ISBN 3-527-40289-6

Printing: Strauss Offsetdruck GmbH, D-69509 Mörlenbach. Bookbinding: J. Schäffer GmbH & Co. KG, D-67269 Grünstadt.
Printed in the Federal Republic of Germany.

WILEY-VCH Verlag Berlin GmbH
Bühringstraße 10
D-13086 Berlin
Federal Republic of Germany

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Preface

How much should a good spectroscopist know about Fourier transforms? How well should a professional who uses them as a tool in his/her work understand their behavior? Our belief is, that a profound insight of the characteristics of Fourier transforms is essential for their successful use, as a superficial knowledge may easily lead to mistakes and misinterpretations. But the more the professional knows about Fourier transforms, the better he/she can apply all those versatile possibilities offered by them.

On the other hand, people who apply Fourier transforms are not, generally, mathematicians. Learning unnecessary details and spending years in specializing in the heavy mathematics which could be connected to Fourier transforms would, for most users, be a waste of time. We believe that there is a demand for a book which would cover understandably those topics of the transforms which are important for the professional, but avoids going into unnecessarily heavy mathematical details. This book is our effort to meet this demand.

We recommend this book for advanced students or, alternatively, post-graduate students of physics, chemistry, and technical sciences. We hope that they can use this book also later during their career as a reference volume. But the book is also targeted to experienced professionals: we trust that they might obtain new aspects in the use of Fourier transforms by reading it through.

Of the many applications of Fourier transforms, we have discussed Fourier transform spectroscopy (FTS) in most depth. However, all the methods of signal and spectral processing explained in the book can also be used in other applications, for example, in nuclear magnetic resonance (NMR) spectroscopy, or ion cyclotron resonance (ICR) mass spectrometry.

We are heavily indebted to Dr. Pekka Saarinen for scientific consultation, for planning problems for the book, and, finally, for writing the last chapter for us. We regard him as a leading specialist of linear prediction in spectroscopy. We are also very grateful to Mr. Matti Hollberg for technical consultation, and for the original preparation of most of the drawings in this book.

Jyrki Kauppinen and Jari Partanen

Turku, Finland, 13th October 2000

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Contents

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Index

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

1 Basic definitions

1.1 Fourier series

If a function $h(t)$, which varies with t, satisfies *the Dirichlet conditions*

1. $h(t)$ is defined from $t = -\infty$ to $t = +\infty$ and is periodic with some period T,
2. $h(t)$ is well-defined and single-valued (except possibly in a finite number of points) in the interval $\left[-\frac{1}{2}T,\ \frac{1}{2}T\right]$,
3. $h(t)$ and its derivative $\mathrm{d}h(t)/\mathrm{d}t$ are continuous (except possibly in a finite number of step discontinuities) in the interval $\left(-\frac{1}{2}T,\ \frac{1}{2}T\right)$, and
4. $h(t)$ is absolutely integrable in the interval $\left[-\frac{1}{2}T,\ \frac{1}{2}T\right]$, that is, $\int_{-T/2}^{T/2} |h(t)|\,\mathrm{d}t < \infty$,

then the function $h(t)$ can be expressed as a *Fourier series* expansion

$$h(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left[a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)\right] = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_0 t}, \tag{1.1}$$

where

$$\begin{cases} \omega_0 &= \dfrac{2\pi}{T} = 2\pi f_0, \\ a_n &= \dfrac{2}{T}\displaystyle\int_{-T/2}^{T/2} h(t)\cos(n\omega_0 t)\,\mathrm{d}t, \\ b_n &= \dfrac{2}{T}\displaystyle\int_{-T/2}^{T/2} h(t)\sin(n\omega_0 t)\,\mathrm{d}t, \\ c_n &= \dfrac{1}{T}\displaystyle\int_{-T/2}^{T/2} h(t)e^{-in\omega_0 t}\,\mathrm{d}t = \dfrac{1}{2}(a_n - ib_n), \\ c_{-n} &= c_n^* = \dfrac{1}{2}(a_n + ib_n). \end{cases} \tag{1.2}$$

f_0 is called the *fundamental frequency* of the system. In the Fourier series, a function $h(t)$ is analyzed into an infinite sum of harmonic components at multiples of the fundamental frequency. The coefficients a_n, b_n and c_n are the *amplitudes* of these harmonic components.

At every point where the function $h(t)$ is continuous the Fourier series converges uniformly to $h(t)$. If the Fourier series is truncated, and $h(t)$ is approximated by a sum of only a finite number of terms of the Fourier series, then this approximation differs somewhat from $h(t)$. Generally, the approximation becomes better and better as more and more terms are included.

At every point $t = t_0$ where the function $h(t)$ has a step discontinuity the Fourier series converges to the average of the limiting values of $h(t)$ as the point is approached from above and from below:

$$\left[\lim_{\varepsilon\to 0+} h(t_0+\varepsilon) + \lim_{\varepsilon\to 0+} h(t_0-\varepsilon)\right] \Big/ 2.$$

Around a step discontinuity, a truncated Fourier series overshoots at both sides near the step, and oscillates around the true value of the function $h(t)$. This oscillation behavior in the vicinity of a point of discontinuity is called the *Gibbs phenomenon*.

The coefficients c_n in Equation 1.1 are the *complex amplitudes* of the harmonic components at the frequencies $f_n = nf_0 = n/T$. The complex amplitudes c_n as a function of the corresponding frequencies f_n constitute a *discrete complex amplitude spectrum*.

Example 1.1: Examine the Fourier series of the *square wave* shown in Figure 1.1.

Solution. Applying Equation 1.2, the square wave can be expressed as the Fourier series

$$\begin{aligned} h(t) &= \frac{4}{\pi}\left[\cos(\omega_0 t) - \frac{1}{3}\cos(3\omega_0 t) + \frac{1}{5}\cos(5\omega_0 t) - \cdots\right] \\ &= \frac{4}{\pi}\sum_{n=1}^{\infty} \frac{\sin(n\pi/2)}{n}\cos(n\omega_0 t). \end{aligned}$$

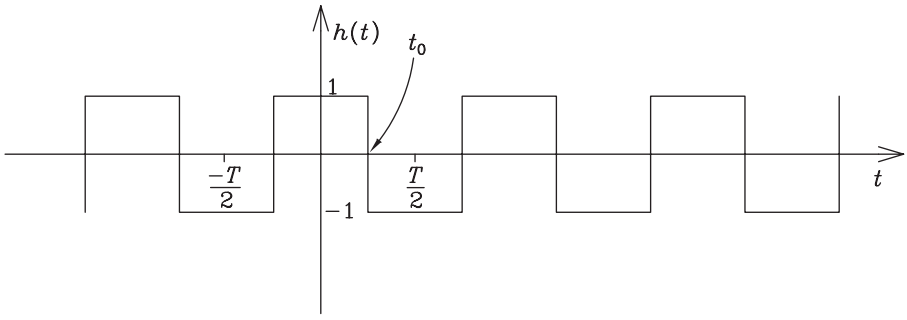


Figure 1.1: Square wave $h(t)$.

If this Fourier series is truncated, and the function is approximated by a finite sum, then this approximation differs from the original square wave, especially around the points of discontinuity. Figure 1.2 illustrates the Gibbs oscillation around the point $t = t_0$ of the square wave of Figure 1.1.

The amplitude spectrum of the square wave of Figure 1.1 is shown in Figure 1.3. The amplitude coefficients of the square wave are $c_n = \frac{1}{2}\, a_n = 0, \frac{2}{\pi}, 0, -\frac{2}{3\pi}, 0, \frac{2}{5\pi}, 0, \ldots$

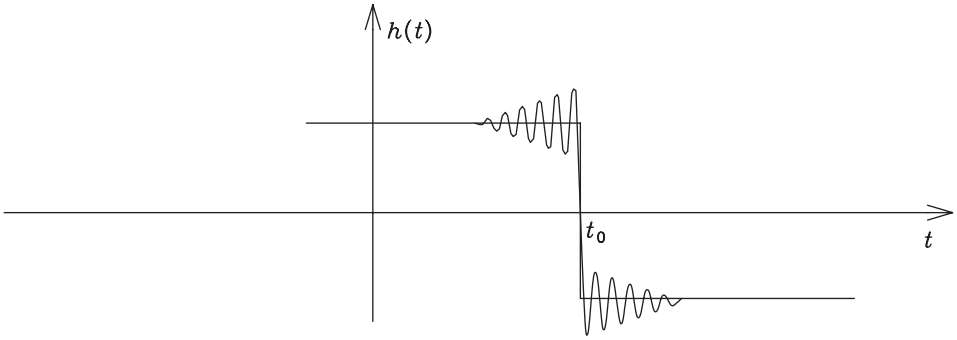


Figure 1.2: The principle how the truncated Fourier series of the square wave $h(t)$ of Fig. 1.1 oscillates around the true value in the vicinity of the point of discontinuity $t = t_0$.

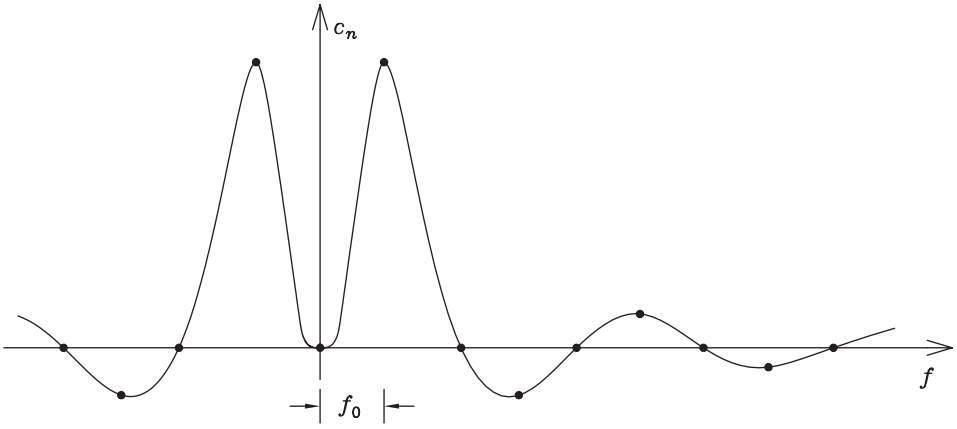


Figure 1.3: Discrete amplitude spectrum of the square wave $h(t)$ of Fig. 1.1, formed by the amplitude coefficients c_n. f_0 is the fundamental frequency.

1.2 Fourier transform

The Fourier series, Equation 1.1, can be used to analyze periodic functions of a period T and a fundamental frequency $f_0 = \frac{1}{T}$. By letting the period tend to infinity and the fundamental frequency to zero, we can obtain a generalization of the Fourier series which also is suitable for analysis of non-periodic functions.

According to Equation 1.1,

$$h(t) = \sum_{n=-\infty}^{\infty} \underbrace{\frac{1}{T} \int_{-T/2}^{T/2} h(t') e^{-i2\pi n f_0 t'} \, \mathrm{d}t'}_{c_n} e^{i12\pi n f_0 t}. \tag{1.3}$$

We shall replace nf_0 by f, and let $T \to \infty$ and $1/T = f_0 = \mathrm{d}f \to 0$. In this case,

$$\sum_{n=-\infty}^{\infty} \frac{1}{T} \to \int_{-\infty}^{\infty} \mathrm{d}f, \tag{1.4}$$

and

$$h(t) = \int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} h(t') e^{-i2\pi f t'} \, \mathrm{d}t' \right]}_{H(f)} e^{i2\pi f t} \, \mathrm{d}f. \tag{1.5}$$

We can interpret this formula as the sum of the waves $H(f)\,\mathrm{d}f\, e^{i2\pi f t}$.

With the help of the notation $H(f)$, we can write Equation 1.5 in the compact form

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{i2\pi f t} \, \mathrm{d}f = \mathcal{F}\{H(f)\}. \tag{1.6}$$

The operation $\mathcal{F}$ is called the *Fourier transform*. From above,

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi f t} \, \mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}. \tag{1.7}$$

The operation $\mathcal{F}^{-1}$ is called the *inverse Fourier transform*.

Functions $h(t)$ and $H(f)$ which are connected by Equations 1.6 and 1.7 constitute a *Fourier transform pair.* Notice that even though we have used as the variables the symbols t and f, which often refer to time [s] and frequency [Hz], the Fourier transform pair can be formed for *any variables, as long as the product of their dimensions is one* (the dimension of one variable is the inverse of the dimension of the other).

In the literature, it is possible to find several, slightly differing ways to define the Fourier integrals. They may differ in the constant coefficients in front of the integrals and in the exponents. In this book we have chosen the definitions in Equations 1.6 and 1.7, because they are the most convenient for our purposes. *In our definition, the exponential functions inside the integrals carry the coefficient* 2π*, because, in this way, we can avoid the coefficients in front of the integrals.* We have noticed that coefficients in front of Fourier integrals are a constant source of mistakes in calculations, and, by our definition, these mistakes can be avoided. Also the theorems of Fourier transform are essentially simpler, if this definition is chosen: in this way even they, except the derivative theorem, have no front coefficients.

The definition of the Fourier transform pair remains sensible, if a constant c is added in front of one integral and its inverse, constant $1/c$, is added in front of the other integral. The product of the front coefficients should equal one. We strongly encourage *not* to use definitions which do not fulfill this condition. An example of this kind of definition, sometimes encountered in literature, is obtained by setting $f = \omega/2\pi$ in Equations 1.6 and 1.7. We obtain

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{i\omega t} \, \mathrm{d}\omega = \mathcal{F}\{H(\omega)\}, \tag{1.8}$$

and

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} \, \mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}. \tag{1.9}$$

We do *not* recommend these definitions, since they easily lead to difficulties.

In our definition, the exponent inside the Fourier transform $\mathcal{F}$ *carries a positive sign, and the exponent inside the inverse Fourier transform* $\mathcal{F}^{-1}$ *carries a negative sign.* It is more common to make the definition vice versa: generally the exponent inside the Fourier transform $\mathcal{F}$ has a negative sign, and the exponent inside the inverse Fourier transform $\mathcal{F}^{-1}$ has a positive sign. Our book mainly discusses symmetric functions, and the Fourier transform and the inverse Fourier transform of a symmetric function are the same. Consequently, the choice of the sign has no scientific meaning. In our opinion, our definition is more logical, simpler, and easier to memorize: $+$ *sign in the exponent corresponds to* $\mathcal{F}$ *and* $-$ *sign in the exponent corresponds to* $\mathcal{F}^{-1}$!

We recommend that, while reading this book, the reader forgets all other definitions and uses only this simplest definition. In other contexts, the reader should always check, which definition of Fourier transforms is being used.

Table 1.1 lists a few important Fourier transform pairs, which will be useful in this book, as well as the full width at half maximum, FWHM, of these functions.

Table 1.1: Fourier transform pairs $h(t)$ and $H(f)$, and the FWHM of these functions.

Name of $h(t)$	$h(t)$	FWHM of $h(t)$	$\mathcal{F}^{-1}, \mathcal{F}$ $\Leftrightarrow$	Name of $H(t)$	$H(f)$	FWHM of $H(f)$
boxcar	$\Pi_{2T}(t) = \begin{cases} 1, & \lvert t\rvert \le T, \\ 0, & \lvert t\rvert > T \end{cases}$	$2T$		sinc	$2T\,\mathrm{sinc}(\pi 2Tf)$	$\dfrac{1.2067}{2T}$
triangular	$\Lambda_T(t) = \begin{cases} 1 - \dfrac{\lvert t\rvert}{T}, & \lvert t\rvert \le T, \\ 0, & \lvert t\rvert > T \end{cases}$	T		sinc^2	$T\,\mathrm{sinc}^2(\pi Tf)$	$\dfrac{1.7718}{2T}$
Lorentzian	$\dfrac{\sigma/\pi}{\sigma^2 + t^2}$	2σ		exponential	$\exp(-\pi 2\sigma\lvert f\rvert)$	$\dfrac{\ln 2}{\pi\sigma}$
Gaussian	$\sqrt{\dfrac{\alpha}{\pi}}\,\exp(-\alpha t^2)$	$2\sqrt{\dfrac{\ln 2}{\alpha}}$		Gaussian	$\exp\left(\dfrac{-\pi^2 f^2}{\alpha}\right)$	$\dfrac{2}{\pi}\sqrt{\alpha \ln 2}$
Dirac's delta	$\delta(t - t_0)$	0	$\mathcal{F}$ $\Rightarrow$	exponential wave	$\exp(i2\pi t_0 f)$	—
Dirac's delta	$\delta(t - t_0)$	0	$\mathcal{F}^{-1}$ $\Rightarrow$	exponential wave	$\exp(-i2\pi t_0 f)$	—

Example 1.2: Applying Fourier transforms, compute the integral $\int_0^\infty \frac{\sin(px)\cos(qx)}{x}\,\mathrm{d}x$, where $p > 0$ and $p \neq q$.

Solution. Knowing that the imaginary part of e^{iqx} is antisymmetric, we can write

$$\begin{aligned}\int_0^\infty \frac{\sin(px)\cos(qx)}{x}\,\mathrm{d}x &= \frac{p}{2}\int_{-\infty}^{\infty}\frac{\sin(px)}{px}e^{iqx}\,\mathrm{d}x = \frac{p}{2}\int_{-\infty}^{\infty}\mathrm{sinc}(px)e^{i2\pi\frac{q}{2\pi}x}\,\mathrm{d}x \\ &= \frac{\pi}{2}h(\frac{q}{2\pi}),\end{aligned}$$

where the function

$$h(t) = \mathcal{F}\{\frac{p}{\pi}\ \mathrm{sinc}(\pi\frac{p}{\pi}f)\}.$$

From Table 1.1, we know that the Fourier transform of a sinc function is a boxcar function. Consequently,

$$\frac{\pi}{2}h(\frac{q}{2\pi}) = \frac{\pi}{2}\Pi_{p/\pi}(\frac{q}{2\pi}) = \begin{cases}\pi/2, & |q| < p,\\ 0, & |q| > p.\end{cases}$$

1.3 Dirac's delta function

Dirac's delta function, $\delta(t)$, also called the *impulse function*, is a concept which is frequently used to describe quantities which are localized in one point. Even though real physical quantities cannot be truly localized in exactly one point, the concept of Dirac's delta function is very useful.

Dirac's delta function is defined with the following equation:

$$\int_{-\infty}^{\infty} F(t)\delta(t-t_0)\,\mathrm{d}t = F(t_0), \tag{1.10}$$

where $F(t)$ is an arbitrary function of t, continuous at the point $t = t_0$.

By inserting the function $F(t) \equiv 1$ in Equation 1.10, we can see that the area of Dirac's delta function is equal to unity, that is,

$$\int_{-\infty}^{\infty} \delta(t-t_0)\,\mathrm{d}t = 1. \tag{1.11}$$

In the usual sense, $\delta(t)$ is not really a function at all. In practice,

$$\lim_{t\to t_0}\delta(t-t_0) = \infty. \tag{1.12}$$

At points $t \neq t_0$ either $\delta(t-t_0) = 0$ or $\delta(t)$ oscillates with infinite frequency.

It can be shown that Dirac's delta function has the following properties:

$$\begin{cases} \delta(-t) & \triangleq & \delta(t), \\ t\delta(t) & \triangleq & 0, \\ \delta(at) & \triangleq & \dfrac{1}{a}\,\delta(t), \quad \text{if } a > 0, \\ \dfrac{\mathrm{d}\delta(t)}{\mathrm{d}t} & \triangleq & -\dfrac{1}{t}\,\delta(t), \\ F(t)\delta(t-t_0) & \triangleq & F(t_0)\delta(t-t_0), \end{cases} \tag{1.13}$$

where the correspondence relation $\triangleq$ means that the value of the integral in Equation 1.10 remains the same whichever side of the relation is inserted in the integral.

The "shape" of Dirac's delta function is not uniquely defined. There are infinitely many representations of $\delta(t)$ which satisfy Equation 1.10. One of them, very useful with Fourier transforms, is

$$\boxed{\delta(t) = \int_{-\infty}^{\infty} e^{i2\pi ts}\,\mathrm{d}s = \mathcal{F}\{1\} = \int_{-\infty}^{\infty} \cos(2\pi ts)\,\mathrm{d}s.} \tag{1.14}$$

Equivalently, we can write

$$\boxed{\delta(t) = \int_{-\infty}^{\infty} e^{-i2\pi ts}\,\mathrm{d}s = \mathcal{F}^{-1}\{1\} = \int_{-\infty}^{\infty} \cos(2\pi ts)\,\mathrm{d}s.} \tag{1.15}$$

A few other useful representations for Dirac's delta function are, for example,

$$\begin{aligned} \delta(t) \quad &\triangleq \quad \lim_{a\to\infty} \frac{\sin(a\pi t)}{\pi t} \triangleq \lim_{\sigma\to 0+} \frac{\sigma/\pi}{t^2+\sigma^2} \\ &\triangleq \quad \lim_{a\to\infty} \frac{a}{2}\,e^{-a|t|} \triangleq \lim_{a\to 0+} \frac{1}{a\sqrt{2\pi}}\,e^{-t^2/(2a^2)}. \end{aligned} \tag{1.16}$$

Two rather simple representations are

$$\delta(t) \triangleq \lim_{a\to 0+} f(t,a) \triangleq \lim_{a\to 0+} g(t,a), \tag{1.17}$$

where $f(t,a)$ and $g(t,a)$ are the functions illustrated in Figure 1.4.

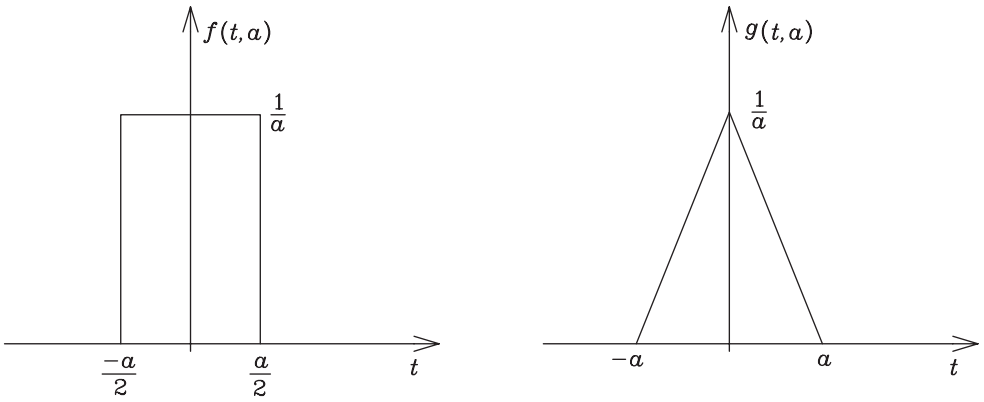


Figure 1.4: Two representations for Dirac's delta function: $\delta(t) \triangleq \lim_{a\to 0+} f(t,a) \triangleq \lim_{a\to 0+} g(t,a)$.

Example 1.3: Using Dirac's delta function in the form of Equation 1.14, prove that Equation 1.6 yields Equation 1.7.

Solution. If $h(t) = \int_{-\infty}^{\infty} H(f) e^{i2\pi f t}\,\mathrm{d}f$, then

$$\begin{aligned}\mathcal{F}^{-1}\{h(t)\} &= \int_{-\infty}^{\infty} h(t) e^{-i2\pi f' t}\,\mathrm{d}t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(f) e^{i2\pi t(f-f')}\,\mathrm{d}f\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} H(f)\left[\int_{-\infty}^{\infty} e^{i2\pi t(f-f')}\,\mathrm{d}t\right]\mathrm{d}f \overset{(1.14)}{=} \int_{-\infty}^{\infty} H(f)\delta(f-f')\,\mathrm{d}f \\ &\overset{(1.10)}{=} H(f').\end{aligned}$$

Consequently,

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi f t}\,\mathrm{d}t,$$

that is, Equation 1.6 $\Rightarrow$ Equation1.7.

Problems

1. Show that the Fourier transform $\mathcal{F}$ and the inverse Fourier transform $\mathcal{F}^{-1}$ of a symmetric function are the same. Also show that these Fourier transforms are symmetric.

2. Compute $\mathcal{F}\{\mathcal{F}\{h(t)\}\}$ and $\mathcal{F}^{-1}\{\mathcal{F}^{-1}\{h(t)\}\}$.

3. Derive the Fourier transforms of the Gaussian curve, *i.e.*,

$$\mathcal{F}\left\{\sqrt{\frac{\alpha}{\pi}}\,\exp\left(-\alpha f^2\right)\right\} \quad \text{and} \quad \mathcal{F}^{-1}\left\{\sqrt{\frac{\alpha}{\pi}}\,\exp\left(-\alpha f^2\right)\right\},$$

   where $\alpha > 0$.

4. Derive the Fourier transform of the Lorentzian function $L(f) = \dfrac{\sigma/\pi}{f^2+\sigma^2}$, $\sigma > 0$. (Function $L(f)$ is symmetric, and hence $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the same.)

   Hint: Use the following integral from mathematical tables: $\displaystyle\int_0^\infty \frac{\cos(Cx)}{x^2+a^2}\,\mathrm{d}x = \frac{\pi}{2a}\,e^{-Ca}$, where $C, a > 0$.

5. Applying Fourier transforms, compute the integral $I = \displaystyle\int_{-T}^{T} (T-|t|)\cos[2\pi f_0(T-t)]\,\mathrm{d}t$.

6. Let us denote $\Pi_{2T'}(t)$ the boxcar function of the width $2T'$, stretching from $-T'$ to T', and of one unit height. The sum of N boxcar functions

$$\frac{1}{N}\,\Pi_{\frac{1}{N}2T}(t),\ \frac{1}{N}\,\Pi_{\frac{2}{N}2T}(t),\ \frac{1}{N}\,\Pi_{\frac{3}{N}2T}(t),\ \ldots,\ \frac{1}{N}\,\Pi_{2T}(t)$$

   is a one-unit high step-pyramidal function. In the limit $N \to \infty$ this pyramidal function approaches a one-unit high triangular function $\Lambda_T(t)$. Determine the inverse Fourier transform of the pyramidal function, and find the inverse Fourier transform of $\Lambda_T(t)$ by letting $N \to \infty$.

   Hint: You may need the trigonometric identity

$$\sin\alpha + \sin(2\alpha) + \cdots + \sin(N\alpha) = \frac{\sin\left[\frac{1}{2}\,(N+1)\alpha\right]\sin\left(\frac{1}{2}\,N\alpha\right)}{\sin(\alpha/2)}.$$

7. Show that the function

$$\delta(t) = \int_{-\infty}^{\infty} e^{i2\pi f t}\,\mathrm{d}f$$

   satisfies the definition of Dirac's delta function by using the Fourier's integral theorem

$$\mathcal{F}^{-1}\mathcal{F}\{F\} = F.$$

8. Show that the function $f(t) = \lim_{\sigma \to 0} \dfrac{\sigma/\pi}{t^2 + \sigma^2}$ satisfies the definition of Dirac's delta function.

9. Let us define the step function $u(t)$ as

$$u(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0. \end{cases}$$

Show that the choice $\delta(t - t_0) = \dfrac{\mathrm{d}}{\mathrm{d}t_0} u(t_0 - t)$ satisfies the condition of Dirac's delta function.

Hint: You can change the order of differentiation and integration.

10. Compute the integral $\int_{-\varepsilon}^{\varepsilon} \delta(t)\,\mathrm{d}t$, where $\varepsilon > 0$, by inserting $\delta(t) = \int_{-\infty}^{\infty} e^{\pm i 2\pi f t}\,\mathrm{d}f$ in it.

11. What are the Fourier transforms (both $\mathcal{F}$ and $\mathcal{F}^{-1}$) of the following functions? Let us assume that $f_0 > 0$.

    (a) $\delta(f)$ (Dirac's delta peak in the origin);

    (b) $\delta(f - f_0)$ (Dirac's delta peak at f_0);

    (c) $\delta(f - f_0) + \delta(f + f_0)$ (Dirac's delta peaks at f_0 and $-f_0$);

    (d) $\delta(f - f_0) - \delta(f + f_0)$ (Dirac's delta peak at f_0 and negative Dirac's delta peak at $-f_0$).

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

2 General properties of Fourier transforms

The general definitions of the Fourier transform $\mathcal{F}$ and the inverse Fourier transform $\mathcal{F}^{-1}$ are

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{i2\pi ft}\,\mathrm{d}f = \mathcal{F}\{H(f)\}, \tag{2.1}$$

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,\mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}. \tag{2.2}$$

$H(f)$ is called the *spectrum* and $h(t)$ is the *signal*. The inverse Fourier transform operator $\mathcal{F}^{-1}$ generates the spectrum from a signal, and the Fourier transform operator $\mathcal{F}$ restores the signal from a spectrum. A single point in the spectrum corresponds to a single exponential wave in the signal, and vice versa. The signal is often defined in the time domain (t-domain), and the spectrum in the frequency domain (f-domain), as above.

Usually, the signal $h(t)$ is real. The spectrum $H(f)$ can still be complex, because

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,\mathrm{d}t = \int_{-\infty}^{\infty} h(t)\cos(2\pi ft)\,\mathrm{d}t - i\int_{-\infty}^{\infty} h(t)\sin(2\pi ft)\,\mathrm{d}t \\ &= \mathcal{F}_{\cos}\{h(t)\} - i\mathcal{F}_{\sin}\{h(t)\} = |H(f)|e^{i\theta(f)}. \end{aligned} \tag{2.3}$$

$\mathcal{F}_{\cos}\{h(t)\}$ and $\mathcal{F}_{\sin}\{h(t)\}$ are called the *cosine transform* and the *sine transform*, respectively, of $h(t)$.

$$|H(f)| = \sqrt{[\mathcal{F}_{\cos}\{h(t)\}]^2 + [\mathcal{F}_{\sin}\{h(t)\}]^2} \tag{2.4}$$

is the *amplitude spectrum* and

$$\theta(f) = -\arctan\frac{\mathcal{F}_{\sin}\{h(t)\}}{\mathcal{F}_{\cos}\{h(t)\}} \tag{2.5}$$

is the *phase spectrum.* (Equation 2.5 is valid only when $-\pi/2 \leq \theta \leq \pi/2$.) The amplitude spectrum and the phase spectrum are illustrated in Figure 2.1. The inverse Fourier transform and the Fourier transform can be expressed with the help of the cosine transform and the sine transform as

$$\mathcal{F}^{-1}\{h(t)\} = \mathcal{F}_{\cos}\{h(t)\} - i\mathcal{F}_{\sin}\{h(t)\} \tag{2.6}$$

and

$$\mathcal{F}\{H(f)\} = \mathcal{F}_{\cos}\{H(f)\} + i\mathcal{F}_{\sin}\{H(f)\}. \tag{2.7}$$

Often, $i\mathcal{F}_{\sin}\{H(f)\}$ equals zero.

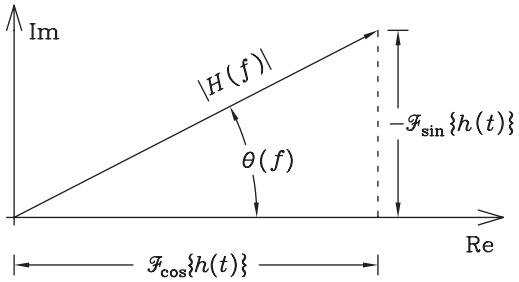


Figure 2.1: The amplitude spectrum $|H(f)|$ and the phase spectrum $\theta(f)$.

In the following, a collection of theorems of Fourier analysis is presented. They contain the most important characteristic features of Fourier transforms.

2.1 Shift theorem

Let us consider how a shift $\pm f_0$ of the frequency of a spectrum $H(f)$ affects the corresponding signal $h(t)$, which is the Fourier transform of the spectrum. We can find this by making a change of variables in the Fourier integral:

$$\begin{aligned}
\mathcal{F}\{H(f \pm f_0)\} &= \int_{-\infty}^{\infty} H(f \pm f_0) e^{i2\pi f t}\, \mathrm{d}f \\
&\overset{\substack{g=f\pm f_0 \\ \mathrm{d}g=\mathrm{d}f}}{=} \int_{-\infty}^{\infty} H(g) e^{i2\pi(g\mp f_0)t}\, \mathrm{d}g = e^{\mp i2\pi f_0 t} \int_{-\infty}^{\infty} H(g) e^{i2\pi g t}\, \mathrm{d}g \\
&= h(t) e^{\pm i2\pi f_0 t}.
\end{aligned}$$

Likewise, we can obtain the effect of a shift $\pm t_0$ of the position of the signal $h(t)$ on the spectrum $H(f)$, which is the inverse Fourier transform of the signal.

These results are called the *shift theorem*. The theorem states how the shift in the position of a function affects its transform. If $H(f) = \mathcal{F}^{-1}\{h(t)\}$, and both t_0 and f_0 are constants, then

$$\boxed{\begin{aligned}
\mathcal{F}\{H(f \pm f_0)\} &= \mathcal{F}\{H(f)\} e^{\mp i2\pi f_0 t} &&= h(t) e^{\mp i2\pi f_0 t}, \\
\mathcal{F}^{-1}\{h(t \pm t_0)\} &= \mathcal{F}^{-1}\{h(t)\} e^{\pm i2\pi f t_0} &&= H(f) e^{\pm i2\pi f t_0}.
\end{aligned}} \tag{2.8}$$

This means that the Fourier transform of a shifted function is the Fourier transform of the unshifted function multiplied by an exponential wave or phase factor. The same holds true for the inverse Fourier transform.

If a function is shifted away from the origin, its transform begins to oscillate at the frequency given by the shift. A shift in the location of a function in one domain corresponds to multiplication by a wave in the other domain.

2.2 Similarity theorem

Let us next examine multiplication of the frequency of a spectrum by a positive real constant a. We shall take the Fourier transform, and apply the change of variables:

$$\begin{aligned}
\mathcal{F}\{H(af)\} &= \int_{-\infty}^{\infty} H(af) e^{i2\pi f t}\,\mathrm{d}f \\
&\overset{\substack{g=af \\ \mathrm{d}g=a\mathrm{d}f}}{=} \int_{-\infty}^{\infty} \frac{1}{a} H(g) e^{i2\pi g t/a}\,\mathrm{d}g \\
&= \frac{1}{a} h(t/a).
\end{aligned}$$

The inverse Fourier transform of a signal can be examined similarly.

We obtain the results that if $H(f) = \mathcal{F}^{-1}\{h(t)\}$, and a is a positive real constant, then

$$\boxed{\mathcal{F}\{H(af)\} = \frac{1}{a} h(t/a),} \tag{2.9}$$

and

$$\boxed{\mathcal{F}^{-1}\{h(at)\} = \frac{1}{a} H(f/a).} \tag{2.10}$$

These statements are called the *similarity theorem*, or the *scaling theorem*, of Fourier transforms. The theorem tells that a contraction of the coordinates in one domain leads to a corresponding stretch of the coordinates in the other domain.

If a function is made to decay faster ($a > 1$), keeping the height constant, then the Fourier transform of the function becomes wider, but lower in height. If a function is made to decay slower ($0 < a < 1$), its Fourier transform becomes narrower, but taller. The same applies to the inverse Fourier transform $\mathcal{F}^{-1}$.

2.3 Modulation theorem

The *modulation theorem* explains the behavior of the Fourier transform when a function is modulated by multiplying it by a harmonic function. A straightforward computation gives that

$$\begin{aligned}
\mathcal{F}\{H(f)\cos(2\pi t_0 f)\} &= \int_{-\infty}^{\infty} H(f)\cos(2\pi t_0 f)e^{i2\pi f t}\,\mathrm{d}f \\
&= \int_{-\infty}^{\infty} H(f)\,\frac{e^{i2\pi t_0 f}+e^{-i2\pi t_0 f}}{2}\,e^{i2\pi f t}\,\mathrm{d}f \\
&= \frac{1}{2}\int_{-\infty}^{\infty} H(f)e^{i2\pi(t+t_0)f}\,\mathrm{d}f + \frac{1}{2}\int_{-\infty}^{\infty} H(f)e^{i2\pi(t-t_0)f}\,\mathrm{d}f \\
&= \frac{1}{2}h(t+t_0)+\frac{1}{2}h(t-t_0).
\end{aligned}$$

A similar result can be obtained for the inverse Fourier transform.

Consequently, we obtain that if $H(f)=\mathcal{F}^{-1}\{h(t)\}$, and t_0 and f_0 are real constants, then

$$\boxed{\mathcal{F}\{H(f)\cos(2\pi t_0 f)\}=\frac{1}{2}h(t+t_0)+\frac{1}{2}h(t-t_0),} \tag{2.11}$$

and

$$\boxed{\mathcal{F}^{-1}\{h(t)\cos(2\pi f_0 t)\}=\frac{1}{2}H(f+f_0)+\frac{1}{2}H(f-f_0).} \tag{2.12}$$

These results are the modulation theorem. The Fourier transform of a function multiplied by $\cos(2\pi t_0 f)$ is the average of two Fourier transforms of the original function, one shifted in the negative direction by the amount t_0, and the other in the positive direction by the amount t_0. The inverse Fourier transform is affected by harmonic modulation in a similar way.

2.4 Convolution theorem

The *convolution* of functions $g(t)$ and $h(t)$ is defined as

$$g(t)*h(t)=\int_{-\infty}^{\infty} g(u)h(t-u)\,\mathrm{d}u. \tag{2.13}$$

This convolution integral is a function which varies with t. It can be thought as the area of the product of $g(u)$ and $h(t-u)$, varying with t. One of the two functions is folded backwards: $h(-u)$ is the mirror image of $h(u)$, reflected about the origin. This is illustrated in Figure 2.2.

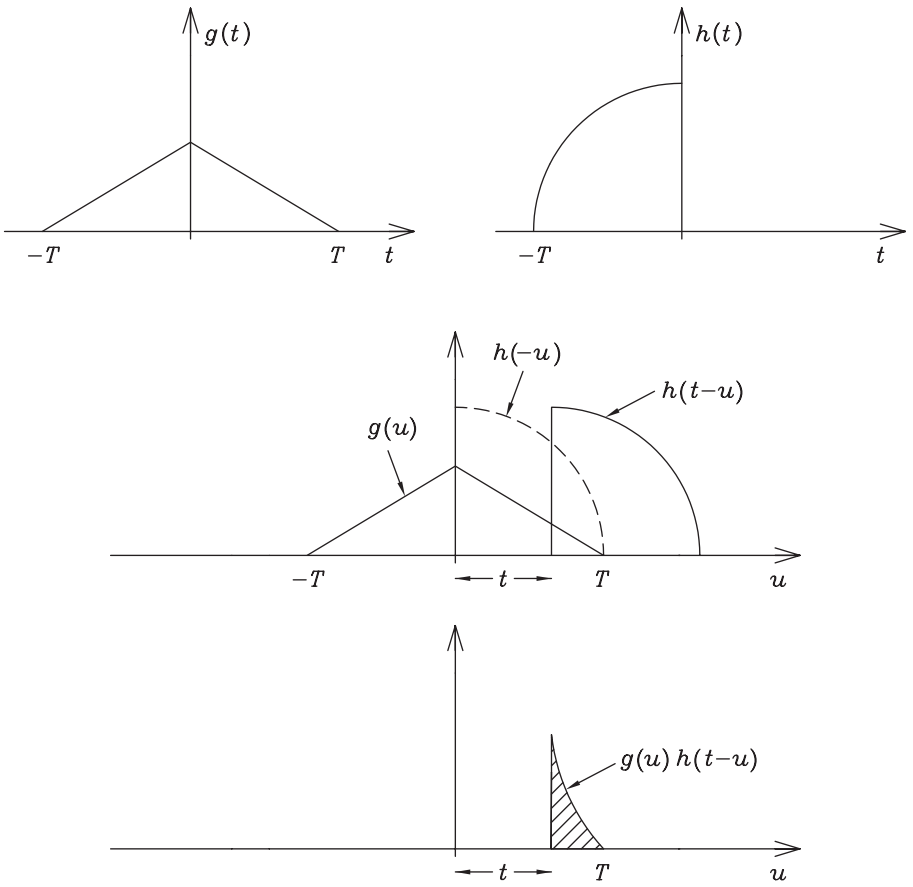


Figure 2.2: The principle of convolution. Convolution of the two functions $g(t)$ and $h(t)$ (upper curves) is the area of the product $g(u)h(t-u)$ (the hatched area inside the lowest curve), as a function of the shift t of $h(-u)$.

The function g can be thought to be an object function, which is convolved by the other, convolving function h. The convolving function is normally a peak-shaped function, and the effect of the convolution is to smear the object function. If also the object function is peak-shaped, then the convolution broadens it.

Convolution is always broader than the broadest of the two functions which are convolved. Convolution is also always smoother than the smoothest of the two functions. Convolution with Dirac's delta function is a special borderline case:

$$\delta(t) * f(t) = f(t), \tag{2.14}$$

and thus convolution with Dirac's delta function is as broad and as smooth as the second function.

Convolution can be shown to have the following properties:

$$\left\{\begin{array}{rcll} h*g & = & g*h & \text{(commutative)},\\ h*(g*k) & = & (h*g)*k & \text{(associative)},\\ h*(g+k) & = & (h*g)+(h*k) & \text{(distributive)}. \end{array}\right. \tag{2.15}$$

Let us examine the Fourier transform and the inverse Fourier transform of a product of two functions. If $H(f)=\mathcal{F}^{-1}\{h(t)\}$, and $G(f)=\mathcal{F}^{-1}\{g(t)\}$, then it is can be shown (Problem 5) that

$$\boxed{\begin{array}{rcl} \mathcal{F}\{H(f)G(f)\} & = & \mathcal{F}\{H(f)\}*\mathcal{F}\{G(f)\} = h(t)*g(t),\\ \mathcal{F}^{-1}\{h(t)*g(t)\} & = & \mathcal{F}^{-1}\{h(t)\}\mathcal{F}^{-1}\{g(t)\} = H(f)G(f) \end{array}} \tag{2.16}$$

and

$$\boxed{\begin{array}{rcl} \mathcal{F}^{-1}\{h(t)g(t)\} & = & H(f)*G(f),\\ \mathcal{F}\{H(f)*G(f)\} & = & h(t)g(t). \end{array}} \tag{2.17}$$

These equations are the *convolution theorem* of Fourier transforms. They can be verified either by a change of variables of integration or by applying Dirac's delta function. The convolution theorem states that convolution in one domain corresponds to multiplication in the other domain.

The Fourier transform of the product of two functions is the convolution of the two individual Fourier transforms of the two functions. The same holds true for the inverse Fourier transform: the inverse Fourier transform of the product of two functions is the convolution of the individual inverse Fourier transforms.

On the other hand, the Fourier transform of the convolution of two functions is the product of the two individual Fourier transforms. And the inverse Fourier transform of the convolution of two functions is the product of the two individual inverse Fourier transforms.

2.5 Power theorem

The *complex conjugate* u^* of a complex number u is obtained by changing the sign of its imaginary part, that is, by replacing i by $-i$.

If $h(t)$ and $H(f)$, and $g(t)$ and $G(f)$ are Fourier transform pairs, then

$$\begin{aligned} \int_{-\infty}^{\infty} h(t)g^*(t)\,\mathrm{d}t &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[H(f)e^{i2\pi ft}\,\mathrm{d}f\right]g^*(t)\,\mathrm{d}t\\ &= \int_{-\infty}^{\infty} H(f)\left[\int_{-\infty}^{\infty} g(t)e^{-i2\pi ft}\,\mathrm{d}t\right]^*\mathrm{d}f\\ &= \int_{-\infty}^{\infty} H(f)G^*(f)\,\mathrm{d}f. \end{aligned}$$

This result,

$$\boxed{\int_{-\infty}^{\infty} h(t)g^*(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} H(f)G^*(f)\,\mathrm{d}f,} \tag{2.18}$$

is called the *power theorem* of Fourier transforms: the overall integral of the product of a function and the complex conjugate of a second function equals the overall integral of the product of the transform of the function and the complex conjugate of the transform of the second function.

2.6 Parseval's theorem

The *Parseval's theorem* can be deduced from the power theorem (Equation 2.18) by choosing that $h = g$. We obtain that

$$\boxed{\int_{-\infty}^{\infty} |h(t)|^2\,\mathrm{d}t = \int_{-\infty}^{\infty} |H(f)|^2\,\mathrm{d}f.} \tag{2.19}$$

The area under the squared absolute value of a function is the same as the area under the squared absolute value of the Fourier transform of the function.

2.7 Derivative theorem

The *derivative theorem* states the form of the Fourier transform of the derivative of a function. Differentiating, with respect to t, both sides of equation

$$h(t) = \mathcal{F}\{H(f)\} = \int_{-\infty}^{\infty} e^{i2\pi ft} H(f)\,\mathrm{d}f, \tag{2.20}$$

we obtain that

$$\frac{\mathrm{d}h(t)}{\mathrm{d}t} = h^{(1)}(t) = \int_{-\infty}^{\infty} (i2\pi f) e^{i2\pi ft} H(f)\,\mathrm{d}f. \tag{2.21}$$

Repetition of this operation k times yields

$$\frac{\mathrm{d}^k h(t)}{\mathrm{d}t^k} = h^{(k)}(t) = \int_{-\infty}^{\infty} (i2\pi f)^k e^{i2\pi ft} H(f)\,\mathrm{d}f. \tag{2.22}$$

On the other hand,

$$\int_{-\infty}^{\infty} (i2\pi f)^k e^{i2\pi ft} H(f)\,\mathrm{d}f = \mathcal{F}\left\{(i2\pi f)^k H(f)\right\}. \tag{2.23}$$

Taking inverse Fourier transforms of both sides of Equations 2.22 and 2.23 yields

$$\boxed{\mathcal{F}^{-1}\left\{\frac{\mathrm{d}^k h(t)}{\mathrm{d}t^k}\right\} = (i2\pi f)^k H(f).} \tag{2.24}$$

If a function is differentiated k times, its inverse Fourier transform is multiplied by $(i2\pi f)^k$.

Likewise, it can be shown that

$$\boxed{\mathcal{F}\left\{\frac{\mathrm{d}^k H(f)}{\mathrm{d}f^k}\right\} = (-i2\pi t)^k h(t).} \tag{2.25}$$

2.8 Correlation theorem

The *cross correlation* between two functions $h(t)$ ja $g(t)$ is defined as

$$x(t) = h(t) \circledast g(t) = \int_{-\infty}^{\infty} h^*(u-t)g(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} h^*(u)g(u+t)\,\mathrm{d}u. \tag{2.26}$$

The cross correlation function is essentially different from the convolution function, because in cross correlation none of the functions is folded backwards. The cross correlation does generally not commute: $h(t) \circledast g(t) \neq g(t) \circledast h(t)$.

The *correlation theorem* states that if $h(t)$ and $H(f)$, and $g(t)$ and $G(f)$, are Fourier transform pairs, then

$$\boxed{\begin{aligned} \mathcal{F}\{H^*(f)G(f)\} &= h(t) \circledast g(t) = \int_{-\infty}^{\infty} h^*(u)g(u+t)\,\mathrm{d}u, \\ \mathcal{F}^{-1}\{h^*(t)g(t)\} &= H(f) \circledast G(f) = \int_{-\infty}^{\infty} H^*(v)G(v+f)\,\mathrm{d}v. \end{aligned}} \tag{2.27}$$

This resembles the convolution theorem, Equations 2.16 and 2.17, but now the first function $H(f)$ of the product $H(f)G(f)$ is replaced by its complex conjugate $H^*(f)$. The correlation theorem may be verified either by change of variables of integration or by applying Dirac's delta function. It can also be derived from the convolution theorem (Problem 8).

2.9 Autocorrelation theorem

If cross correlation of a function (Equation 2.26) is taken with the function itself, the operation is called *autocorrelation*. The autocorrelation function of function $h(t)$ is

$$a(t) = h(t) \circledast h(t) = \int_{-\infty}^{\infty} h^*(u-t)h(u)\,\mathrm{d}u. \tag{2.28}$$

If we set $H(f) = G(f)$ in the correlation theorem (Equation 2.27), we obtain that

$$\boxed{\mathcal{F}\left\{|H(f)|^2\right\} = h(t) \circledast h(t).} \tag{2.29}$$

Similarly,

$$\boxed{\mathcal{F}^{-1}\left\{|h(t)|^2\right\} = H(f) \circledast H(f).} \tag{2.30}$$

These equations are the *autocorrelation theorem* of Fourier transforms.

The *power spectrum* $|H(f)|^2$ of the signal $h(t)$ is the square of the amplitude spectrum $|H(f)|$ in Equation 2.4. It is equal to the inverse Fourier transform of the autocorrelation function:

$$\boxed{|H(f)|^2 = \mathcal{F}^{-1}\{h(t) \circledast h(t)\}.} \tag{2.31}$$

The power spectrum is used when the phase spectrum $\theta(f)$ can be omitted. This is the case in the examination of noise, shake, seismograms, and so forth.

Example 2.1: Tex Willer shoots six shots in one second and then loads his revolver for four seconds. These periods of five seconds repeat identically for the whole duration of one minute of a gunfight. Calculate the power spectrum of the sounds of the shooting.

Solution. We can assume that the sounds of the shots are so short that they can be considered Dirac's delta peaks. The inverse Fourier transform of one Dirac's delta peak at $t = t_0$ is the f-domain function

$$\mathcal{F}^{-1}\{\delta(t-t_0)\} = \int_{-\infty}^{\infty} \delta(t-t_0)e^{-i2\pi tf}\,\mathrm{d}t = e^{-i2\pi t_0 f}.$$

Since a power spectrum is not affected by a shift, which is easy to see by absolute squaring of the shift theorem in Equation 2.8, we can choose the origin freely. Let us define that $t = 0$ at the beginning of the gunfight. Let us denote $\tau = 0.2\ s$, and $T = 5\ s$. The spectrum of the 12 periods of six shots is

$$\begin{aligned} &\left[e^{-i2\pi 0f} + e^{-i2\pi \tau f} + \cdots + e^{-i2\pi 5\tau f}\right] + e^{-i2\pi Tf}\left[e^{-i2\pi 0f} + \cdots + e^{-i2\pi 5\tau f}\right] \\ &+ \cdots + e^{-i2\pi 11Tf}\left[e^{-i2\pi 0f} + \cdots + e^{-i2\pi 5\tau f}\right] \\ = \;& \sum_{n=0}^{11} e^{-i2\pi nTf} \sum_{m=0}^{5} e^{-i2\pi m\tau f} = S_1\, S_2. \end{aligned}$$

We can calculate the sums:

$$S_1 = \frac{1 - e^{-i2\pi 12Tf}}{1 - e^{-i2\pi Tf}} = e^{-i\pi 11Tf} \frac{\sin(12\pi Tf)}{\sin(\pi Tf)},$$

and

$$S_2 = \frac{1 - e^{-i2\pi 6\tau f}}{1 - e^{-i2\pi \tau f}} = e^{-i\pi 5\tau f} \frac{\sin(6\pi \tau f)}{\sin(\pi \tau f)}.$$

Consequently, the power spectrum is

$$|S_1|^2 |S_2|^2 = \left[\frac{\sin(12\pi Tf)}{\sin(\pi Tf)}\right]^2 \left[\frac{\sin(6\pi \tau f)}{\sin(\pi \tau f)}\right]^2.$$

Example 2.2: Compute the Fourier transform $\mathcal{F}\{G(f)\}$, where

$$G(f) = \begin{cases} \sin^2(\pi f/f_m), & |f| \le f_m, \\ 0, & |f| > f_m. \end{cases}$$

Solution. $G(f)$ is a $\sin^2$ function truncated by a boxcar function:

$$G(f) = \Pi_{2f_m}(f)\ \sin^2(\pi f/f_m).$$

We know that $\sin^2(A) = \frac{1}{2} - \frac{1}{2}\cos(2A)$. Consequently,

$$\begin{aligned} g(t) &= \mathcal{F}\{G(f)\} = \mathcal{F}\{\Pi_{2f_m}(f)\ \sin^2(\pi f/f_m)\} \\ &= \frac{1}{2}\mathcal{F}\{[1 - \cos(2\pi f/f_m)]\ \Pi_{2f_m}(f)\}. \end{aligned}$$

Applying the modulation theorem, and writing $h(t) = \mathcal{F}\{\Pi_{2f_m}(f)\} = 2f_m \operatorname{sinc}(2\pi f_m t)$, we obtain that

$$\begin{aligned} g(t) &= \frac{1}{2}h(t) - \frac{1}{4}h(t - \frac{1}{f_m}) - \frac{1}{4}h(t + \frac{1}{f_m}) \\ &= f_m \operatorname{sinc}(2\pi f_m t) - \frac{f_m}{2} \operatorname{sinc}\left[2\pi f_m (t - \frac{1}{f_m})\right] \\ &\quad - \frac{f_m}{2} \operatorname{sinc}\left[2\pi f_m (t + \frac{1}{f_m})\right]. \end{aligned}$$

Problems

1. Show that if a function f is symmetric about a point a and a function g is symmetric about a point b, then $f * g$ is symmetric about the point $a + b$.

2. Show that successive convolution and cross correlation have the following property:

$$(h * g) \circledast e = h \circledast (g \circledast e).$$

3. Show that autocorrelation function $h \circledast h$ is always conjugate symmetric, *i.e.*,

$$(h \circledast h)(-t) = [h(t) \circledast h(t)]^* ,$$

where $(h \circledast h)(-t)$ means the value of $h \circledast h$ at the point $-t$.

4. Compute the convolution function $\Pi_{2T}(t) * \Pi_{2T}(t)$, where $\Pi_{2T}(t)$ is the boxcar function

$$\Pi_{2T}(t) = \begin{cases} 1, & |t| \le T, \\ 0, & |t| > T. \end{cases}$$

Plot the result.

5. Prove that $\mathcal{F}\{H(f)G(f)\} = \mathcal{F}\{H(f)\} * \mathcal{F}\{G(f)\}$ (the convolution theorem).

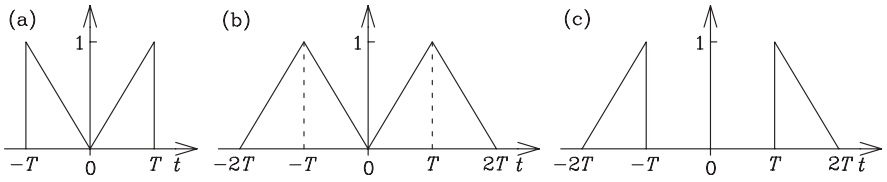
6. Show that the power spectrum of a real signal is always symmetric.

7. Formulate the modulation theorem for sine modulation, *i.e.*, determine the transform

$$\mathcal{F}^{-1}\{h(t)\sin(2\pi f_0 t)\}.$$

8. Let us assume that we know the convolution theorem $\mathcal{F}^{-1}\{h(t) * g(t)\} = H(f)G(f)$. Derive from it the correlation theorem $\mathcal{F}^{-1}\{h(t) \circledast g(t)\} = H^*(f)G(f)$.

9. Find the inverse Fourier transforms of the following functions:



10. Find the Fourier transform of the triangular function

$$\Lambda_T(t) = \begin{cases} 1 - \dfrac{|t|}{T}, & |t| \le T, \\ 0, & |t| > T. \end{cases}$$

Use the Fourier transform of a boxcar function and the convolution theorem. It is recommended that you solve Problem 4 first. (This triangular function is symmetric, and hence $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the same.)

11. Applying the derivative theorem, compute $\mathcal{F}^{-1}\left\{t \exp\left(-\pi t^2\right)\right\}$.

12. A signal consists of two cosine waves, which both have an amplitude A. The frequencies of the waves are f_1 and f_2. Derive the inverse Fourier transform of a differentiated signal,
    (a) by first differentiating the signal and then making the transform $\mathcal{F}^{-1}$,
    (b) by first transforming the original signal and then applying the derivative theorem.

13. Applying Fourier transforms, compute the following integrals:
    (a) $\int_{-\infty}^{\infty} \operatorname{sinc}^4 x \, \mathrm{d}x$,
    (b) $\int_{-\infty}^{\infty} \exp\left(-\pi x^2\right) \cos(2\pi a x) \, \mathrm{d}x$, where a is a real constant.

14. Applying Fourier transforms, compute the integral $\int_{-\infty}^{\infty} \frac{\sin^3 x}{x^3} \, \mathrm{d}x$.

    Hint: Use the power theorem.

15. Applying Fourier transforms, compute the integral $\int_{0}^{\infty} \frac{x^2 \, \mathrm{d}x}{(x^2 + a^2)^2}$, where a is a constant.

    Hint: Use the Parseval's theorem and the derivative theorem.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

3 Discrete Fourier transform

In practical measurements we do not deal with functions which are expressed as explicit mathematical expressions whose Fourier transforms are known. Instead, Fourier transforms are computed numerically. In practice, the measurement of a signal usually gives us a finite number of data, measured at discrete points. Consequently, also the integrals of Fourier transforms must be approximated by finite sums. The integral from $-\infty$ to $+\infty$ is replaced by a sum from $-N$ to $N-1$. A Fourier transform calculated in this way is called a *discrete Fourier transform*.

Calculation of a discrete Fourier transform is possible, if we record the signal $h(t)$ at $2N$ equally spaced sampling points

$$t_j = j\Delta t, \qquad j = -N, -N+1, -N+2, \ldots, -1, 0, 1, \ldots, N-1. \tag{3.1}$$

Generally, the recorded signal is a real function. If the signal is real and symmetric, then, according to Equation 2.2, also the spectrum $H(f)$ is real and symmetric, and

$$H(-f) = H(f). \tag{3.2}$$

The spectrum calculated from the discrete signal samples is given by a discrete approximation of $H(f)$ in Equation 2.2. The obtained spectrum is

$$H_T^{\Delta t}(f) = \Delta t \sum_{j=-N}^{N-1} h(j\Delta t) e^{-i2\pi f j \Delta t}, \tag{3.3}$$

where $T = N\Delta t$.

It is clear that a signal which consists of data at a finite number of discrete points cannot contain the same amount of information as an infinitely long, continuous signal. This, inevitably, leads to some distortions, compared to the true case. In the following, we shall examine these distortions.

3.1 Effect of truncation

The sum in Equation 3.3 covers only a finite segment $(-T, T)$ of the signal $h(t)$, unlike the Fourier integral, which extends from $-\infty$ to $+\infty$. Let us consider what is the effect of this approximation, the *truncation* of the signal, on the spectrum.

The truncated signal can be thought as the product of the original signal $h(t)$ and a truncation function, which is the *boxcar function* $\Pi_{2T}(t)$

$$\Pi_{2T}(t) = \begin{cases} 1, & |t| \le T, \\ 0, & |t| > T. \end{cases} \tag{3.4}$$

The boxcar function is illustrated in Figure 3.1.

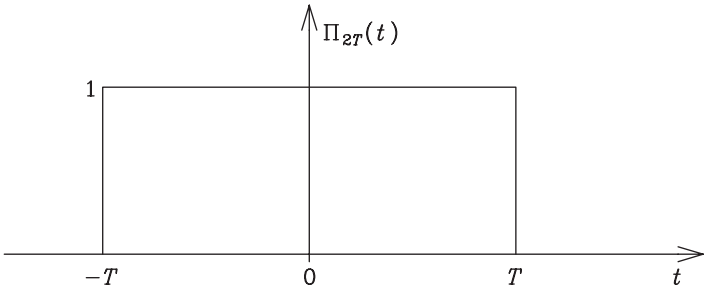


Figure 3.1: The boxcar signal truncation function $\Pi_{2T}(t)$.

The spectrum of the truncated signal is the convolution

$$\begin{aligned} H_T(f) &= \mathcal{F}^{-1}\{\Pi_{2T}(t)h(t)\} = \mathcal{F}^{-1}\{\Pi_{2T}(t)\} \overset{(2.17)}{*} \mathcal{F}^{-1}\{h(t)\} \\ &= \mathcal{F}^{-1}\{\Pi_{2T}(t)\} * H(f). \end{aligned} \tag{3.5}$$

The true spectrum $H(f)$ is convolved by the inverse Fourier transform of the truncation function. The inverse Fourier transform of $\Pi_{2T}(t)$ is

$$\mathcal{F}^{-1}\{\Pi_{2T}(t)\} = \int_{-T}^{T} e^{-i2\pi ft}\,\mathrm{d}t = 2T\,\mathrm{sinc}(2\pi fT) = W_T(f), \tag{3.6}$$

where the sinc function

$$\mathrm{sinc}\,x = \frac{\sin x}{x}. \tag{3.7}$$

(In the literature, one can sometimes find the definition $\mathrm{sinc}\,x = \sin(\pi x)/x$, but this definition would lead to impractical expressions in our case.) The function $W_T(f)$ is called the *instrumental function of truncation*. It is the sinc function shown in Figure 3.2. Thus, the *effect of truncation of the signal on the spectrum is that the true spectrum is convolved by a sinc function*:

$$H_T(f) = 2T\,\mathrm{sinc}(2\pi fT) * H(f). \tag{3.8}$$

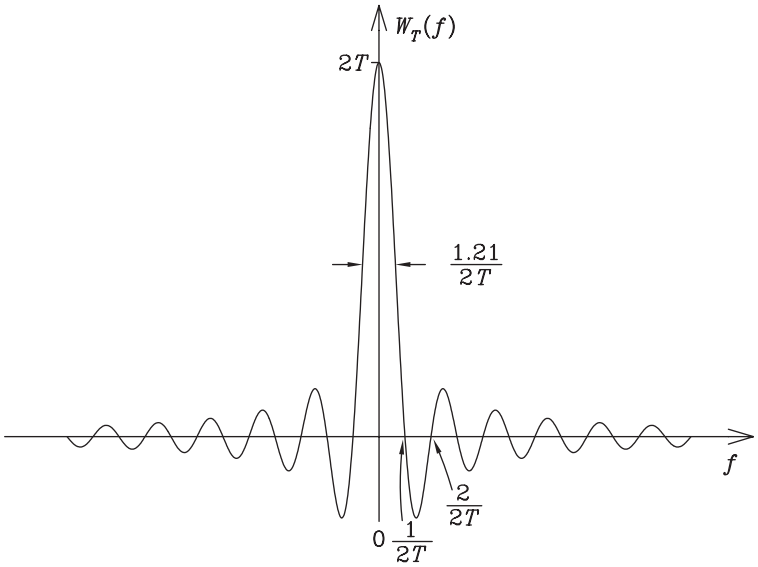


Figure 3.2: The inverse Fourier transform of the boxcar function $\Pi_{2T}(t)$, which is the instrumental function of truncation $W_T(f) = \mathcal{F}^{-1}\{\Pi_{2T}(t)\} = 2T\ \mathrm{sinc}(2\pi f T)$.

Example 3.1: Examine the cosine wave $h(t) = \cos(2\pi f_0 t)$, its inverse Fourier transform, and the effect of truncation of the wave.

Solution. With the help of the Euler's formula, and the linearity of Fourier transforms, we obtain that

$$\mathcal{F}^{-1}\{h(t)\} = \mathcal{F}^{-1}\{\cos(2\pi f_0 t)\} = \frac{1}{2}\mathcal{F}^{-1}\{e^{i2\pi f_0 t}\} + \frac{1}{2}\mathcal{F}^{-1}\{e^{-i2\pi f_0 t}\}. \tag{3.9}$$

According to Equation 1.14, $\mathcal{F}\{\delta(f)\} = 1$. This and the shift theorem (Equation 2.8) yield

$$\mathcal{F}\{\delta(f \pm f_0)\} = \mathcal{F}\{\delta(f)\}e^{\mp i2\pi f_0 t} = e^{\mp i2\pi f_0 t}. \tag{3.10}$$

From Equations 3.9 and 3.10 we see that the inverse Fourier transform of $h(t) = \cos(2\pi f_0 t)$ is $H(f) = \frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0)$. The cosine wave function and its inverse Fourier transform, in the absence of truncation, are shown in Figure 3.3. This same result would also be given by the modulation theorem, Equation 2.12, by setting $h(t) = 1$, and using the fact that $\mathcal{F}^{-1}\{1\} = \delta(f)$.

Generally, the convolution of $\delta(f \pm f_0)$ with another function is

$$\delta(f \pm f_0) * G(f) = \int_{-\infty}^{\infty} \delta(u \pm f_0)G(f - u)\,\mathrm{d}u = G(f \pm f_0). \tag{3.11}$$

Thus, the spectrum of a truncated cosine wave is

$$H_T(f) = 2T\,\mathrm{sinc}(2\pi f T) * H(f) = T\,\mathrm{sinc}\,[2\pi(f - f_0)T] + T\,\mathrm{sinc}\,[2\pi(f + f_0)T]\,. \tag{3.12}$$

This function, consisting of two sinc functions, is shown in Figure 3.4. The effect of the truncation of a cosine wave is that Dirac's delta functions of the spectrum are smeared to sinc functions.

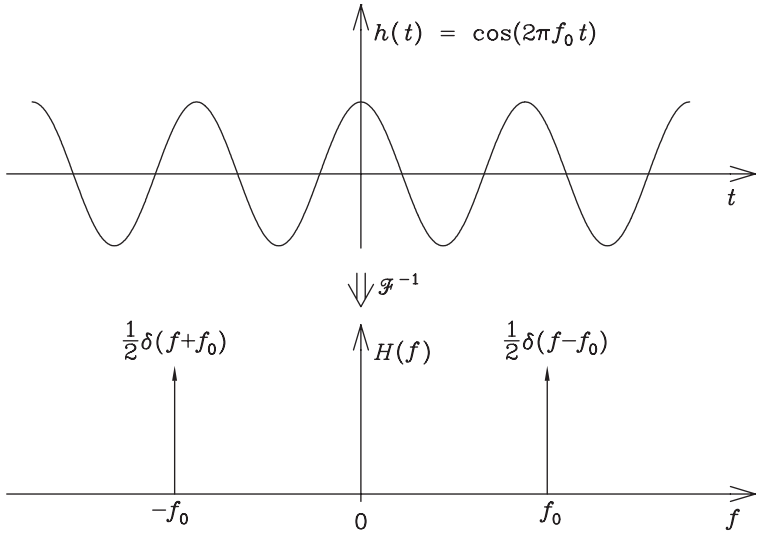


Figure 3.3: Infinitely long cosine wave $h(t) = \cos(2\pi f_0 t)$, and its inverse Fourier transform $H(f)$, which consists of two Dirac's delta functions $\frac{1}{2}\,\delta(f + f_0)$ and $\frac{1}{2}\,\delta(f - f_0)$.

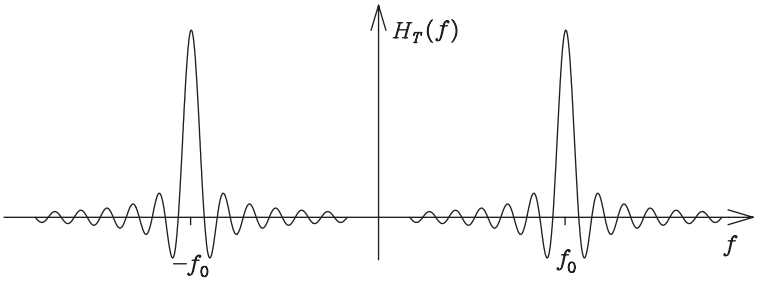


Figure 3.4: The spectrum of a truncated cosine wave $H_T(f) = 2T\,\mathrm{sinc}(2\pi f T) * H(f)$, which consists of two sinc functions $T\,\mathrm{sinc}\left[2\pi(f + f_0)T\right]$ and $T\,\mathrm{sinc}\left[2\pi(f - f_0)T\right]$.

3.2 Effect of sampling

In the calculation of the spectrum in Equation 3.3, the signal $h(j\Delta t)$ is not continuous but discrete. Let us now consider the effect of discretization on the spectrum.

If samples of the signal $h(t)$ are recorded at points separated by a sampling interval Δt, assuming no truncation of the signal, the calculated spectrum becomes

$$H^{\Delta t}(f) = \Delta t \sum_{j=-\infty}^{\infty} h(j\Delta t) e^{-i2\pi f j \Delta t}. \tag{3.13}$$

Clearly,

$$\exp\left[-i2\pi\left(f + \frac{1}{\Delta t}\right) j\Delta t\right] = \exp(-i2\pi f j \Delta t),$$

if j is an integer. Consequently, we can see that $H^{\Delta t}\left(f + \frac{1}{\Delta t}\right) = H^{\Delta t}(f)$. Thus, *the spectrum of a signal sampled with the sampling interval Δt is periodic with the period* $1/(\Delta t)$.

At every discrete sample point $t_j = j\Delta t$, the waves $e^{i2\pi f t}$ and $e^{i2\pi(f+n\frac{1}{\Delta t})t}$, where n is an integer, always obtain the same value. It is natural that these waves cannot be distinguished in the calculated spectrum, because there is no sufficient information. In the calculated spectrum, the contributions of all the waves at frequencies $f + n\frac{1}{\Delta t}$ are therefore superimposed. This phenomenon is demonstrated in Figure 3.5 for cosine waves.

The spectrum of a sampled signal can be written as

$$H^{\Delta t}(f) = \sum_{k=-\infty}^{\infty} H\left(f - \frac{k}{\Delta t}\right). \tag{3.14}$$

This means that the spectrum of a sampled signal consists of a set of functions $H(f)$, repeated at intervals $1/(\Delta t)$. The functions $H(f - \frac{k}{\Delta t})$ with various k are called the *spectral orders* of $h(t)$.

If the signal, in addition to being discrete, is also truncated, we obtain that

$$\begin{aligned} H_T^{\Delta t}(f) &= \sum_{k=-\infty}^{\infty} H_T\left(f - \frac{k}{\Delta t}\right) \overset{(3.8)}{\underset{(3.11)}{=}} \sum_{k=-\infty}^{\infty} 2T \operatorname{sinc}\left[2\pi\left(f - \frac{k}{\Delta t}\right) T\right] * H(f) \\ &= W_T^{\Delta t}(f) * H(f), \end{aligned} \tag{3.15}$$

where $W_T^{\Delta t}$ is the *instrumental function of discrete (numerical) transform*

$$\boxed{W_T^{\Delta t}(f) = \sum_{k=-\infty}^{\infty} 2T \operatorname{sinc}\left[2\pi\left(f - \frac{k}{\Delta t}\right) T\right].} \tag{3.16}$$

The effects of truncation and sampling on a continuous spectrum are demonstrated in Figure 3.6. In the figure, it is assumed that the true continuous spectrum is limited to a

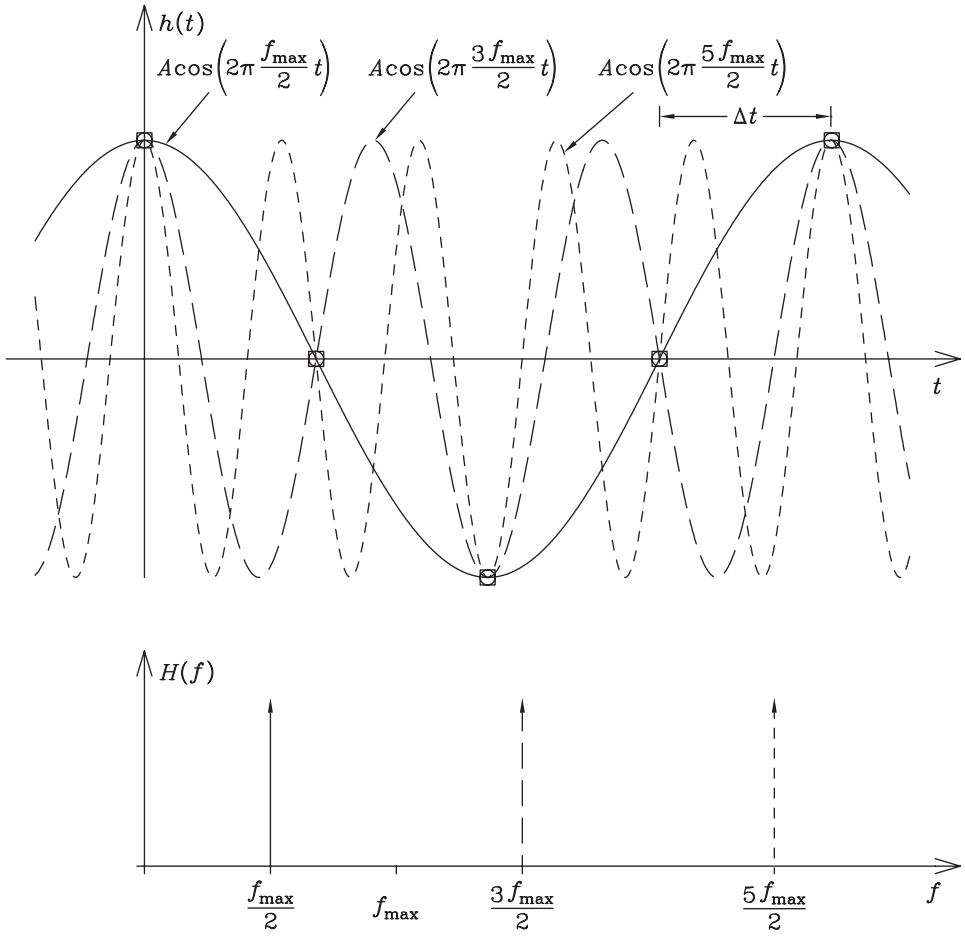


Figure 3.5: A signal $h(t)$ consists of cosine waves, whose spectrum is $H(f)$. The signal is sampled with the sampling interval $1/(2f_{\max})$, which is one quarter of the period of the lowest frequency $f_{\max}/2$. Cosine waves shown in the figure cannot be distinguished at the sample points, and in the calculated discrete spectrum they will all be superimposed.

band $[-f_{\max},\ f_{\max}]$. The corresponding signal $h(t) = \mathcal{F}\{H(f)\}$ is truncated with a boxcar function at the point $T(\gg 1/f_{\max})$. The convolution of the instrumental function of truncation and the true spectrum is $W_T(f)*H(f) = 2T\ \mathrm{sinc}(2\pi fT)*H(f)$. Since the truncation boxcar is large ($T \gg 1/f_{\max}$), the instrumental function $W_T(f)$ is narrow (see Figure 3.2), and the spectrum is barely distorted by truncation. Samples of the signal are taken in the interval $\Delta t(< 1/2f_{\max})$. Because of the sampling, the function $W_T(f) * H(f)$ is repeated in the intervals $1/(\Delta t)$. The spectrum of the truncated, discrete signal is shown in the lower part of Figure 3.6.

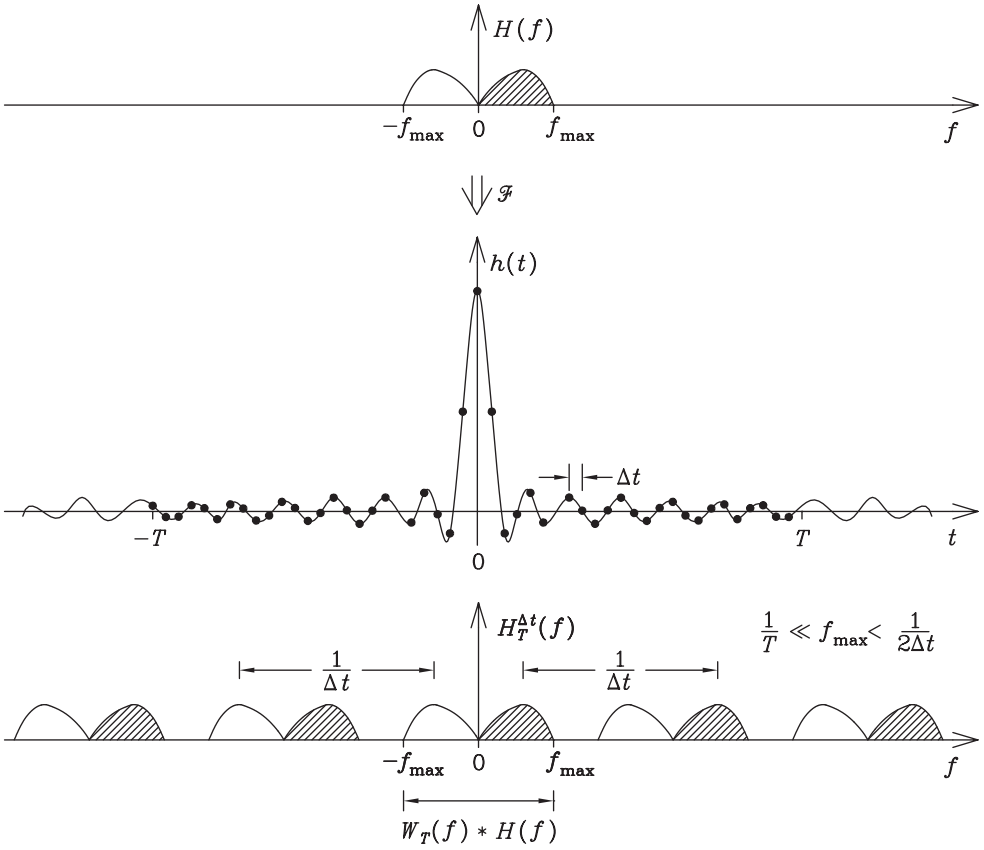


Figure 3.6: The true continuous spectrum $H(f)$ and the corresponding signal $h(t) = \mathcal{F}\{H(f)\}$. The signal is truncated with a boxcar function at the point T, and samples of the signal are taken in the interval Δt. The lower part of the picture shows the spectrum $H_T^{\Delta t}(f)$ of the truncated, discrete signal.

The *critical sampling interval* is

$$\boxed{\Delta t = \frac{1}{2 f_{\max}},} \tag{3.17}$$

where $f_{\max}$ is the maximum frequency of the true spectrum. If the sampling interval is smaller than this critical interval, then the spectral orders of the sampled signal do not overlap, and the spectrum is of the type shown in Figure 3.6, where the repeating functions $W_T(f) * H(f)$ are clearly distinct. The spectrum $H_T^{\Delta t}(f)$ of a signal which is sampled at exactly the critical sampling interval is shown in Figure 3.7. In this case, the period of the spectrum $H_T^{\Delta t}(f)$ is exactly equal to the bandwidth $2 f_{\max}$ of the true spectrum.

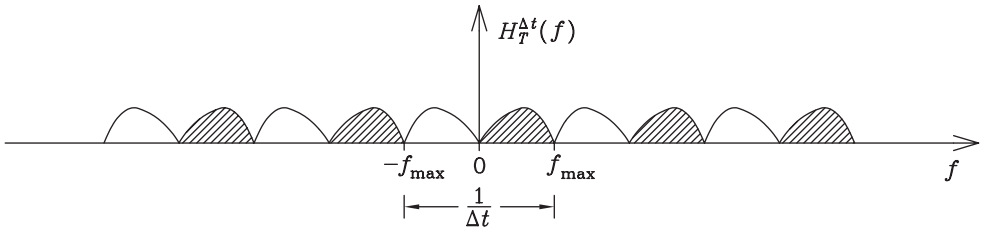


Figure 3.7: The spectrum $H_T^{\Delta t}$ of a truncated, discrete signal, when the sampling interval is exactly the critical sampling interval, $\Delta t = 1/(2f_{\max})$.

The sampling is called undersampling, if the sampling interval $\Delta t > 1/(2f_{\max})$. In this case, the period $1/(\Delta t)$ of the spectrum $H_T^{\Delta t}(f)$ is smaller than the bandwidth $2f_{\max}$ of the true spectrum $H(f)$. Then, the portion of the spectrum with $|f| > 1/(2\Delta t)$ is *aliased* into the basic period $[-1/(2\Delta t), 1/(2\Delta t)]$ and thus overlaps with the spectral information originally located in this interval. Thereby, a distorted spectrum $H_T^{\Delta t}(f)$ is obtained. Figure 3.8 demonstrates the spectrum in the case where the sampling interval has been larger than the critical sampling interval.

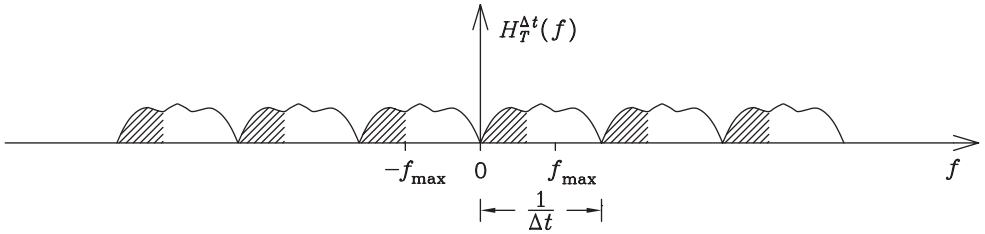


Figure 3.8: The spectrum $H_T^{\Delta t}$ which is distorted by aliasing because the sampling interval is larger than the critical sampling interval, *i.e.*, $\Delta t > 1/(2f_{\max})$.

The *critical sampling frequency*

$$\boxed{f_N = \frac{1}{\Delta t} = 2f_{\max}} \tag{3.18}$$

is called the *Nyquist frequency*.

Example 3.2: Examine the effect of discrete sampling on the inverse Fourier transform of the truncated cosine wave.

Solution. The inverse Fourier transform of a continuous cosine wave was shown in Figure 3.3. It is the monochromatic spectrum $\mathcal{F}^{-1}\{\cos(2\pi f_0 t)\} = \frac{1}{2}\delta(f + f_0) + \frac{1}{2}\delta(f - f_0) = H(f)$. The effect of truncation of the cosine wave is that Dirac's delta functions of the spectrum are

replaced by sinc functions, as shown in Figure 3.4. The effect of a sampling interval Δt of the cosine wave is that the sinc functions of the spectrum are repeated in the intervals $1/(\Delta t)$. The spectrum of the truncated, discrete cosine wave is shown in Figure 3.9.

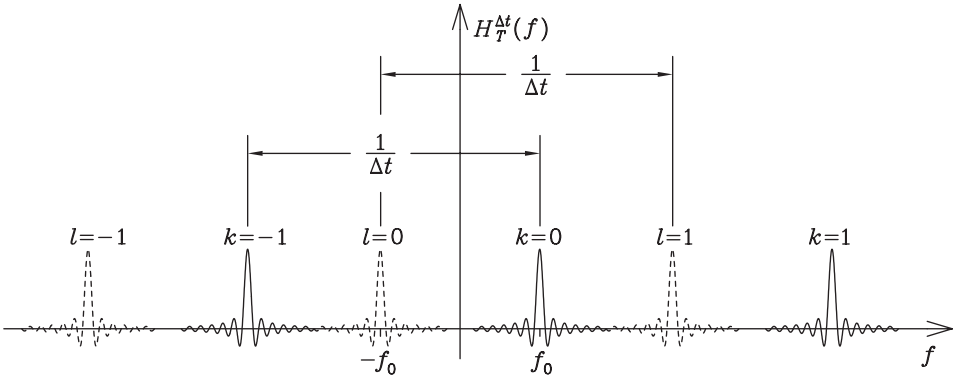


Figure 3.9: The spectrum $H_T^{\Delta t}(f) = W_T^{\Delta t}(f) * \frac{1}{2}[\delta(f + f_0) + \delta(f - f_0)] = \sum_{k=-\infty}^{\infty} T \operatorname{sinc}\left[2\pi\left(f - f_0 - \frac{k}{\Delta t}\right)T\right] + \sum_{l=-\infty}^{\infty} T \operatorname{sinc}\left[2\pi\left(f + f_0 - \frac{l}{\Delta t}\right)T\right]$ of the truncated, discrete (sampling interval Δt) cosine wave $\cos(2\pi f_0 t)$.

3.3 Discrete spectrum

The spectrum $H_T^{\Delta t}(f)$ is a continuous function of f, and it can be computed at any value of f. In practice, however, it is computed only at a finite number of points. If the signal consists of $2N$ signal data, it is sufficient to calculate $2N$ spectral data in order to preserve all information that we have. Computation of more spectral data is pure interpolation. In practice, not only the signal is discrete but also the spectrum is discrete.

If the spectrum is calculated at the points

$$f_k = \pm k\Delta f, \tag{3.19}$$

where k is an integer, then the corresponding signal $h(t) = \mathcal{F}\{H(f)\}$ is periodic with the period $1/(\Delta f)$. This is analogous to what was discussed above, only now the meanings of $h(t)$ and $H(f)$ have been changed. Analogously to Equation 3.17, we can find the *critical computing interval*

$$\Delta f = \frac{1}{2T} \tag{3.20}$$

for the spectrum of a signal whose period is $2T$. The critical computing interval is the minimum computing interval which must be used in the calculation of the spectrum in such

a way that all information in the signal segment $(-T, T)$ is preserved. This situation is illustrated in Figure 3.10. The number of data points in the period $2f_{\max}$ (critical sampling) of the spectrum or in the period $2T$ of the signal is

$$2f_{\max}/(\Delta f) = 2f_{\max}2T = 2T/(\Delta t) = 2N, \tag{3.21}$$

which ensures the conservation of the information.

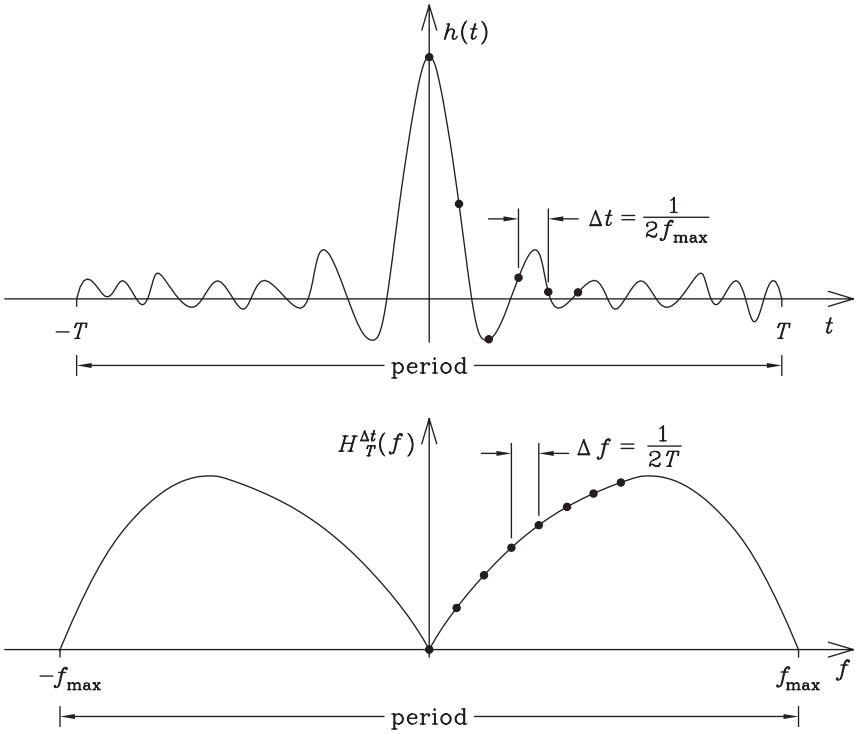


Figure 3.10: One period of the signal $h(t)$ and one period of its spectrum $H_T^{\Delta t}(f)$ in the case where all information is preserved in the discrete Fourier transforms. Both the signal and the spectrum consist of $2N$ data, $T = N\Delta t$, $\Delta t = 1/(2f_{\max})$ and $\Delta f = 1/(2T)$. In this case $f_{\max}/(\Delta f) = 2Tf_{\max} = 2T/(2\Delta t) = N$. (Compare Fig. 4.1.)

If the computing interval $\Delta f > 1/(2T)$, some information is lost, and if $\Delta f < 1/(2T)$, an interpolation is made without increasing information.

Since the signal in discrete Fourier transform consists of discrete data, it can be regarded as a signal vector $\boldsymbol{h}$, whose length is the number of sampled data points. In the same way, the spectrum which is calculated at discrete points can be treated as a vector $\boldsymbol{H}$, whose length is the number of computed data. $\boldsymbol{H}$ is the discrete inverse Fourier transform of $\boldsymbol{h}$.

Example 3.3: Show that the discrete inverse Fourier transform of the discrete convolution

$$h'_k = \Delta t \sum_{j=0}^{M-1} g_j h_{k-j}$$

is given by

$$H'_p = G_p H_p,$$

where the vectors $\boldsymbol{H}'$ (consisting of data H'_p), $\boldsymbol{H}$ (consisting of data H_p), and $\boldsymbol{G}$ (consisting of data G_p), are the discrete inverse Fourier transforms of the vectors $\boldsymbol{h}'$ (consisting of data h'_k), $\boldsymbol{h}$ (consisting of data h_k), and $\boldsymbol{g}$ (consisting of data g_k), respectively. M is the number of data in each vector. The vectors are regarded as infinitely long vectors with the M data repeating periodically.

Solution. Let us compute the discrete inverse Fourier transform of $\boldsymbol{h}'$:

$$\begin{aligned}
H'_p &= \Delta t \sum_{k=0}^{M-1} h'_k e^{-i2\pi pk/M} \\
&= \Delta t^2 \sum_{k=0}^{M-1} \sum_{j=0}^{M-1} g_j h_{k-j} e^{-i2\pi pk/M} \\
&\overset{r=k-j}{=} \Delta t^2 \sum_{j=0}^{M-1} \sum_{r=-j}^{M-j-1} g_j h_r e^{-i2\pi pr/M} e^{-i2\pi pj/M} \\
&\overset{\text{(periodicity)}}{=} \Delta t \sum_{j=0}^{M-1} g_j e^{-i2\pi pj/M} \Delta t \sum_{r=0}^{M-1} h_r e^{-i2\pi pr/M} \\
&= G_p H_p.
\end{aligned}$$

This result is the convolution theorem for the discrete Fourier transform.

Example 3.4: Due to limitations of the central memory of the computer, the signal vector $\boldsymbol{h}$ (number of data $2N$) must be divided into four partial vectors before inverse Fourier transformation. These partial vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ are each transformed separately. Each of the partial vectors consists of every fourth piece of the original data, $\boldsymbol{a}$ starting from h_0, $\boldsymbol{b}$ starting from h_1, $\boldsymbol{c}$ starting from h_2, and $\boldsymbol{d}$ starting from h_3. How is the discrete inverse Fourier transform of the original signal vector $\boldsymbol{h}$ obtained from the transforms of the partial vectors?

Solution. Let us denote $w_M = e^{i2\pi/M}$. The discrete inverse Fourier transforms of the partial vectors are ($k = 0, \ldots, \frac{N}{2} - 1$):

$$A_k = \Delta t \sum_{j=0}^{\frac{N}{2}-1} a_j w_{\frac{N}{2}}^{-jk} = \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j} w_{\frac{N}{2}}^{-jk},$$

$$B_k = \Delta t \sum_{j=0}^{\frac{N}{2}-1} b_j w_{\frac{N}{2}}^{-jk} = \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+1} w_{\frac{N}{2}}^{-jk},$$

$$C_k = \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+2} w_{\frac{N}{2}}^{-jk}, \text{ and}$$

$$D_k = \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+3} w_{\frac{N}{2}}^{-jk}.$$

The discrete inverse Fourier transforms of the original signal vector is:

$$\begin{aligned} H_k &= \Delta t \sum_{j=0}^{2N-1} h_j w_{2N}^{-jk} \\ &= \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j} w_{2N}^{-4jk} + w_{2N}^{-k} \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+1} w_{2N}^{-4jk} \\ &\quad + w_{2N}^{-2k} \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+2} w_{2N}^{-4jk} + w_{2N}^{-3k} \Delta t \sum_{j=0}^{\frac{N}{2}-1} h_{4j+3} w_{2N}^{-4jk}. \end{aligned}$$

Since $w_{2N}^{-4jk} = w_{\frac{N}{2}}^{-jk}$, we obtain that

$$H_k = \begin{cases} A_k + w_{2N}^{-k} B_k + w_{2N}^{-2k} C_k + w_{2N}^{-3k} D_k, & k = 0, \ldots, \frac{N}{2} - 1, \\ A_{k-\frac{N}{2}} + w_{2N}^{-k} B_{k-\frac{N}{2}} + w_{2N}^{-2k} C_{k-\frac{N}{2}} + w_{2N}^{-3k} D_{k-\frac{N}{2}}, & k = \frac{N}{2}, \ldots, N - 1, \\ A_{k-N} + w_{2N}^{-k} B_{k-N} + w_{2N}^{-2k} C_{k-N} + w_{2N}^{-3k} D_{k-N}, & k = N, \ldots, \frac{3N}{2} - 1, \\ A_{k-\frac{3N}{2}} + w_{2N}^{-k} B_{k-\frac{3N}{2}} + w_{2N}^{-2k} C_{k-\frac{3N}{2}} + w_{2N}^{-3k} D_{k-\frac{3N}{2}}, & k = \frac{3N}{2}, \ldots, 2N - 1. \end{cases}$$

Note that since the period of the vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ is $N/2$ data, indices that are too large have been reduced by subtracting suitable multiples of $N/2$.

Example 3.5: Find the optimal computing interval for the spectrum of the truncated cosine wave.

Solution. If the cosine wave is truncated at the point $t = T$, then the spectrum consists of the sinc functions of Figure 3.4. According to Equation 3.20, the critical computing interval of the spectrum is $\Delta f = \frac{1}{2T}$. As shown in Figure 3.11, this computing interval gives the data points for the sinc function in an optimal way.

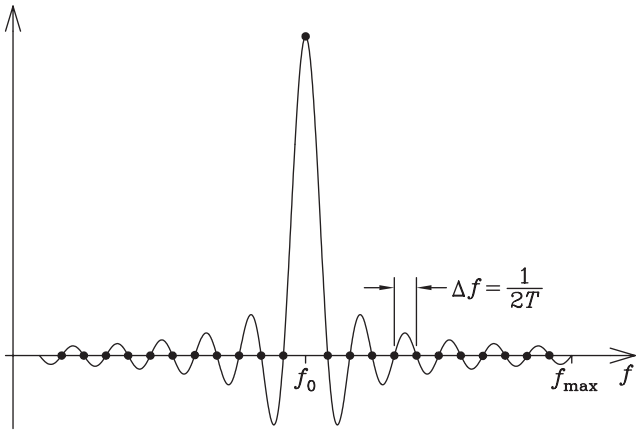
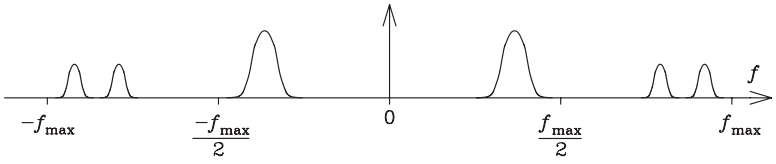


Figure 3.11: Optimally chosen data points for the sinc function. This means that the computing interval is $\Delta f = 1/(2T)$. Here f_0 is an integer multiple of Δf, *i.e.*, the sinc function is situated exactly at one of the spectral sampling points.

Problems

1. The inverse Fourier transformation of a signal gave the following symmetric spectrum. Outline the plot of the spectrum which is obtained if only every second datum of the signal is picked out, and these data, whose number is only half of the original, are transformed.



2. The detector takes samples of the signal at constant intervals at the points $k\Delta t$, where k is an integer. Let $f_{\max} = 1/(2\Delta t)$. f_0 is a frequency in the interval $-f_{\max} \leq f_0 < f_{\max}$.

   (a) Show that the frequency $f_0 + p2f_{\max}$, where p is an integer, is aliased at f_0.

   (b) Let us assume that we have an integrating detector and the kth sample is obtained by integrating the signal from $k\Delta t$ to $(k+1)\Delta t$. Show that also in this case the frequency $f_0 + p2f_{\max}$ is aliased at f_0, but it has been damped, the more the larger $|p|$ is.

   Hint: Consider exponential waves $e^{i2\pi ft}$.

3. The spectrum $G(f)$ of a time-dependent signal $g(t)$ is band-limited in such a way that $G(f) = 0$ at $|f| \geq f_0$.

   (a) Show that if $F \geq f_0$, then $G = \Pi_{2F}\left(\frac{1}{2F}\,\text{Ш}_{2F} * G\right)$, where

   $$\Pi_{2F}(f) = \begin{cases} 1, & |f| \leq F, \\ 0, & |f| > F, \end{cases}$$

   and $\text{Ш}_{2F}(f)$ is the comb function of period $2F$, *i.e.*,

   $$\text{Ш}_{2F}(f) = \sum_{j=-\infty}^{\infty} 2F\,\delta(f - j2F).$$

   (b) Applying the result of (a), show that if the function $g(t)$ is known at points $k\Delta t$, where k is an arbitrary integer and $\Delta t \leq 1/(2f_0)$, then these samples define the value of the function $g(t)$ everywhere. (This result is called the *sampling theorem*.)

   Hint: $\mathcal{F}\{\text{Ш}_{2F}\} = \mathcal{F}^{-1}\{\text{Ш}_{2F}\} = 2F\,\text{Ш}_{1/(2F)}$.

4. Radio Turku broadcasts at the frequency 94.3 MHz and Radio Hundred broadcasts at the frequency 100.1 MHz. Find the three smallest sampling intervals Δt which alias these stations on each other.

5. The discrete inverse Fourier transform converts the vector $\boldsymbol{h}$ of length M, consisting of data h_j, to the vector $\boldsymbol{H}$ of length M, consisting of data H_k, where

   $$H_k = \Delta t \sum_{j=0}^{M-1} h_j w_M^{-kj}, \qquad k = 0, \ldots, M-1,$$

   and $w_M = e^{i2\pi/M}$. Since this transform is linear, it corresponds to multiplication of the vector by a matrix. Formulate the matrices of the $\mathcal{F}$ and $\mathcal{F}^{-1}$ transforms, and show that the product of these matrices is the identity matrix.

6. Let us define the norm of the sample vector of a complex function in t-domain as

   $$\|\boldsymbol{C}\| = \sqrt{\Delta t\, \boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}},$$

   where superscript $^{\mathrm{H}}$ means conjugate transposition. Prove the following, discrete form of the Parseval's theorem:

   Let $\boldsymbol{H} = \Delta t \boldsymbol{W}^{-}\boldsymbol{h}$, where $\Delta t \boldsymbol{W}^{-}$ is the matrix of the discrete inverse Fourier transform ($\mathcal{F}^{-1}$ transform, see the previous problem). Then $\|\boldsymbol{H}\|^2 = \|\boldsymbol{h}\|^2$.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

4 Fast Fourier transform (FFT)

4.1 Basis of FFT

The number of sampled values available for the computation of a discrete Fourier transform is often very large. The computation of the transform is an extremely laborious task, if a direct method of computation is employed. The computation is significantly easier, if it is simplified by a suitable algorithm. The most practical algorithm for this purpose is the *fast Fourier transform*, FFT. In the following, we shall examine this algorithm, starting from the basic definition of the Fourier transform pair.

The Fourier transform pair $H(f)$ and $h(t)$ was defined as

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{i2\pi f t}\, \mathrm{d}f = \mathcal{F}\{H(f)\}, \tag{4.1}$$

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi f t}\, \mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}, \tag{4.2}$$

where $H(f)$ is the spectrum and $h(t)$ is the signal. For the computation of the discrete transform, the signal data h_j were assumed to be recorded in equal intervals at points

$$t_j = j\Delta t, \qquad j = -N, -N+1, -N+2, \ldots, -1, 0, 1, \ldots, N-1, \tag{4.3}$$

where j and N are integers. If $T = N\Delta t$, the spectrum, computed from the data h_j, is given by Equation 3.15:

$$H_T^{\Delta t}(f) = \sum_{k=-\infty}^{\infty} 2T \operatorname{sinc}\left[2\pi\left(f - \frac{k}{\Delta t}\right)T\right] * H(f). \tag{4.4}$$

The spectrum $H_T^{\Delta t}(f)$ is periodic with the period $1/(\Delta t)$, and smoothed by the sinc function so that the fine structure smaller than $1/(2T)$ is partly lost.

All necessary information must be included in one period of the spectrum $H_T^{\Delta t}(f)$. For this purpose, we must, as was explained in the previous section, use at least the critical sampling frequency of the signal

$$\frac{1}{\Delta t} = 2f_{\max}, \tag{4.5}$$

where $2f_{\max}$ is the Nyquist frequency. If also the spectrum $H(f)$ is computed in equal intervals, at points

$$f_k = \pm k\Delta f, \tag{4.6}$$

where k is an integer, then the signal $h(t)$ must also be periodic, its period being $1/(\Delta f)$. The $2N$ data points of Equation 4.3 cover exactly one period in the signal domain (t-domain), if

$$\frac{1}{\Delta f} = 2T, \tag{4.7}$$

that is,

$$\Delta f = \frac{1}{2T}. \tag{4.8}$$

The number of data points in one period of the spectrum $H_T^{\Delta t}(f)$ is

$$\frac{2f_{\max}}{\Delta f} = 2T \underbrace{2f_{\max}}_{1/(\Delta t)} = 2T\,\frac{1}{\Delta t} = 2N, \tag{4.9}$$

which is the same number as the number of data points in one period of the signal.

The situation where the discrete transform is used in such a way that both the signal and the spectrum are known at $2N$ points in equal intervals is illustrated in Figure 4.1. (The true spectrum $H(f)$ or the true signal $h(t)$ may coincide with $H_T^{\Delta t}(f)$ or $h_{f_{\max}}^{\Delta f}(t)$, respectively, in any period.)

If the signal is known at $2N$ points in the intervals $\Delta t = 1/(2f_{\max})$ and the spectrum at $2N$ points in the intervals $\Delta f = 1/(2T)$, then the Fourier transform pair of Equations 4.1 and 4.2 can be written in the discrete form (we include in the sums exactly one period, if we take the summation from $-N$ to $N-1$, *not* to N)

$$h_j = \sum_{k=-N}^{N-1} H_k e^{i\pi jk/N}\,\Delta f, \tag{4.10}$$

$$H_k = \sum_{j=-N}^{N-1} h_j e^{-i\pi jk/N}\,\Delta t, \tag{4.11}$$

where

$$\begin{cases} h_j = h_{f_{\max}}^{\Delta f}(j\Delta t), & j = -N, \ldots, N-1, \\ H_k = H_T^{\Delta t}(k\Delta f), & k = -N, \ldots, N-1, \end{cases} \tag{4.12}$$

and the exponent has been rewritten, since

$$2\pi f t = 2\pi \underbrace{k\Delta f}_{f}\,\underbrace{j\Delta t}_{t} = 2\pi kj\,\frac{\Delta t}{2T} = \pi jk/N.$$

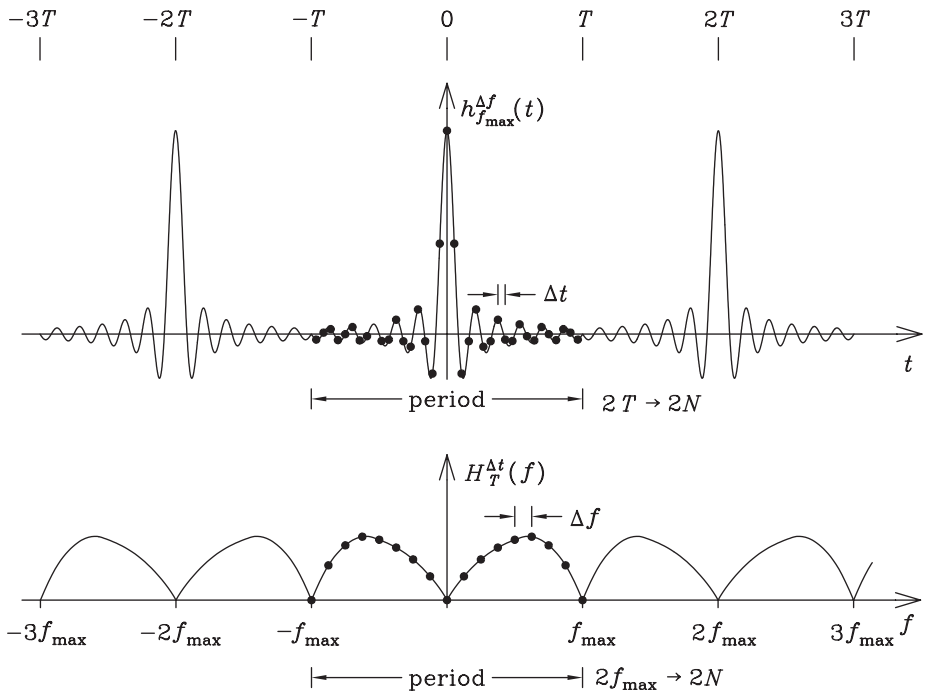


Figure 4.1: A discrete Fourier transform pair where the signal $h^{\Delta f}_{f_{\max}}(t)$ and the spectrum $H^{\Delta t}_{T}(f)$ are both known in equal intervals. In the figure $\Delta t = 1/(2f_{\max})$ and $\Delta f = 1/(2T)$.

We can write the discrete spectral values

$$\begin{aligned} H_k &= \Delta t \sum_{j=-N}^{N-1} h_j e^{-i\pi jk/N} = \Delta t \sum_{j=-N}^{-1} h_j e^{-i\pi jk/N} + \Delta t \sum_{j=0}^{N-1} h_j e^{-i\pi jk/N} \\ &= \Delta t \sum_{j=-N+2N}^{2N-1} h_j e^{-i\pi jk/N} + \Delta t \sum_{j=0}^{N-1} h_j e^{-i\pi jk/N} \\ &= \Delta t \sum_{j=0}^{2N-1} h_j e^{-i\pi jk/N}. \end{aligned} \tag{4.13}$$

This means that if the signal is extended periodically, so that always $h_j = h_{j+2N}$, then the sampling of the signal can be shifted half the period, N, as many times as we wish. Actually, the start of the sampling of the signal could be shifted anywhere, if the length of the sampling is the same as the period of the signal. In order to avoid negative indices, the sampling is often started with $j = 0$ as in the last form of Equation 4.13. This possibility of shifting the sampling is illustrated in Figure 4.2. Analogously, the same possibility of shifting half the period applies to the spectrum.

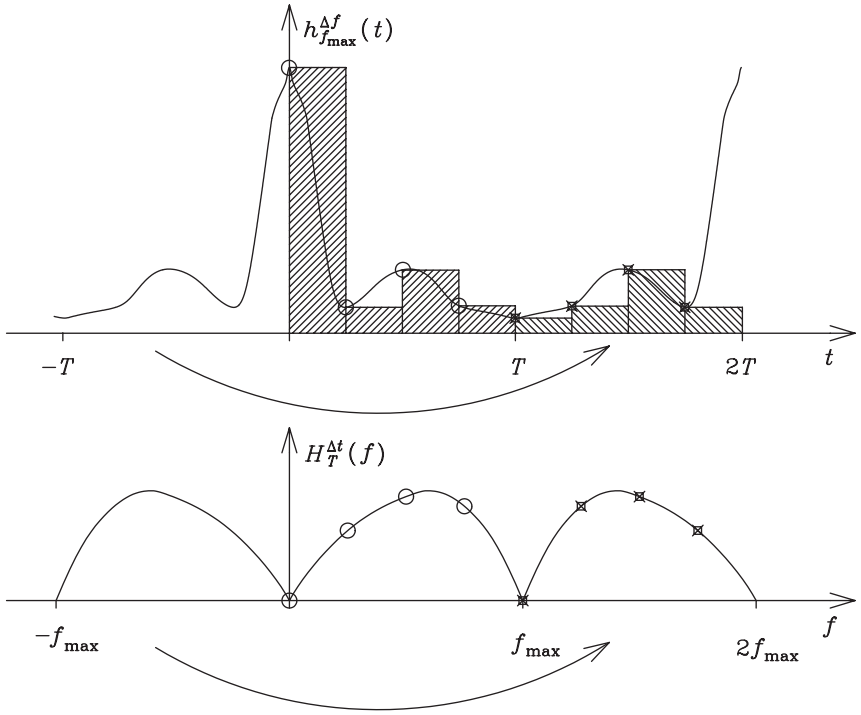


Figure 4.2: The sampling of the signal $h^{\Delta f}_{f_{\max}}(t)$ can be shifted half of the period. The same applies to the spectrum $H^{\Delta t}_{T}(f)$. Note that the sums in Equations 4.10 and 4.11 were calculated from $-N$ to $N-1$, *not* to N, in order to not include the columns *behind* points $t = 2T$ or $f = 2f_{\max}$.

Let us now divide the signal data h_j into two parts, one with even j, and the other with odd j,

$$\left.\begin{array}{rcl} x_l &=& h_{2l}, \\ y_l &=& h_{2l+1}, \end{array}\right\} \qquad l = 0, 1, 2, \ldots, N-1. \tag{4.14}$$

We shall denote

$$\left.\begin{array}{rcccl} A_k &=& \mathcal{F}^{-1}\{x_l\} &=& \Delta t \displaystyle\sum_{l=0}^{N-1} x_l e^{-i2\pi kl/N}, \\ B_k &=& \mathcal{F}^{-1}\{y_l\} &=& \Delta t \displaystyle\sum_{l=0}^{N-1} y_l e^{-i2\pi kl/N}, \end{array}\right\} \quad k = 0, 1, 2, \ldots, N-1. \tag{4.15}$$

A_k and B_k are N-transforms, that is, they are sums of N components.

Now we can write the discrete spectrum, which is a $2N$-transform, as (compare with Example 3.4)

$$\begin{aligned} H_k &= \Delta t \sum_{j=0}^{2N-1} h_j e^{-i\pi kj/N} = \Delta t \sum_{l=0}^{N-1} \underbrace{x_l}_{h_{2l}} e^{-i\pi k(2l)/N} + \Delta t \sum_{l=0}^{N-1} \underbrace{y_l}_{h_{2l+1}} e^{-i\pi k(2l+1)/N} \\ &= A_k + e^{-i\pi k/N} B_k. \end{aligned} \tag{4.16}$$

We can see that the $2N$-transform can be expressed with the help of two N-transforms.

Very important and useful is, that after we have computed A_k and B_k to obtain H_k until the index $k = N-1$, we can obtain the remaining values of H_k at higher indices using the same, already calculated values A_k and B_k, because

$$\begin{aligned} H_{k+N} &= \Delta t \sum_{l=0}^{N-1} x_l \underbrace{e^{-i2\pi l(k+N)/N}}_{e^{-i2\pi lk/N}\underbrace{e^{-i2l\pi}}_{1}} + \Delta t \sum_{l=0}^{N-1} y_l \underbrace{e^{-i\pi(2l+1)(k+N)/N}}_{e^{-i2\pi kl/N}e^{-i\pi k/N}\underbrace{e^{-i(2l+1)\pi}}_{-1}} \\ &= A_k - e^{-i\pi k/N} B_k. \end{aligned} \tag{4.17}$$

This means that the computation of the values of the two N-transforms A_k and B_k actually gives us two values H_k and H_{k+N} of the $2N$-transform, which saves a lot of work, since calculation of an N-transform is much less laborious than calculation of a $2N$-transform.

If we denote $w_N^k = e^{i2\pi k/N}$, then

$$\left\{ \begin{array}{rcl} w_N^{-(k\pm N)} &=& w_N^{-k}, \\ w_N^{-[k\pm(N/2)]} &=& -w_N^{-k}, \end{array} \right. \tag{4.18}$$

and we obtain

$$\boxed{\left. \begin{array}{rcl} H_k &=& A_k + w_{2N}^{-k} B_k, \\ H_{k+N} &=& A_k - w_{2N}^{-k} B_k \\ &=& A_k + w_{2N}^{-(k+N)} B_k, \end{array} \right\} \quad k = 0, 1, 2, \ldots, N-1.} \tag{4.19}$$

These formulae are the basis of the fast Fourier transform. With them we can obtain a $2N$-transform by the calculation of two N-transforms.

4.2 Cooley–Tukey algorithm

If $2N = 2^m$, where m is a positive integer, then we can repeat the division of the data into two parts (Equation 4.19) m times. Finally, we arrive in the situation where we need to calculate the Fourier transform of only one single datum, and it is the datum itself. With the help of Equation 4.19 we compute N times the Fourier transform of two data, using those $2N$ Fourier transforms of one datum. Next, we compute $N/2$ Fourier transforms of four data using the N Fourier transforms of two data, then the $N/4$ Fourier transforms of eight data, and so on until we have one Fourier transform of $2N$ data. This method is called the *Cooley–Tukey algorithm*. The calculation of a $2N$-transform, where $2N = 2^m$, takes place in m successive steps, using Equation 4.19 m times.

Let us, as an example, examine the case, where $2N = 2^3 = 8$. In this case $m = 3$. The data h_j must then be divided to two groups, even and odd, $m = 3$ times. This is shown in Table 4.1.

Table 4.1: The grouping of the data h_j for FFT. If the number of data is 8, then the division to two groups must be done three times.

Memory cell	Data	1. phase	2. phase	3. phase
0	h_0	h_0	h_0	h_0
1	h_1	h_2	h_4	h_4
2	h_2	h_4	h_2	h_2
3	h_3	h_6	h_6	h_6
4	h_4	h_1	h_1	h_1
5	h_5	h_3	h_5	h_5
6	h_6	h_5	h_3	h_3
7	h_7	h_7	h_7	h_7

An easy way to make this grouping of the data by computer is to use *binary inversion*. This is illustrated in Table 4.2 in the case $m = 3$. If the binary number of a memory cell is inverted, the new binary number is the same as the index j of the datum h_j which should be positioned in this memory cell. The binary inversion may be used, if $2N = 2^m$, where m is an integer.

After we have divided the data h_j into groups in the manner explained, we are able to employ Equation 4.19 m times. This procedure in the case where $m = 3$ is illustrated by the net in Figure 4.3.

Table 4.2: Binary inversion in the case where the number of data is 8 and the data are positioned in the memory cells $0, \ldots, 7$.

Memory cell	Bin	Inversion	Bin	Data
0	000	→	000	h_0
1	001	→	100	h_4
2	010	→	010	h_2
3	011	→	110	h_6
4	100	→	001	h_1
5	101	→	101	h_5
6	110	→	011	h_3
7	111	→	111	h_7

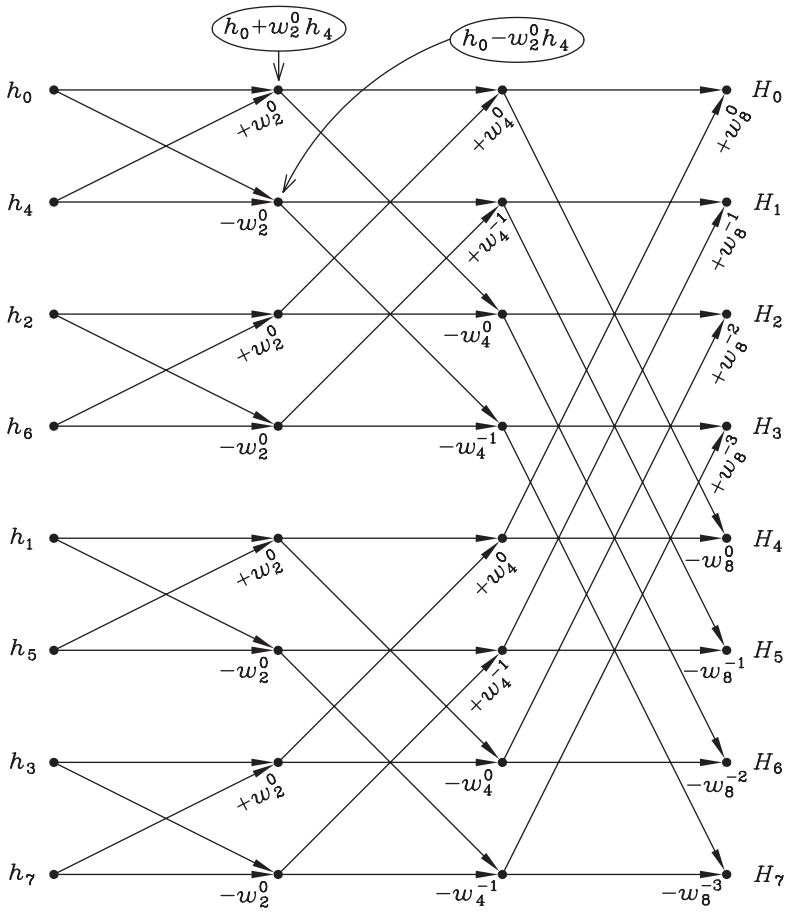


Figure 4.3: The 2^3 net, which illustrates the FFT procedure in the case where the number of data h_j is $N = 8$. $w_N = e^{i2\pi/N}$.

4.3 Computation time

If the discrete Fourier transform is computed directly, employing Equation 4.11, we are said to use the *direct Fourier transform*. The total computation time of the direct Fourier transform is proportional to

$$2N \times \underbrace{2N\left(\text{computation time of } w_{2N}^{-jk} + \text{multiplication time} + \text{addition time}\right)}_{H_k}.$$

In the FFT, the binary inverted data are replaced m times, so that $2N = 2^m$. The total computation time is proportional to

$$m \times 2N(\text{computation time of } w_N^{-k} + \text{multiplication time} + \text{addition time})$$
$$+\text{inversion time}.$$

The ratio of the computation times of direct Fourier transform and fast Fourier transform is, approximately (ignoring the inversion time),

$$\frac{T_{\text{direct}}}{T_{\text{FFT}}} \approx \frac{2N}{m} = \frac{2N}{\log_2(2N)}. \tag{4.20}$$

Example 4.1: The number of signal data is $2N = 2^{20} = 1\,048\,576$, and the computation time taken by FFT is $T_{\text{FFT}} = 1$ hour. What would be the time taken by direct Fourier transform?

Solution. The ratio of the computation times is

$$T_{\text{direct}}/T_{\text{FFT}} \approx 52\,429,$$

and this yields

$$T_{\text{direct}} = 6 \text{ years}.$$

Example 4.2: Write a PASCAL program for fast Fourier transform.

Solution.

```
const
  vectlength = 8192;                    { = 2^13 = the maximum number of data points }
  pi         = 3.14159265358979323846 2643;

type
  realarray  = array[0..vectlength - 1] of real;
  cmplxarray = record
                 re :realarray;
                 im:realarray
```

```
        end;

procedure fft(var d:cmplxarray; n, isign:integer);

    (* Procedure fft takes the fast Fourier transform of a complex data vector of length n (indexed
    from 0 to n − 1) in d in the direction isign. The result is stored back in d. *)

var
    i, istep, j, m, mmax                :integer;
    ct, st, z, theta, wstpr, wstpi, tr, ti:real;

 begin   (* fft *)
   with d do
     begin
       j: = 0;
       for i: = 0 to n − 1 do
         begin
           if i < j then
             begin
               z: = re[i];
               re[i]: = re[j];
               re[j]: = z;
               z: = im[i];
               im[i]: = im[j];
               im[j]: = z
             end;
           m: = n div 2;
           while (j > m) and (m >= 0) do
             begin
               j: = j − m − 1;
               m: = m div 2
             end;
           j: = j + m + 1
         end;
       mmax: = 1;
       while mmax < n do
         begin
           istep: = 2 ∗ mmax;
           for m: = 1 to mmax do
             begin
               i: = m − 1;
               theta: = isign ∗ pi ∗ (m − 1)/mmax;
               ct: = cos(theta);
               st: = sin(theta);
               while i < n do
                 begin
                   j: = i + mmax;
                   tr: = re[i];
```

```
                ti: = im[i];
                wstpr: = ct * re[j] + st * im[j];
                wstpi: = ct * im[j] − st * re[j];
                re[i]: = tr + wstpr;
                im[i]: = ti + wstpi;
                re[j]: = tr − wstpr;
                im[j]: = ti − wstpi;
                i: = i + istep
              end
            end;
          mmax: = istep
        end;
      for i: = 0 to n − 1 do
        begin
          re[i]: = re[i]/sqrt(n);
          im[i]: = im[i]/sqrt(n)
        end
    end
end;    (* fft *)
```

Problems

1. FFT employs the Nyquist sampling. This means that $2N$ data of the signal are taken in constant intervals at points $0, \Delta t, \ldots, (2N-1)\Delta t$, and the discrete spectrum is calculated in constant intervals at points $0, \Delta f, \ldots, (2N-1)\Delta f$, where $N\Delta f = f_{\max} = 1/(2\Delta t)$. Show that

   (a) the product of the data intervals is $\Delta t \Delta f = \dfrac{1}{2N}$, and

   (b) the product of the lengths of the sampled and computed periods is $2T\, 2f_{\max} = 2N$.

2. In order to save computer memory, FFT of the real data h_j, where $j = 0, \ldots, 2N-1$, is calculated by forming N complex numbers c_k whose real parts are the even signal data and imaginary parts the odd signal data:

$$c_k = h_{2k} + ih_{2k+1}, \qquad k = 0, \ldots, N-1.$$

   Find the formula which gives the discrete transform of the original data h_j with the help of the transform of the data c_k.

3. The discrete inverse Fourier transform $\boldsymbol{H}$,

$$H_k = \Delta t \sum_{j=0}^{2N-1} h_j w_{2N}^{-kj}, \qquad k = 0, \ldots, 2N-1,$$

where $w_{2N} = \exp\left(\dfrac{i2\pi}{2N}\right)$, of the vector $\boldsymbol{h} = (h_0, h_1, \ldots, h_{2N-1})$ is computed by FFT. What is the inverse Fourier transform vector in the following cases?

(a) $h_0 = \dfrac{1}{\Delta t}$, other components are zero, (this is discrete Dirac's delta function)

(b) $h_N = \dfrac{1}{\Delta t}$, other components are zero,

(c) $h_p = h_{2N-p} = \dfrac{1}{2\Delta t}$ at one value of the index p $(0 < p < N)$, but other components are zero,

(d) $h_p = \dfrac{1}{2\Delta t}$, $h_{2N-p} = -\dfrac{1}{2\Delta t}$ at one value of the index p $(0 < p < N)$, but other components are zero.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

5 Other integral transforms

5.1 Laplace transform

The inverse *Laplace transform* $H_0(s) = \mathcal{L}_0^{-1}\{h(t)\}$ of a signal $h(t)$ is defined as

$$\boxed{H_0(s) = \int_0^\infty h(t)e^{-st}\,\mathrm{d}t = \mathcal{L}_0^{-1}\{h(t)\}.} \tag{5.1}$$

s is the *complex angular frequency*, which can be divided into the real and the imaginary part:

$$s = \alpha + i\underbrace{2\pi f}_{\omega}. \tag{5.2}$$

Inserting this in Equation 5.1 yields

$$H_0(s) = \int_0^\infty h(t)e^{-\alpha t}e^{-i2\pi f t}\,\mathrm{d}t = \int_{-\infty}^\infty u(t)h(t)e^{-\alpha t}e^{-i2\pi f t}\,\mathrm{d}t = \mathcal{F}^{-1}\{u(t)h(t)e^{-\alpha t}\}, \tag{5.3}$$

where $u(t)$ is the *unit step* function, illustrated in Figure 5.1:

$$u(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{5.4}$$

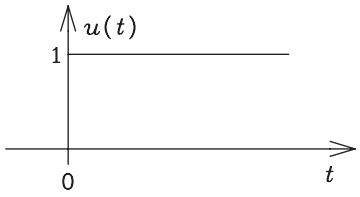


Figure 5.1: Unit step function $u(t)$.

We can see that the inverse Laplace transform $\mathcal{L}_0^{-1}$ of a signal $h(t)$ is the same as the inverse Fourier transform $\mathcal{F}^{-1}$ of the signal $u(t)h(t)e^{-\alpha t}$, which starts at $t = 0$ and where $h(t)$ has been multiplied by an exponential function. In other words, the inverse Laplace transform of a signal $h(t)$ is the spectrum of the signal $h(t)e^{-\alpha t}$, if $h(t) = 0$ at $t < 0$.

Laplace transform can be used in several cases where Fourier transform does not, mathematically, exist. In practice, it is not seldom that one encounters a signal which does not have an inverse Fourier transform which would be an ordinary function. The signal may, however, have an inverse Laplace transform.

Let us, as an example, examine the constant function $h(t) \equiv 1$, which is illustrated in Figure 5.2. The inverse Fourier transform of this constant function is Dirac's delta function, $\mathcal{F}^{-1}\{1\} = \delta(f)$, which is not an ordinary function. Similarly, the inverse Fourier transform of the unit step $u(t)$ is not an ordinary function. However, these functions do have an inverse Laplace transform

$$H_0(s) = \mathcal{L}_0^{-1}\{1\} = \int_0^\infty 1e^{-st}\,\mathrm{d}t = \frac{1}{s} = \mathcal{L}_0^{-1}\{u(t)\} \tag{5.5}$$

which is unambiguous at $\mathrm{Re}\{s\} > 0$.

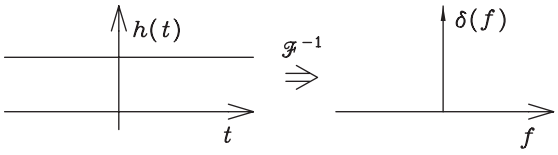


Figure 5.2: Constant function $h(t) \equiv 1$ and its inverse Fourier transform $\delta(f)$.

If $h(t) = 0$ at $t < 0$, then, at points s where $\alpha = 0$,

$$\mathcal{L}_0^{-1}\{h(t)\} = \int_0^\infty h(t)e^{-\alpha t}e^{-i2\pi ft}\,\mathrm{d}t = \int_{-\infty}^\infty h(t)e^{-i2\pi ft}\,\mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}, \tag{5.6}$$

that is, the Laplace transform is equal to the Fourier transform.

From Equation 5.3 we can see that the Fourier transform of a Laplace transform is

$$u(t)h(t)e^{-\alpha t} = \mathcal{F}\{H_0(s)\} = \int_{-\infty}^\infty H_0(s)e^{i2\pi ft}\,\mathrm{d}f. \tag{5.7}$$

Let us, in this last integration, change variables, and use the variable $s = \alpha + i2\pi f$. An increase of f means an increase of the imaginary part of the new variable s, whereas α remains constant. Hence, $\mathrm{d}s = i2\pi\,\mathrm{d}f$. Integration with respect to the new variable is made along a

line in the direction of the imaginary axis, from $\alpha - i\infty$ to $\alpha + i\infty$. The value of the integral is independent of the choice of the constant α. We obtain

$$u(t)h(t) = \frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} H_0(s) e^{st} \, \mathrm{d}s = \mathcal{L}_0\{H_0(s)\}. \tag{5.8}$$

Since this integration of $H(s)$ gives again the signal $h(t)$ (at $t \geq 0$), it is the inverse transform of $\mathcal{L}_0^{-1}$, the Laplace transform $\mathcal{L}_0$.

Table 5.1 lists the inverse Laplace and inverse Fourier transforms of some signals. In practice, it is very seldom necessary actually to compute the integral of Equation 5.8 or even of Equation 5.1. Instead, transforms are found by using transform tables and applying theorems, the most important of which are included in Table 5.1.

Table 5.1: Inverse Laplace and inverse Fourier transforms of a few signals. In the table, $\mathcal{L}_0^{-1}\{h(t)\} = H_0(s)$, and $\mathcal{F}^{-1}\{h(t)\} = H(f)$. In the first row, the term $-h(0^+)$ in $sH_0(s) - h(0^+)$ is omitted, if the derivative $\mathrm{d}h/\mathrm{d}t$ is assumed to contain Dirac's delta peak $\lim_{a \to 0+} h(a)\delta(t)$ and if $h(t) = 0$ at $t < 0$. In the sixth row, $h(t - a) = 0$ at $t < a$.

Signal	Inverse Laplace transform $\mathcal{L}_0^{-1}$	Inverse Fourier transform $\mathcal{F}^{-1}$
$\mathrm{d}h(t)/\mathrm{d}t$	$sH_0(s) - h(0^+)$	$2\pi f i H(f)$
$\int_{-\infty}^{t} h(\tau)\,\mathrm{d}\tau$	$H_0(s)/s$	—
$th(t)$	$-\mathrm{d}H_0(s)/\mathrm{d}s$	$-\mathrm{d}H(f)/(2\pi i \mathrm{d}f)$
$e^{-at}h(t)$	$H_0(s + a)$ $(\mathrm{Re}\{s\} > -a)$	—
$h(t/a)$ $(a > 0)$	$aH_0(as)$	$aH(af)$
$h(t - a)$ $(a > 0)$	$e^{-as}H_0(s)$	$e^{-i2\pi f a}H(f)$
e^{-at}	$1/(s + a)$ $(\mathrm{Re}\{s\} > -a)$	—
t^n $(n \geq 0)$	$n!/s^{n+1}$	—

An alternative integral which is frequently used to define the inverse Laplace transform is

$$\boxed{H_\infty(s) = \int_{-\infty}^{\infty} h(t)e^{-st}\,\mathrm{d}t = \mathcal{L}_\infty^{-1}\{h(t)\} = \mathcal{F}^{-1}\{h(t)e^{-\alpha t}\}.} \tag{5.9}$$

In this integral, the integration stretches from $-\infty$ to $+\infty$. If this definition is used, Equation 5.8 becomes

$$\boxed{h(t) = \frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty} H_\infty(s)e^{st}\,\mathrm{d}s = \mathcal{L}_\infty\{H_\infty(s)\}.} \tag{5.10}$$

This is the inverse transform of $\mathcal{L}_\infty^{-1}$, the Laplace transform $\mathcal{L}_\infty$.

Example 5.1: (a) Show that the inverse Laplace transform of a periodic function h, whose period is p, that is, $h(t+p) = h(t)$ at all t, is

$$\mathcal{L}_0^{-1}\{h(t)\} = \frac{1}{1-e^{-ps}}\int_0^p h(t)e^{-st}\,\mathrm{d}t.$$

(b) Derive the $\mathcal{L}_0^{-1}$ transform of the square wave

$$h(t) = \begin{cases} V_0, & 2nT < t \le (2n+1)T \qquad (n = 0, 1, 2, \ldots), \\ -V_0, & (2n+1)T < t \le (2n+2)T \qquad (n = 0, 1, 2, \ldots), \\ 0, & t < 0. \end{cases}$$

Solution. (a) Let us define the function h_p as

$$h_p = \begin{cases} h(t), & 0 \le t < p, \\ 0 & \text{otherwise.} \end{cases}$$

Applying the linearity of the Laplace transforms, as well as the shift theorem in the sixth row of Table 5.1, we obtain that

$$\begin{aligned} \mathcal{L}_0^{-1}\{h(t)\} &= \mathcal{L}_0^{-1}\{h_p(t) + h_p(t-p) + h_p(t-2p) + \ldots\} \\ &= (1 + e^{-ps} + e^{-2ps} + \ldots)\,\mathcal{L}_0^{-1}\{h_p(t)\} \\ &= \frac{1}{1-e^{-ps}}\int_0^p h(t)e^{-st}\,\mathrm{d}t. \end{aligned}$$

(b) Applying the result of (a),

$$\begin{aligned} H_0(s) &= \frac{1}{1-e^{-2Ts}}\left(\int_0^T V_0 e^{-st}\,\mathrm{d}t - \int_T^{2T} V_0 e^{-st}\,\mathrm{d}t\right) \\ &= \frac{V_0}{1-e^{-2Ts}}\frac{1}{s}(1-e^{-Ts}+e^{-2Ts}-e^{-Ts}) \\ &= \frac{V_0}{s}\frac{(1-e^{-Ts})^2}{(1-e^{-Ts})(1+e^{-Ts})} \\ &= \frac{V_0}{s}\frac{(1-e^{-Ts})}{(1+e^{-Ts})} \\ &= \frac{V_0}{s}\tanh(Ts/2). \end{aligned}$$

Example 5.2: Applying Laplace transforms, find the solution for

$$\begin{cases} \dfrac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) \equiv 1, & t \geq 0, \\ y(0) = y_0. \end{cases}$$

Solution. Let us denote $Y(s) = \mathcal{L}_0^{-1}\{y(t)\}$. Taking the inverse Laplace transform of both sides of $\mathrm{d}y(t)/\mathrm{d}t + y(t) = 1$, and applying the derivative theorem in the first row of Table 5.1, we obtain

$$sY(s) - y(0^+) + Y(s) = \frac{1}{s}.$$

Solving for $Y(s)$ gives

$$\begin{aligned} Y(s) &= \frac{\frac{1}{s} + y(0^+)}{s+1} = \frac{1+sy(0^+)}{s(s+1)} \\ &= \frac{1}{s} + \frac{y(0^+)-1}{s+1}. \end{aligned}$$

$\mathcal{L}_0$ transform (that is, reading Table 5.1 from right to left) gives

$$y(t) = 1 + (y_0 - 1)e^{-t}.$$

5.2 Transfer function of a linear system

A *linear system* is a system where the input signal and the output signal are related by a linear differential equation with constant coefficients. Let us denote $h_{\text{in}}(t)$ the input signal and $h_{\text{out}}(t)$ the output signal, as illustrated in Figure 5.3. The system can be described by an equation

$$\begin{aligned} & \left[A_n \frac{\mathrm{d}^n}{\mathrm{d}t^n} + \cdots + A_2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} + A_1 \frac{\mathrm{d}}{\mathrm{d}t} + A_0 \right] h_{\text{out}}(t) \\ = & \left[B_m \frac{\mathrm{d}^m}{\mathrm{d}t^m} + \cdots + B_2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} + B_1 \frac{\mathrm{d}}{\mathrm{d}t} + B_0 \right] h_{\text{in}}(t). \end{aligned} \tag{5.11}$$

The *derivative theorem* of the Laplace transform $\mathcal{L}_\infty^{-1}$ states that

$$\boxed{\mathcal{L}_\infty^{-1}\left\{ \frac{\mathrm{d}^k h(t)}{\mathrm{d}\, t^k} \right\} = s^k \mathcal{L}_\infty^{-1}\{h(t)\}.} \tag{5.12}$$

Equation 5.11 can be solved in a very practical way by taking the inverse Laplace transform $\mathcal{L}_\infty^{-1}$ of both sides of the equation, and applying the derivative theorem. We obtain

$$\begin{aligned} & \left(A_n s^n + \cdots + A_2 s^2 + A_1 s + A_0 \right) \mathcal{L}_\infty^{-1}\{h_{\text{out}}(t)\} \\ = & \left(B_m s^m + \cdots + B_2 s^2 + B_1 s + B_0 \right) \mathcal{L}_\infty^{-1}\{h_{\text{in}}(t)\}. \end{aligned} \tag{5.13}$$

The *transfer function* of a system can be defined as the relation of the inverse Laplace transform of the output signal to the inverse Laplace transform of the input signal. Applying Equation 5.13, we obtain that the transfer function of a linear system is

$$\boxed{G(s) = \frac{H_\infty^{\text{out}}(s)}{H_\infty^{\text{in}}(s)} = \frac{\mathcal{L}_\infty^{-1}\{h_{\text{out}}(t)\}}{\mathcal{L}_\infty^{-1}\{h_{\text{in}}(t)\}} \overset{(5.13)}{=} \frac{\sum_{j=0}^{m} B_j s^j}{\sum_{k=0}^{n} A_k s^k}.} \tag{5.14}$$

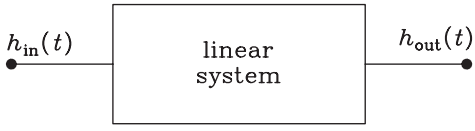


Figure 5.3: Schematic illustration of a linear system. $h_{\text{in}}(t)$ is the input signal and $h_{\text{out}}(t)$ is the output signal.

At the points where the real part of s is zero, that is, $\alpha = 0$ and $s = i2\pi f$, the Laplace transform $\mathcal{L}_\infty^{-1}$ can be replaced by the Fourier transform $\mathcal{F}^{-1}$, and the transfer function is

$$\boxed{G(f) = \frac{H_{\text{out}}(f)}{H_{\text{in}}(f)} = \frac{\mathcal{F}^{-1}\{h_{\text{out}}(t)\}}{\mathcal{F}^{-1}\{h_{\text{in}}(t)\}} = \frac{\sum_{j=0}^{m} B_j (i2\pi f)^j}{\sum_{k=0}^{n} A_k (i2\pi f)^k}.} \tag{5.15}$$

The *impulse response* of a system is the Laplace transform of the transfer function:

$$\boxed{g(t) = \mathcal{L}_\infty\{G(s)\} = \frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} G(s) e^{st}\, \mathrm{d}s.} \tag{5.16}$$

At points where $\alpha = 0$ and $s = 2\pi i f$

$$g(t) = \mathcal{F}\{G(f)\} = \int_{-\infty}^{\infty} G(f) e^{i2\pi f t}\, \mathrm{d}f. \tag{5.17}$$

The definition of the transfer function, Equation 5.14, yields

$$H_\infty^{\text{out}}(s) = G(s) H_\infty^{\text{in}}(s). \tag{5.18}$$

Consequently, applying the convolution theorem of the Laplace transform, we obtain that the output signal is the convolution of the input signal and the impulse response of the system:

$$\boxed{h_{\text{out}}(t) = g(t) * h_{\text{in}}(t).} \tag{5.19}$$

If the input signal is Dirac's delta peak or impulse, $h_{\text{in}}(t) = \delta(t)$, then the output signal is the impulse response of the system,

$$h_{\text{out}}(t) = g(t) * \delta(t) = g(t). \tag{5.20}$$

This is illustrated in Figure 5.4. In principle, using Dirac's delta peak as input would give a method to determine the impulse response. However, for example in electronics, it may be very difficult to produce a sufficiently narrow peak which would work as an infinitely high and narrow Dirac's delta peak.

Unlike Dirac's delta peak, a unit step function $u(t)$ is easy to produce in practice. If $h_{\text{in}}(t) = u(t)$, then $h_{\text{out}}(t) = u(t) * g(t)$. This output signal of a unit step function is called the *step response*. Let us examine its derivative (for differentiation of a step, see Problem 1.9 or 12.7):

$$\frac{\mathrm{d}h_{\text{out}}(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}[u(t) * g(t)] = \frac{\mathrm{d}u(t)}{\mathrm{d}t} * g(t) = \delta(t) * g(t) = g(t), \tag{5.21}$$

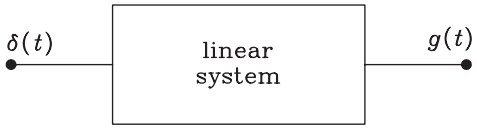


Figure 5.4: If the input signal of a linear system is a delta peak, $h_{\text{in}}(t) = \delta(t)$, then the output signal is the impulse response of the system, $h_{\text{out}}(t) = g(t)$.

because, for convolution,

$$\frac{\mathrm{d}}{\mathrm{d}t}(f * g) = \frac{\mathrm{d}f}{\mathrm{d}t} * g = f * \frac{\mathrm{d}g}{\mathrm{d}t}. \tag{5.22}$$

Consequently, the impulse response of the system can be obtained by using a unit step function, $h_{\text{in}}(t) = u(t)$, as the input signal, and calculating the derivative of the output signal,

$$\boxed{g(t) = \frac{\mathrm{d}h_{\text{out}}(t)}{\mathrm{d}t}}. \tag{5.23}$$

Figure 5.5 shows the situation schematically.

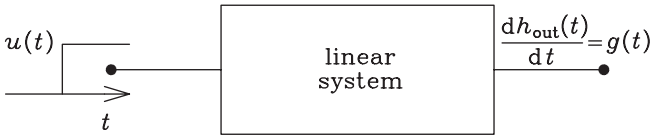


Figure 5.5: If the input signal of a linear system is a unit step function, $h_{\text{in}}(t) = u(t)$, then the impulse response of the system is the derivative of the output signal, $\mathrm{d}h_{\text{out}}(t)/\mathrm{d}t = g(t)$.

Example 5.3: Find the impulse response and the transfer function of the *integrating RC circuit* of Figure 5.6.

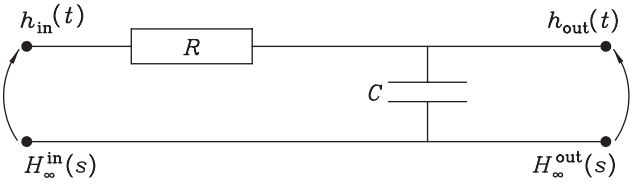


Figure 5.6: Schematic illustration of an integrating RC circuit. The input signal is $h_{\text{in}}(t)$ and the output signal $h_{\text{out}}(t)$. The spectrum of the input is $H_{\text{in}}(f) = H_{\infty}^{\text{in}}(s)$, where $s = i2\pi f$ ($\alpha = 0$). The impedance of capacitor C is $Z_C = 1/(i2\pi fC) = 1/(sC)$.

Solution. If the input signal voltage of the integrating RC circuit is a unit step function, then the output signal voltage, the step response, is

$$h_{\text{out}}(t) = 1 - e^{-t/(RC)} \quad (t \geq 0). \tag{5.24}$$

These input and output functions are illustrated in Figure 5.7.

The impulse response of the circuit can be calculated from the step response:

$$g(t) = \frac{\mathrm{d}h_{\text{out}}(t)}{\mathrm{d}t} = \frac{\mathrm{d}\left[1 - e^{-t/(RC)}\right]}{\mathrm{d}t} = \frac{1}{RC}\, e^{-t/(RC)}. \tag{5.25}$$

The transfer function of the circuit is

$$\begin{aligned} G(s) &= \mathcal{L}_{\infty}^{-1}\{g(t)\} = \int_0^{\infty} \frac{1}{RC}\, e^{-t/(RC)} e^{-st}\, \mathrm{d}t \\ &= \int_0^{\infty} \frac{1}{RC}\, \frac{-1}{\left(\frac{1}{RC} + s\right)}\, e^{-t\left(\frac{1}{RC}+s\right)} = \frac{1}{1 + RCs}, \end{aligned} \tag{5.26}$$

in accordance with Table 5.1. Above we have assumed that $g(t) = 0$ at $t < 0$, and hence $\mathcal{L}_0^{-1}\{g(t)\} = \mathcal{L}_{\infty}^{-1}\{g(t)\}$. The transfer function has the amplitude

$$|G(s)| = \sqrt{G^*(s)G(s)} = \frac{1}{\sqrt{1 + (2\pi f RC)^2}} \tag{5.27}$$

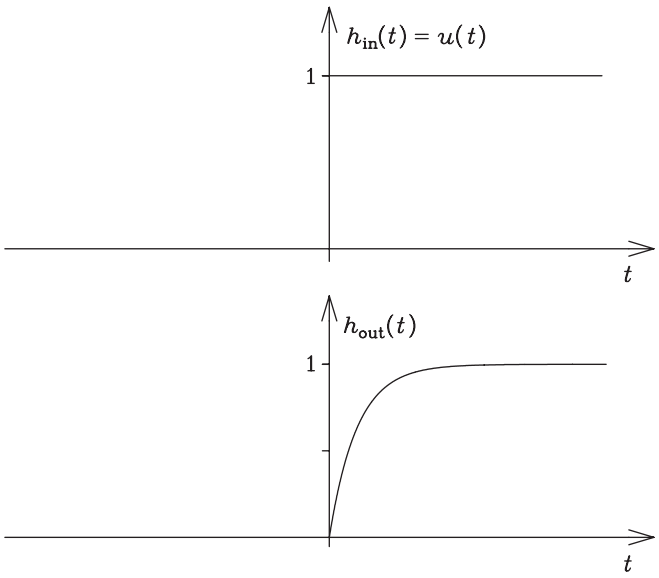


Figure 5.7: If the input signal of an integrating RC circuit is the unit step function, $h_{\text{in}}(t) = u(t)$ (upper curve), then the output signal is $h_{\text{out}}(t) = 1 - e^{-t/(RC)}$ $(t \geq 0)$ (lower curve).

and the phase

$$\phi(s) = \arctan\left\{\frac{\operatorname{Im} G(s)}{\operatorname{Re} G(s)}\right\} = \arctan(-RCs). \tag{5.28}$$

The transfer function of an electric circuit can also be found more directly. Let us examine the circuit in Figure 5.8, which consists of two impedances Z_1 and Z_2. If the input signal voltage $h_{\text{in}}(t)$ is a sinusoidal wave, of a frequency $f = \dfrac{\omega}{2\pi} = \dfrac{s}{i2\pi}$, then, according to traditional electronics, the output signal voltage is also sinusoidal and is given by

$$h_{\text{out}}(t) = \frac{Z_2}{Z_1 + Z_2} h_{\text{in}}(t). \tag{5.29}$$

Taking the Laplace transform $\mathcal{L}_\infty^{-1}$ of both sides of this equation, and remembering that Z_1 and Z_2 are constants, independent of time, we obtain that

$$H_\infty^{\text{out}}(s) = \frac{Z_2}{Z_1 + Z_2} H_\infty^{\text{in}}(s). \tag{5.30}$$

Now, since $h_{\text{in}}(t)$ and $h_{\text{out}}(t)$ are both sinusoids with the same frequency f, $H_\infty^{\text{in}}(s)$ and $H_\infty^{\text{out}}(s)$ are both Dirac's delta peaks at the same point $s = i2\pi f$. On the other hand, Z_1 and Z_2 both depend on angular frequency $\omega = 2\pi f$ and thereby on the point s where $H_\infty^{\text{in}}(s)$ and $H_\infty^{\text{out}}(s)$ are nonzero. Therefore we may write

$$H_\infty^{\text{out}}(s) = \frac{Z_2(s)}{Z_1(s) + Z_2(s)} H_\infty^{\text{in}}(s). \tag{5.31}$$

Since this holds for every s separately, it is also valid for a general input, wherefrom we obtain

$$G(s) = \frac{H_\infty^{\text{out}}(s)}{H_\infty^{\text{in}}(s)} = \frac{Z_2(s)}{Z_1(s) + Z_2(s)}. \tag{5.32}$$

Let us calculate the transfer function of the integrating circuit by applying Equation 5.32. In this case, $Z_1 = R$ and $Z_2 = 1/(i2\pi fC) = 1/(sC)$. Consequently,

$$G(s) = \frac{\frac{1}{sC}}{R + \frac{1}{sC}} = \frac{1}{1 + sRC}. \tag{5.33}$$

This is, indeed, the same result as in Equation 5.26.

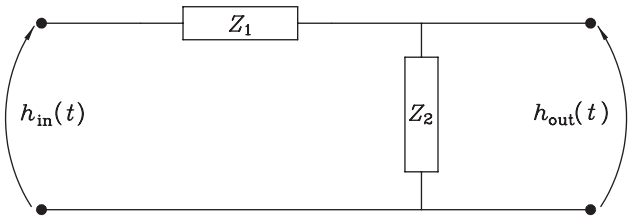


Figure 5.8: Schematic illustration of a simple electric circuit, consisting of two impedances Z_1 and Z_2. The input signal is $h_{\text{in}}(t)$ and the output signal $h_{\text{out}}(t)$.

Example 5.4: Find the impulse response and the transfer function of the *differentiating RC circuit* of Figure 5.9.

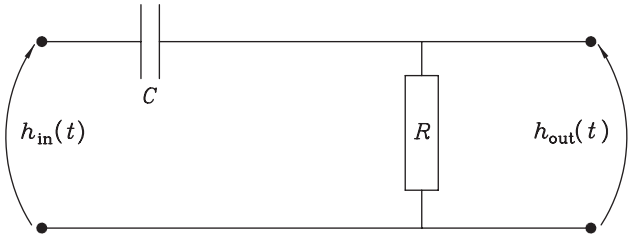


Figure 5.9: Schematic illustration of a differentiating RC circuit. The input signal is $h_{\text{in}}(t)$ and the output signal $h_{\text{out}}(t)$.

Solution. If the input signal of the differentiating RC circuit is a unit step function $h_{\text{in}}(t) = u(t)$, then the circuit gives the output signal

$$h_{\text{out}} = e^{-t/(RC)}, \quad (t \geq 0), \tag{5.34}$$

shown in Figure 5.10.

The impulse response of the circuit is ($g(t) = 0$ at $t < 0$)

$$g(t) = \frac{\mathrm{d}h_{\text{out}}(t)}{\mathrm{d}t} = -\frac{1}{RC}\, e^{-t/(RC)} + \delta(t), \quad t \geq 0. \tag{5.35}$$

The transfer function of the circuit is

$$\begin{aligned} G(s) &= \mathcal{L}_{\infty}^{-1}\{g(t)\} = \int_0^{\infty} -\frac{1}{RC}\, e^{-t/(RC)} e^{-st}\, \mathrm{d}t + \int_{-\infty}^{\infty} \delta(t) e^{-st}\, \mathrm{d}t \\ &= -\frac{1}{1+RCs} + 1 = \frac{RCs}{1+RCs}. \end{aligned} \tag{5.36}$$

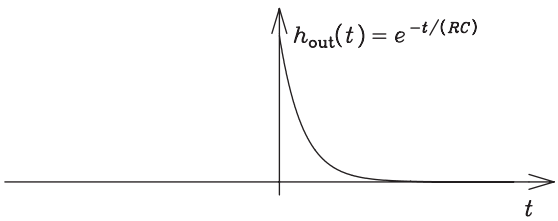


Figure 5.10: Output signal of a differentiating RC circuit is $h_{\text{out}}(t) = e^{-t/(RC)}$ (at $t \geq 0$), if the input signal is the unit step function.

We can check this result by applying Equation 5.32. In this case,

$$G(s) = \frac{R}{R + \frac{1}{sC}} = \frac{RCs}{1 + RCs}. \tag{5.37}$$

Example 5.5: Find the impulse response and the transfer function of the *RCL circuit* of Figure 5.11.

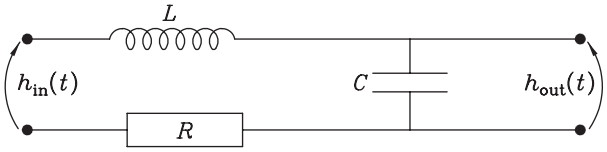


Figure 5.11: Schematic illustration of an RCL circuit. The input signal is $h_{\text{in}}(t)$ and the output signal $h_{\text{out}}(t)$.

Solution. Applying Equation 5.32, we obtain that the transfer function of the RCL circuit is

$$G(s) = \frac{\frac{1}{sC}}{sL + \frac{1}{sC} + R} = \frac{1}{s^2LC + sRC + 1} = \frac{\frac{1}{LC}}{s^2 + \frac{sR}{L} + \frac{1}{LC}}. \tag{5.38}$$

We divide this expression into partial fractions:

$$G(s) = \frac{\frac{1}{LC}}{s^2 + \frac{sR}{L} + \frac{1}{LC}} = \frac{A}{s+a} - \frac{A}{s+b}, \tag{5.39}$$

where $-a$ and $-b$ are the roots of the denominator, that is,

$$\begin{cases} a &= +\dfrac{R}{2L} + \dfrac{1}{2}\sqrt{\dfrac{R^2}{L^2} - \dfrac{4}{LC}}, \\ b &= +\dfrac{R}{2L} - \dfrac{1}{2}\sqrt{\dfrac{R^2}{L^2} - \dfrac{4}{LC}}, \end{cases} \tag{5.40}$$

and $A(s+b) - A(s+a) = 1/(LC)$, or

$$A = \frac{1}{LC(b-a)}. \tag{5.41}$$

Applying the rule in the seventh row of Table 5.1, we can calculate the impulse response of the circuit:

$$\begin{aligned} g(t) &= \mathcal{L}_\infty\{G(s)\} = \mathcal{L}_\infty\left\{\frac{A}{s+a}\right\} - \mathcal{L}_\infty\left\{\frac{A}{s+b}\right\} \\ &\overset{\mathcal{L}_0=\mathcal{L}_\infty}{=} Ae^{-at} - Ae^{-bt} = \frac{1}{LC(b-a)}\left(e^{-at} - e^{-bt}\right). \end{aligned} \tag{5.42}$$

5.3 z transform

z transform is another integral transform which is in general use in electronics and signal processing. The z transform is the same as the Laplace transform $\mathcal{L}_{\infty}^{-1}$ of a discrete signal.

Let us examine the Laplace transform

$$H_{\infty}(s) = \int_{-\infty}^{\infty} h(t)e^{-st}\,\mathrm{d}t = \mathcal{L}_{\infty}^{-1}\{h(t)\} \tag{5.43}$$

of a signal $h(t)$, which is discrete, but infinitely long (not truncated). The Laplace transform of the discrete signal is

$$H_{\infty}^{\Delta t}(s) = \sum_{j=-\infty}^{\infty} \Delta t\, h(j\,\Delta t)e^{-sj\,\Delta t} = \sum_{j=-\infty}^{\infty} \Delta t\, h(j\,\Delta t)z^{-j}, \tag{5.44}$$

where $z = e^{s\,\Delta t}$, and $s = \alpha + i2\pi f$. The variable z can also be written as

$$z = \underbrace{e^{\alpha\,\Delta t}}_{=r} e^{i2\pi f\,\Delta t} = re^{i2\pi f\,\Delta t}. \tag{5.45}$$

If $|z| = 1$, that is, $r = 1$, or $\alpha = 0$, then the z transform is the same as the Fourier transform of a discrete signal.

If the data are abbreviated as $h_j = \Delta t\, h(j\,\Delta t)$, then the z transform can be written as

$$H_{\infty}^{\Delta t}(z) = \sum_{j=-\infty}^{\infty} h_j z^{-j}. \tag{5.46}$$

A one-sided z transform is also in frequent use. The one-sided z transform is the $\mathcal{L}_0^{-1}$ transform of a discrete signal,

$$H_0^{\Delta t}(z) = \sum_{j=0}^{\infty} h_j z^{-j}. \tag{5.47}$$

This is the same as the two-sided transform in Equation 5.46, if the signal $h(t) = 0$ at $t < 0$.

Generally, it is possible to apply a continuous transform (or integration) to a discrete signal, if the signal is written as

$$h(t) = h_0\delta(t) + h_1\delta(t - \Delta t) + \cdots + h_k\delta(t - k\,\Delta t) + \cdots. \tag{5.48}$$

If we take the continuous Laplace transform $\mathcal{L}_{\infty}^{-1}$ of a term in the sum of Equation 5.48, we obtain

$$\mathcal{L}_{\infty}^{-1}\{h_k\delta(t - k\,\Delta t)\} = \int_{-\infty}^{\infty} h_k\delta(t - k\,\Delta t)e^{-st}\,\mathrm{d}t = h_k e^{-s\,\Delta t\,k} = h_k z^{-k}. \tag{5.49}$$

The *shift theorem* of the Laplace transform states that if a signal is shifted by Δt, then the spectrum is multiplied by a phase factor $e^{-s\,\Delta t} = z^{-k}$. This can be seen in Equation 5.49.

Example 5.6: What is the z transform of the vector $h_j = 1$ for all $j \geq 0$, that is, the vector $(1, 1, 1, \ldots)$?

Solution. Applying Equation 5.46, the z transform is

$$H_\infty^{\Delta t}(z) = 1 + z^{-1} + z^{-2} + \cdots = 1/(1 - z^{-1}). \tag{5.50}$$

Example 5.7: What is the z transform of the vector $h_0 = 0$ for $j = 0$ and $h_j = 1$ for all $j \geq 1$, that is, the vector $(0, 1, 1, 1, \ldots)$?

Solution. Applying the shift theorem, the z transform is

$$H_\infty^{\Delta t}(z) = z^{-1}/(1 - z^{-1}). \tag{5.51}$$

Problems

1. In the theory of (one-sided) Laplace transforms, functions are assumed to be zero at negative values of t. Hence the definition of convolution achieves the form

   $$h(t) * g(t) = \int_0^t h(u)g(t-u)\,\mathrm{d}u.$$

   Let us denote $\mathcal{L}_0^{-1}\{h(t)\} = H_0(s)$ and $\mathcal{L}_0^{-1}\{g(t)\} = G_0(s)$. Prove the convolution theorem of Laplace transforms

   $$\mathcal{L}_0^{-1}\{h * g\} = H_0 G_0.$$

2. Let us assume that a function $h(t)$ is continuous and differentiable in the interval from 0 to ∞, and it has no step, even in the origin. Let us also assume that $\mathcal{L}_0^{-1}\{h(t)\}$ exists. Show that in this case

   $$\mathcal{L}_0^{-1}\{h^{(1)}(t)\} = s\mathcal{L}_0^{-1}\{h(t)\} - h(0),$$

   where $h^{(1)}(t) = \mathrm{d}h(t)/\mathrm{d}t$. Also show that if the function $h(t)$ has in the origin a step from 0 to $h(0^+) = \lim_{a \to 0+} h(a)$ (in which case the derivative of the function is in the origin $h^{(1)}(0) = h(0^+)\delta(t)$), then the rule is reformulated as

   $$\mathcal{L}_0^{-1}\{h^{(1)}(t)\} = s\mathcal{L}_0^{-1}\{h(t)\}.$$

   (The effect of the term $h(0^+)$ is then included in $h^{(1)}$.)

3. (a) Compute $\mathcal{L}_0^{-1}\{\delta(t-t_0)\}$, where $t_0 \geq 0$.

   (b) Compute $\mathcal{L}_0^{-1}\left\{\mathcal{L}_0^{-1}\{\delta(t)\}\right\}$.

   (c) Which function has the function $\dfrac{bs}{s+a}$ as its $\mathcal{L}_0^{-1}$ transform?

4. Let us denote $\mathcal{L}_\infty^{-1}\{h(t)\}$ the two-sided inverse Laplace transform of a function $h(t)$, *i.e.*,

$$\mathcal{L}_\infty^{-1}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-st}\,\mathrm{d}t.$$

   Compute the two-sided inverse Laplace transform of the function $h(t) = e^{-a|t|}$, where a is a real constant. At which values of s and a does the transform exist (as an ordinary function)?

5. (a) Compute $\mathcal{L}_0^{-1}\{\cos(at)\}$, where a is a real constant.

   (b) Applying the result of (a), and the rule $\mathcal{L}_0^{-1}\{h^{(1)}(t)\} = sH_0(s) - h(0^+)$, compute $\mathcal{L}_0^{-1}\{\sin(at)\}$.

6. Applying the Laplace transform $\mathcal{L}_0^{-1}$, solve the differential equation

$$h^{(2)}(t) + h(t) = t, \quad h(0) = 0, \quad h^{(1)}(0) = 2,$$

   where $h^{(1)}(t) = \mathrm{d}h(t)/\mathrm{d}t$, and $h^{(2)}(t) = \mathrm{d}^2h(t)/\mathrm{d}t^2$.

   Hint: $\mathcal{L}_0^{-1}\{\sin(at)\} = \dfrac{a}{s^2+a^2}$.

7. Applying the Laplace transforms, solve the pair of differential equations

$$\begin{cases} f^{(1)}(t) + 2f(t) &= 2g(t), \\ g^{(1)}(t) &= f^{(1)}(t) - 2f(t) \end{cases} \quad \text{at initial values} \quad \begin{cases} f(0) &= 0, \\ g(0) &= 1. \end{cases}$$

8. A vertical string (string constant k) is fixed from its upper end. A mass m is attached to the lower end. In addition to the string force, the string is affected by a periodic vertical force $F_0\sin(pt)$. Let us denote y the downward deviation of the string from the equilibrium position. The equation of motion is then

$$m\,\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = F_0\sin(pt) - ky, \qquad \text{or} \qquad \frac{\mathrm{d}^2y}{\mathrm{d}t^2} = K\sin(pt) - \omega_0^2 y,$$

   where $K = F_0/m$ and $\omega_0 = \sqrt{k/m}$. Applying the Laplace transforms, solve $y(t)$ with the initial values $y(0) = 0$ and $y^{(1)}(0) = 0$. How does the solution behave in the resonance $p^2 = \omega_0^2$?

9. The impulse response of an electric circuit is a boxcar-shaped signal of height $1/T$ and of duration from 0 to T. What is the output of the circuit, if the input signal is a boxcar-shaped voltage pulse of height U and of duration $2T$?

10. The output and the input of a linear system are connected by the differential equation

$$\left(4 - \frac{\mathrm{d}^2}{\mathrm{d}t^2}\right) h_{\mathrm{out}}(t) \equiv \frac{\mathrm{d}}{\mathrm{d}t}\, h_{\mathrm{in}}(t).$$

Find the impulse response $g(t)$ of the system.

Hint: $\mathcal{L}_{\infty}^{-1}\left\{e^{-a|t|}\right\} = \dfrac{1}{a-s} + \dfrac{1}{a+s}$ (two-sided Laplace transform).

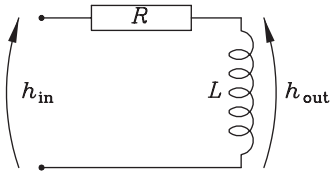
11. What is the output of a linear circuit, if

$$h_{\mathrm{in}}(t) = \begin{cases} Ae^{-\beta_0 t}, & t \geq 0, \\ 0, & t < 0, \end{cases} \qquad \text{and} \qquad G(s) = \frac{1}{s/\beta_1 + 1}\,?$$

12. Find the impulse response of the following RL circuit,

(a) by formulating the transfer function $G(s)$ and applying Laplace transforms, and

(b) without Laplace transforms, by first solving the current i from the differential equation $L\mathrm{d}i/\mathrm{d}t + Ri = h_{\mathrm{in}}$, where h_{in} is the step function u.



13. The impulse response of a linear system is the signal e^{-Ct}, which starts at $t = 0$ and damps out exponentially. Two such systems are combined in a series, so that the output of the first system is the input of the second system. What is the impulse response of the combined system?

14. The z transforms of the data sequences $\{h_j\}$ and $\{g_j\}$ are $H_{\infty}^{\Delta t}(z)$ and $G_{\infty}^{\Delta t}(z)$, respectively. Let us define the data sequence $\{e_j\}$ as $e_j = \sum_{k=-\infty}^{\infty} h_k g_{j-k}$. Show that the z transform of $\{e_j\}$ is $E_{\infty}^{\Delta t}(z) = H_{\infty}^{\Delta t}(z) G_{\infty}^{\Delta t}(z)$. (This result is the convolution theorem of the z transforms.)

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

6 Fourier transform spectroscopy (FTS)

A. A. Michelson invented in the 1880s a simple interferometer, which basically consists of a beamsplitter and two plane mirrors. Initially, the interferometer was used to study the speed of light and to fix the standard meter with the wavelength of a spectral line. The use of the Michelson interferometer as a spectrometer was started more than half a century later. In 1949, astrophysicist Peter Fellgett applied the interferometer to measure the light from celestial bodies, and computed the first Fourier transform spectrum.

6.1 Interference of light

A monochromatic electromagnetic wave with an electric field strength E is given by

$$E = Ae^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}, \tag{6.1}$$

where A is the amplitude of the wave, $\boldsymbol{k}$ is the wave vector, $\boldsymbol{r}$ is the position vector, ω is the angular frequency, and t is time. In this expression,

$$k = |\boldsymbol{k}| = \frac{2\pi}{\lambda} = 2\pi\nu, \tag{6.2}$$

where λ is the wavelength of the electromagnetic wave. Both k and ν are called the wavenumber. It is conventional to express the wavenumber in the unit $1/\mathrm{cm}$ instead of $1/\mathrm{m}$.

Usually, we are not able to measure E as a function of time, because the angular frequency ω is too high. We measure the intensity, that is, the power of the electromagnetic wave per area, instead. The intensity I is proportional to the square of the amplitude:

$$I \propto A^2. \tag{6.3}$$

Let us examine the interference of two electromagnetic waves $E_1 = A_1 e^{i(k_x x-\omega t)}$ and $E_2 = A_2 e^{i(k_x x-\omega t+\delta)}$, propagating in the same direction x. The waves, their amplitudes, and the resultant amplitude are illustrated in Figure 6.1. The phase difference of the two waves is δ. The square of the amplitude of the resulting wave is

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2\cos\delta. \tag{6.4}$$

If I_1 and I_2 are the intensities of the two interfering waves, the intensity of the interfered wave is

$$\boxed{I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\cos\delta\,.} \tag{6.5}$$

If the two interfering waves have equal intensity, that is, $I_1 = I_2 = I_0$, then the intensity of the resulting wave is simply

$$I = 2I_0(1 + \cos\delta). \tag{6.6}$$

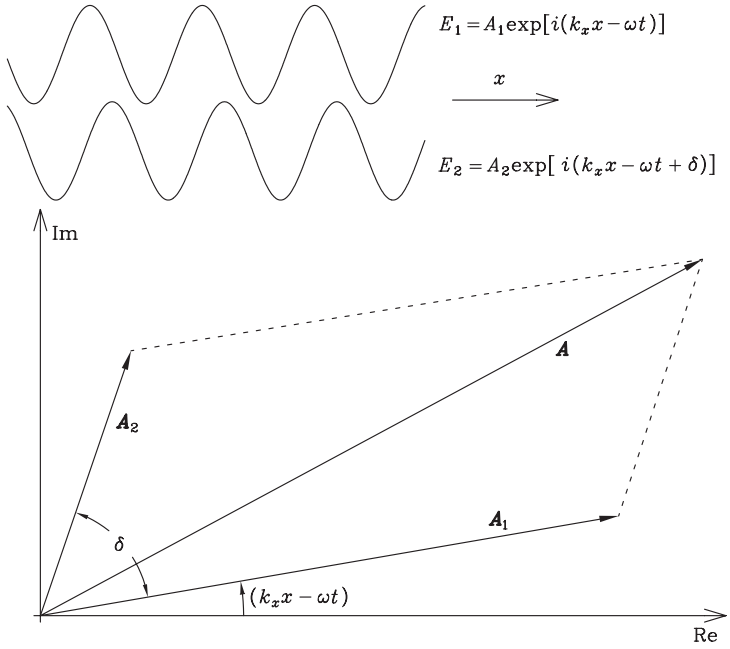


Figure 6.1: Interference of the two waves $E_1 = A_1 e^{i(k_x x - \omega t)}$ and $E_2 = A_2 e^{i(k_x x - \omega t + \delta)}$. A is the resultant amplitude.

6.2 Michelson interferometer

The Michelson interferometer is very simple, but its principles of function can be applied to other interferometers used in Fourier transform spectroscopy, FTS. The basic optical layout of the Michelson interferometer is illustrated in Figure 6.2. The incident beam is divided by a beamsplitter B into two parts: the reflected beam travels to a fixed mirror M_1 and back, and the transmitted beam travels to a moving mirror M_2 and back. When the two beams return to the beamsplitter they interfere.

Effectively, the Michelson interferometer produces two coherent images S′ and S″ of the real source S_0, as shown in Figure 6.3. The beamsplitter forms an image S_0' of the source S_0, and an image M_2' of the moving mirror M_2. S′ is the image of S_0', formed by the mirror M_1, and S″ is the image of S_0', formed by the image mirror M_2'. Let us denote d the distance between the mirror M_1 and the image mirror M_2'. Then the distance of the two images S′ and S″ is $2d$.

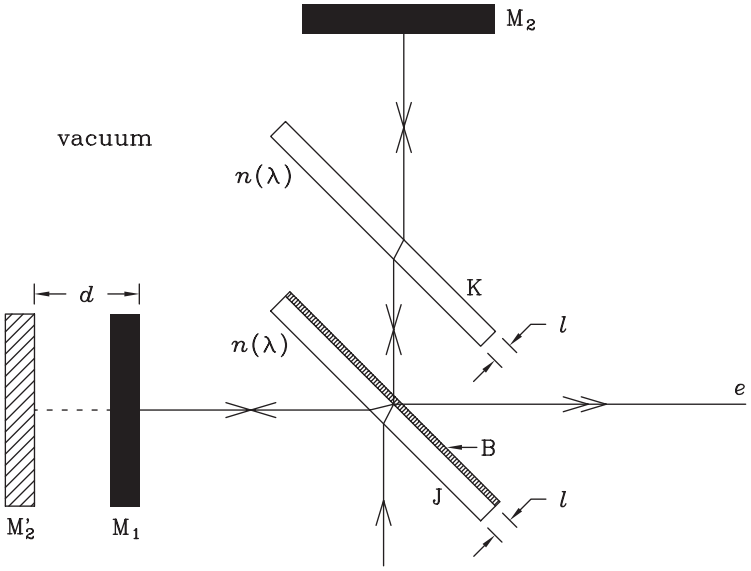


Figure 6.2: Optical arrangement of the Michelson interferometer. M_1 and M_2 are a fixed and a moving mirror, respectively. B is the half-transparent surface of a beamsplitter J, and K is a compensator plate. The beamsplitter and the compensator plate have the same thickness l and the same refractive index $n(\lambda)$. M_2' is the image of M_2, formed by the surface B. The distance of M_1 and M_2' is d. The ray leaves the interferometer along the path e, which is the optical axis.

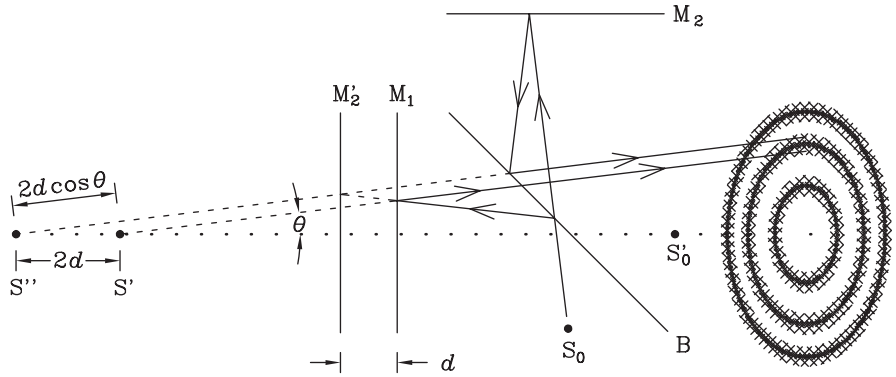


Figure 6.3: The interference pattern formed by a Michelson interferometer. The beamsplitting surface B forms the image M_2' of the moving mirror M_2, and the image S_0' of the light source S_0. S' is the image of S_0', formed by the fixed mirror M_1, and S'' is the image of S_0', formed by M_2'. The path difference of the two rays, travelling at the angle θ, is $2d\cos\theta$.

Figure 6.3 also shows the rays travelling at the angle θ with respect to the optical axis of the interferometer. The optical path difference of the rays coming from the two coherent images is $x = 2nd\cos\theta$, where n is the refractive index of the medium. If the beamsplitter is perfect, it divides the beam into two parts with equal intensity I_0. The output intensity of the interferometer is then, according to Equation 6.6,

$$I = 2I_0(1 + \cos\delta), \tag{6.7}$$

where the phase difference of the two rays is now

$$\delta = \frac{2\pi x}{\lambda} = \frac{2\pi}{\lambda}\, 2nd\cos\theta. \tag{6.8}$$

An interference fringe pattern is obtained. Intensity maxima are obtained with the condition

$$\boxed{\begin{array}{rclcrcl} \delta & = & l2\pi, & & l & = & 0, \pm1, \pm2, \pm3, \ldots, \\ 2nd\cos\theta & = & l\lambda, & & l & = & 0, \pm1, \pm2, \pm3, \ldots \end{array}} \tag{6.9}$$

If $\theta = 0$ (the rays propagate in the direction of the optical axis), $n = 1$ (the medium is vacuum), and S_0 is a monochromatic point source, then the optical path difference of the two rays is $x = 2d$, and the intensity is

$$F = 2I_0\left[1 + \cos(2\pi\nu\underbrace{2d}_{x})\right]. \tag{6.10}$$

This intensity in the optical axis of the interferometer is shown in Figure 6.4 as a function of the distance d. Maxima are obtained in the intervals $\lambda/2$ of the distance d, that is, in the intervals λ of the optical path difference x.

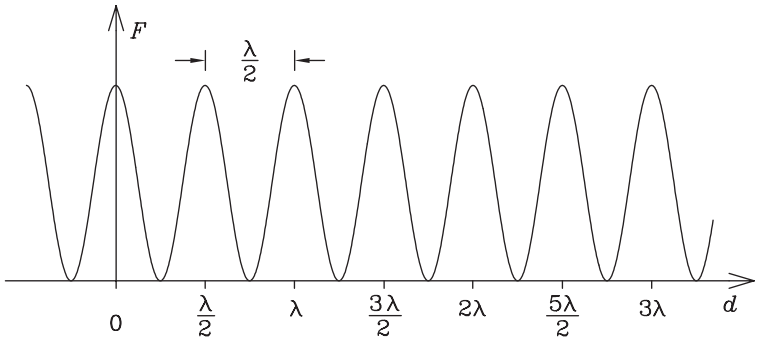


Figure 6.4: Output intensity F in the optical axis of a Michelson interferometer as a function of the distance d.

Let us now assume that the spectrum of the light source S_0 is continuous and consists of a wide band of wavenumbers, so that the spectrum of the beams in both branches of the interferometer is $E(\nu)$, shown in Figure 6.5. We shall assume that the source is still a point source. At a given optical path difference x, the interference signal from the infinitesimal spectral element between ν and $\nu + \mathrm{d}\nu$ is, according to Equation 6.10,

$$\mathrm{d}F(x, \nu) = 2E(\nu)\,[1 + \cos(2\pi\nu x)]\,\mathrm{d}\nu. \tag{6.11}$$

Consequently, the total signal from the whole spectral band is

$$F(x) = 2\int_0^\infty E(\nu)\,[1 + \cos(2\pi\nu x)]\,\mathrm{d}\nu. \tag{6.12}$$

$F(x)$ is called an *interference record*. An interference record is the total interference signal of the whole spectral band, measured as the function of the optical path difference x. A typical interference record is shown in Figure 6.6.

If we subtract the constant term

$$\frac{1}{2}F(0) = 2\int_0^\infty E(\nu)\,\mathrm{d}\nu \tag{6.13}$$

from the interference record $F(x)$, we obtain

$$I(x) = F(x) - \frac{1}{2}F(0) = 2\int_0^\infty E(\nu)\cos(2\pi\nu x)\,\mathrm{d}\nu. \tag{6.14}$$

$I(x)$ is called an *interferogram*.

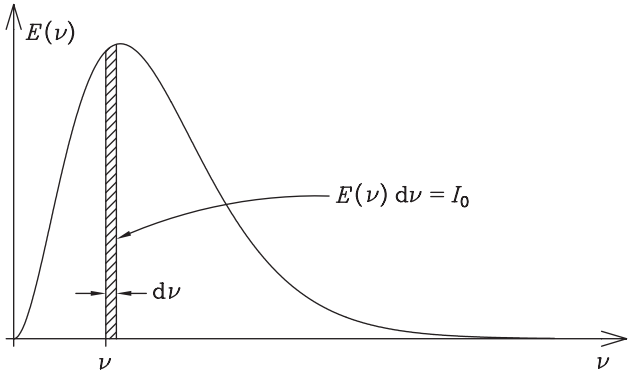


Figure 6.5: A wide-band continuous spectrum $E(\nu)$ and an infinitesimal monochromatic section of the width $\mathrm{d}\nu$.

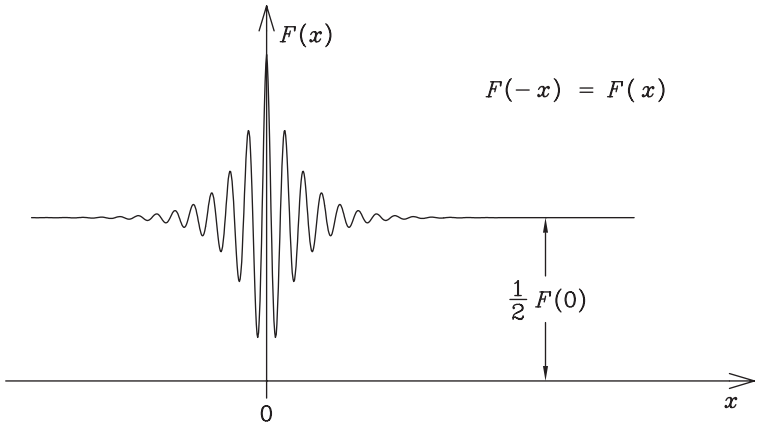


Figure 6.6: An interference record $F(x)$.

If we define $E(-\nu) = E(\nu)$, the computation is simplified, and we obtain

$$I(x) = \int_{-\infty}^{\infty} E(\nu)\cos(2\pi\nu x)\,\mathrm{d}\nu = \int_{-\infty}^{\infty} E(\nu)e^{i2\pi\nu x}\,\mathrm{d}\nu = \mathcal{F}\{E(\nu)\}, \tag{6.15}$$

where $\mathcal{F}$ is the Fourier transform. Hence, $I(x)$ and $E(\nu)$ form a Fourier transform pair, and they can be written as

$$\boxed{\begin{aligned} I(x) &= \int_{-\infty}^{\infty} E(\nu)e^{i2\pi\nu x}\,\mathrm{d}\nu &&= \mathcal{F}\{E(\nu)\}, \\ E(\nu) &= \int_{-\infty}^{\infty} I(x)e^{-i2\pi\nu x}\,\mathrm{d}x &&= \mathcal{F}^{-1}\{I(x)\}. \end{aligned}} \tag{6.16}$$

By denoting $E(\nu)$ the spectrum in one branch of the interferometer, in which case the whole spectrum of the source is $2E(\nu)$, we avoided extra coefficients in front of the integrals. We can also see that

$$\mathcal{F}\{\mathcal{F}^{-1}\{I(x)\}\} = \mathcal{F}^{-1}\{\mathcal{F}\{I(x)\}\} = I(x). \tag{6.17}$$

Since the signal of the interferometer as the function of the optical path difference and the spectrum as a function of wavenumber form a Fourier transform pair, we can apply all the properties of Fourier transforms in the calculation of the spectrum. We have all the theorems of Chapter 2 at our disposal, if we merely *replace t by the distance x* [cm] *and f by the wavenumber ν* $\left[\mathrm{cm}^{-1}\right]$.

In FTS, the output of a measurement is generally a signal in x-domain, and the most interesting information is obtained from the spectrum in ν-domain, which is calculated from the signal by inverse Fourier transform.

Example 6.1: The data collection system of a Fourier transform spectrometer makes a phase error of 5 μm. This means that every measured optical path difference value is 5 μm too large. How does this distort the spectral lines at wavenumbers 1 000 cm^{-1}, 1 500 cm^{-1} and 2 000 cm^{-1}?

Solution. The spectrum is $E(\nu) = \mathcal{F}^{-1}\{I(x)\}$. Let us denote the measured optical path difference x. The true optical path difference is $x - \Delta x$. The spectrum with the phase error is, applying the shift theorem,

$$E'(\nu) = \mathcal{F}^{-1}\{I'(x)\} = \mathcal{F}^{-1}\{I(x - \Delta x)\} = e^{-i2\pi\,\Delta x \nu}\, E(\nu).$$

At $\nu = 1\,000$ cm^{-1}: $e^{-i2\pi\,\Delta x \nu} = e^{-i\pi} = -1$.
At $\nu = 1\,500$ cm^{-1}: $e^{-i2\pi\,\Delta x \nu} = e^{-i3\pi/2} = i$.
At $\nu = 2\,000$ cm^{-1}: $e^{-i2\pi\,\Delta x \nu} = e^{-i2\pi} = 1$.

6.3 Sampling and truncation in FTS

In Chapter 3 we discussed the general theory of discrete Fourier transforms. In the following, we shall examine the same phenomena in the case of FTS. This section is a good example how well the general theory fits a true application.

In practice, we cannot compute the spectrum by Equation 6.16, because the interferogram $I(x)$ is sampled only at a set of discrete points with a sampling interval Δx:

$$x_j = j\,\Delta x, \qquad j = -N, -N+1, -N+2, \ldots, -1, 0, 1, \ldots, N-1. \tag{6.18}$$

The data are collected only in the finite region from $x = -L$ to $x = L$, where $L = N\,\Delta x$. We shall denote the interferogram samples by $I(x_j) = I_j$. We shall have to apply the discrete Fourier spectrum, where the integral is replaced by a sum, that is, $\int_{-\infty}^{\infty} \mathrm{d}x \to \Delta x \sum_{-N}^{N-1}$.

The spectrum is then given by

$$E_L^{\Delta x}(\nu) = \Delta x \sum_{j=-N}^{N-1} I_j e^{-i2\pi\nu j\,\Delta x}. \tag{6.19}$$

A truncated, continuous interferogram $I_L(x)$ can be written as

$$I_L(x) = \Pi_{2L}(x) I(x), \tag{6.20}$$

where Π_{2L} is the boxcar function

$$\Pi_{2L}(x) = \begin{cases} 1, & |x| \le L, \\ 0, & |x| > L. \end{cases} \tag{6.21}$$

The spectrum computed from the truncated, continuous interferogram is

$$\begin{aligned} E_L(\nu) &= \mathcal{F}^{-1}\{\Pi_{2L}(x)I(x)\} = \mathcal{F}^{-1}\{\Pi_{2L}(x)\} * \mathcal{F}^{-1}\{I(x)\} \\ &= W_L(\nu) * E(\nu) = \int_{-\infty}^{\infty} W_L(u)E(\nu-u)\,\mathrm{d}u, \end{aligned} \tag{6.22}$$

where

$$W_L(\nu) = \mathcal{F}^{-1}\{\Pi_{2L}(x)\} = \int_{-\infty}^{\infty} \Pi_{2L}(x)e^{-i2\pi\nu x}\,\mathrm{d}x = 2L\,\mathrm{sinc}(2\pi\nu L). \tag{6.23}$$

The spectrum computed from a sampled, infinitely long (not truncated) interferogram is

$$E^{\Delta x}(\nu) = \Delta x \sum_{j=-\infty}^{\infty} I_j e^{-i2\pi\nu j\,\Delta x}. \tag{6.24}$$

Remembering the properties of the discrete Fourier transform, we can see that the discrete sampling of an interferogram with the sampling interval Δx gives a periodic spectrum $E^{\Delta x}(\nu)$ with the period $1/(\Delta x)$, that is, $E^{\Delta x}\left(\nu - \frac{k}{\Delta x}\right) = E^{\Delta x}(\nu)$ with all integers k. The spectrum computed from the sampled, not truncated interferogram is

$$\boxed{E^{\Delta x}(\nu) = \sum_{k=-\infty}^{\infty} E\left(\nu - \frac{k}{\Delta x}\right).} \tag{6.25}$$

Combining the effects of truncation and discrete sampling, we obtain the spectrum $E_L^{\Delta x}(\nu)$ of a sampled, truncated interferogram

$$\begin{aligned} E_L^{\Delta x}(\nu) &= \sum_{k=-\infty}^{\infty} E_L\left(\nu - \frac{k}{\Delta x}\right) = \sum_{k=-\infty}^{\infty} \delta\left(\nu - \frac{k}{\Delta x}\right) * E_L(\nu) \\ &\overset{(6.22)}{=} \sum_{k=-\infty}^{\infty} \delta\left(\nu - \frac{k}{\Delta x}\right) * \underbrace{2L\,\mathrm{sinc}(2\pi\nu L)}_{W_L(\nu)} * E(\nu) \\ &= \left\{\sum_{k=-\infty}^{\infty} 2L\,\mathrm{sinc}\left[2\pi\left(\nu - \frac{k}{\Delta x}\right)L\right]\right\} * E(\nu) \\ &= W_L^{\Delta x}(\nu) * E(\nu), \end{aligned} \tag{6.26}$$

where

$$\boxed{W_L^{\Delta x}(\nu) = \sum_{k=-\infty}^{\infty} 2L\,\mathrm{sinc}\left[2\pi\left(\nu - \frac{k}{\Delta x}\right)L\right]} \tag{6.27}$$

is the instrumental profile due to the truncation of the interferogram and the discrete sampling. It can also be called the *instrumental function of a Fourier transform spectrometer*.

Let us examine the monochromatic spectrum consisting of spectral lines at the wavenumbers $+\nu_0$ and $-\nu_0$. The spectrum is

$$E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0). \tag{6.28}$$

The corresponding signal is the interferogram

$$I(x) = \mathcal{F}\{E(\nu)\} = e^{i2\pi\nu_0 x} + e^{-i2\pi\nu_0 x} = 2\cos(2\pi\nu_0 x). \tag{6.29}$$

The monochromatic spectrum and the signal, which is a cosine wave, are shown in Figure 6.7. If this signal is recorded only from $x = -L$ to $x = L$, that is, the interferogram is truncated, then the computed spectrum becomes

$$\begin{aligned} E_L(\nu) &= W_L(\nu) * [\delta(\nu - \nu_0) + \delta(\nu + \nu_0)] \\ &= W_L(\nu - \nu_0) + W_L(\nu + \nu_0) \\ &= 2L\,\mathrm{sinc}\,[2\pi(\nu - \nu_0)L] + 2L\,\mathrm{sinc}\,[2\pi(\nu + \nu_0)L]\,. \end{aligned} \tag{6.30}$$

Thus, because of truncation, the computed spectrum consists of two sinc functions instead of two sharp lines, as shown in Figure 6.8.

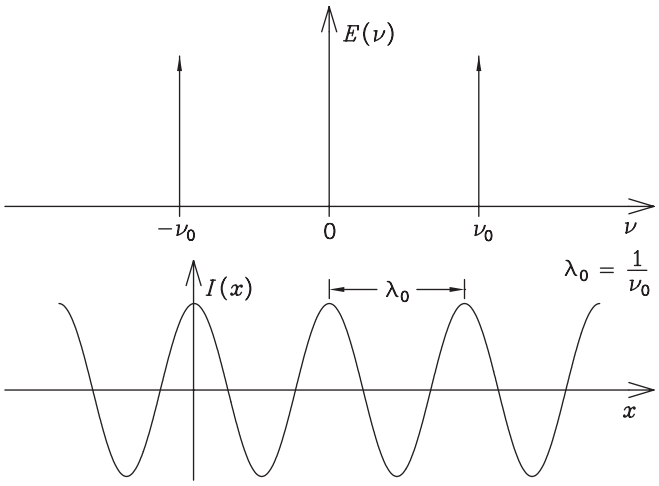


Figure 6.7: Monochromatic spectrum $E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0)$, and its Fourier transform, the signal $I(x) = 2\cos(2\pi\nu_0 x)$.

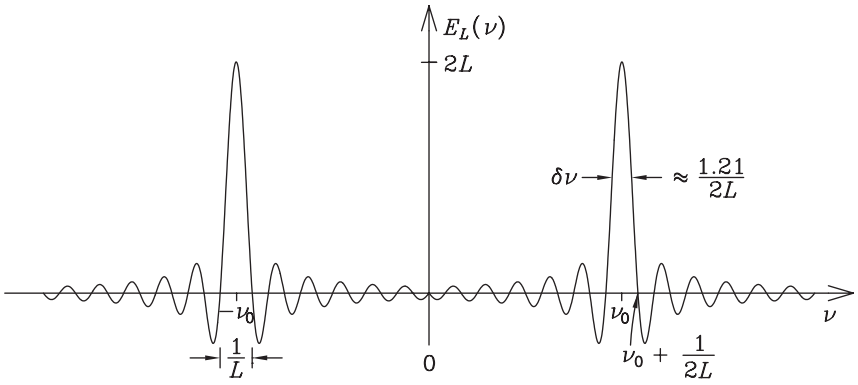


Figure 6.8: A spectrum $E_L(\nu)$ which consists of sinc functions instead of monochromatic lines, because the signal has been recorded only in the finite region from $-L$ to L. The theoretical resolution is approximately $1.207/(2L)$.

The *theoretical resolution* of a spectrum, if L is the maximum optical path difference of the interferometer, is the full width at half maximum, FWHM, of the sinc function, which is

$$\boxed{\delta\nu \approx 1.207/(2L).} \tag{6.31}$$

If the truncated cosine wave is measured only at discrete points, then the computed spectrum becomes

$$\begin{aligned} E_L^{\Delta x}(\nu) &= W_L^{\Delta x}(\nu) * [\delta(\nu+\nu_0) + \delta(\nu-\nu_0)] = W_L^{\Delta x}(\nu-\nu_0) + W_L^{\Delta x}(\nu+\nu_0) \\ &= \sum_{k=-\infty}^{\infty} 2L\,\mathrm{sinc}\left[2\pi\left(\nu+\nu_0-\frac{k}{\Delta x}\right)L\right] \\ &\quad + \sum_{l=-\infty}^{\infty} 2L\,\mathrm{sinc}\left[2\pi\left(\nu-\nu_0-\frac{l}{\Delta x}\right)L\right]. \end{aligned} \tag{6.32}$$

This spectrum is shown in Figure 6.9. It consists of two sets of periodic series of sinc functions $\sum_k$ and $\sum_l$ with the period $1/(\Delta x)$.

A true spectrum $E(\nu)$ may have, instead of monochromatic lines, a wide spectral band in some wavenumber region $(-\nu_{\max}, \nu_{\max})$, as shown in Figure 6.10. Let us, in addition, assume that the length of the sampling region is large: $1/L \ll 2\nu_{\max}$. In this case, the sinc-widening (see Equation 6.31) of a wide-band spectrum due to signal truncation is negligible, and

$$W_L(\nu) * E(\nu) = 2L\,\mathrm{sinc}(2\pi\nu L) * E(\nu) \approx E(\nu).$$

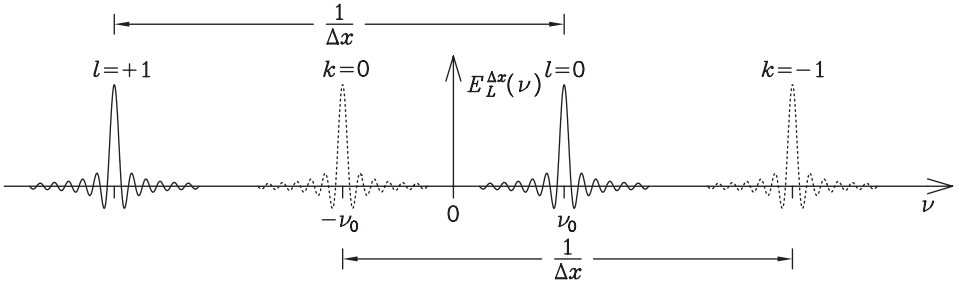


Figure 6.9: The spectrum $E_L^{\Delta x}(\nu)$, which is obtained, if the true spectrum is the monochromatic spectrum $\delta(\nu + \nu_0) + \delta(\nu - \nu_0)$, but the signal has been truncated and sampled at discrete points. The spectrum consists of the two series $\sum_{k=-\infty}^{\infty} 2L \operatorname{sinc}\left[2\pi\left(\nu + \nu_0 - \frac{k}{\Delta x}\right)L\right]$ and $\sum_{l=-\infty}^{\infty} 2L \operatorname{sinc}\left[2\pi\left(\nu - \nu_0 - \frac{l}{\Delta x}\right)L\right]$.

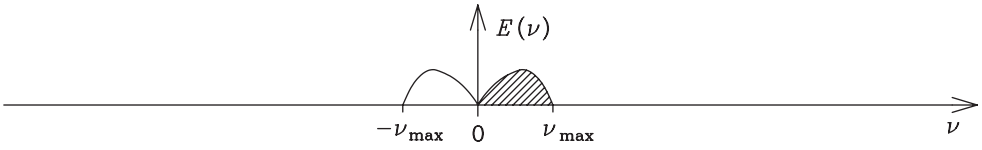


Figure 6.10: Spectrum $E(\nu)$ which has a wide band in the wavenumber region $(-\nu_{\max}, \nu_{\max})$.

If the sampling period is short, then the spectral band is not wider than the period of the computed spectrum caused by discrete sampling, that is, $2\nu_{\max} \leq 1/(\Delta x)$. In this case, the computed spectrum $E_L^{\Delta x}(\nu)$ consists of periodically recurring functions

$$W_L(\nu) * E(\nu) = 2L \operatorname{sinc}(2\pi\nu L) * E(\nu),$$

as shown in Figure 6.11.

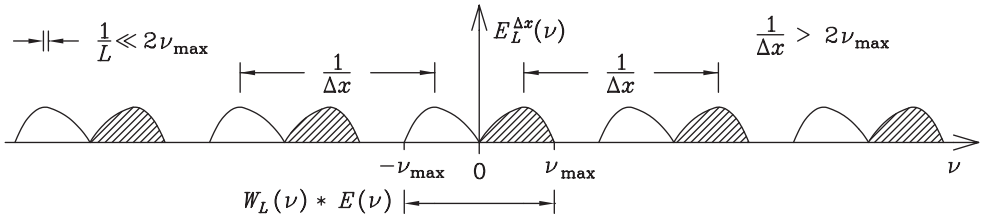


Figure 6.11: The spectrum $E_L^{\Delta x}(\nu)$, if $1/(\Delta x) > 2\nu_{\max}$. The spectrum in the region $(-\nu_{\max}, \nu_{\max})$ is $W_L(\nu) * E(\nu) = 2L\,\mathrm{sinc}(2\pi\nu L) * E(\nu) \approx E(\nu)$, if $1/L \ll \nu_{\max}$.

The critical sampling interval of an interferometric signal is the inverse of the Nyquist frequency:

$$\boxed{(\Delta x)_{\text{Nyquist}} = \frac{1}{2\nu_{\max}}}\,, \tag{6.33}$$

where $\nu_{\max}$ is the maximum wavenumber of the spectral band. If this optimal sampling interval is used, then the computed spectrum $E_L^{\Delta x}(\nu)$ contains no free regions, as demonstrated in Figure 6.12. If the sampling interval is larger, that is, $\Delta x > 1/(2\nu_{\max})$, then the spectrum $E_L^{\Delta x}(\nu)$ is distorted due to aliasing, because the different spectral orders (different l and k) overlap (see Figure 3.8).

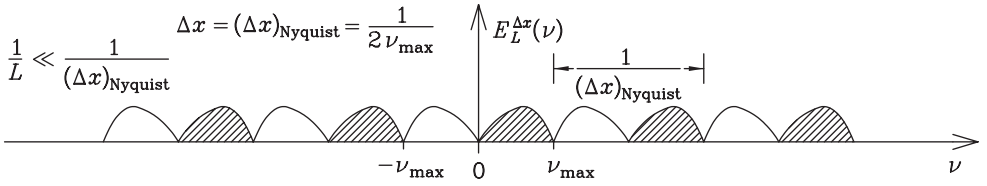


Figure 6.12: The spectrum $E_L^{\Delta x}(\nu)$, when the critical sampling interval $\Delta x = (\Delta x)_{\text{Nyquist}} = 1/(2\nu_{\max})$ is used.

6.4 Collimated beam and extended light source

In a practical interferometer, light coming from the source is collected by a collimator, which may be, for example, a convex lens, as shown in Figure 6.13. In this way, a larger proportion of the light energy can be utilized. In the figure, the collimator lens L_1 turns the spherical wave coming from the source into a plane wave. In front of the detector there is another lens L_2 which focuses the plane wave to the detector. The interferometer in Figure 6.13 with two convex lenses is called the Twyman–Green interferometer.

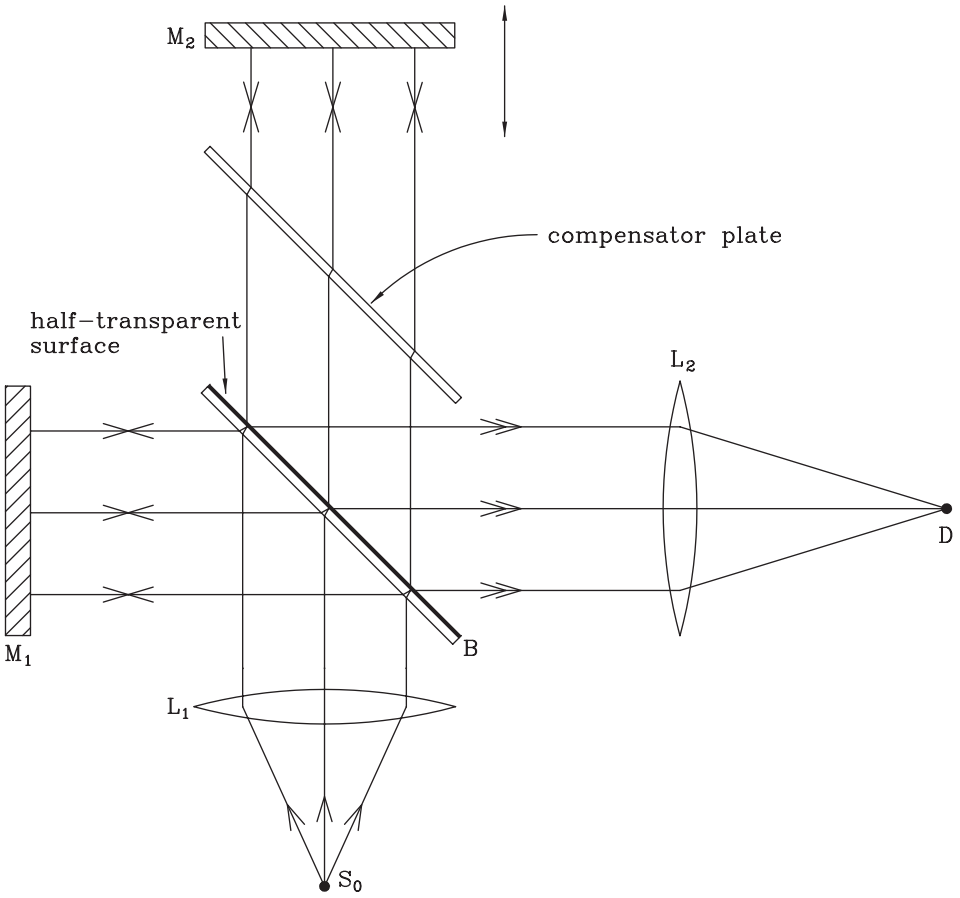


Figure 6.13: The optical layout of the Twyman–Green interferometer. S_0 is a light source, D is a detector, B is a beam splitter, M_1 is a fixed mirror, and M_2 is a moving mirror. L_1 and L_2 are convex lenses.

Instead of a lens, the collimator may be a mirror. Figure 6.14 shows an interferometer whose collimator is a parabolic mirror which is off-axis, that is, the mirror is out of the center point of the paraboloid. The mirror focusing the beam to the detector is another off-axis parabolic mirror.

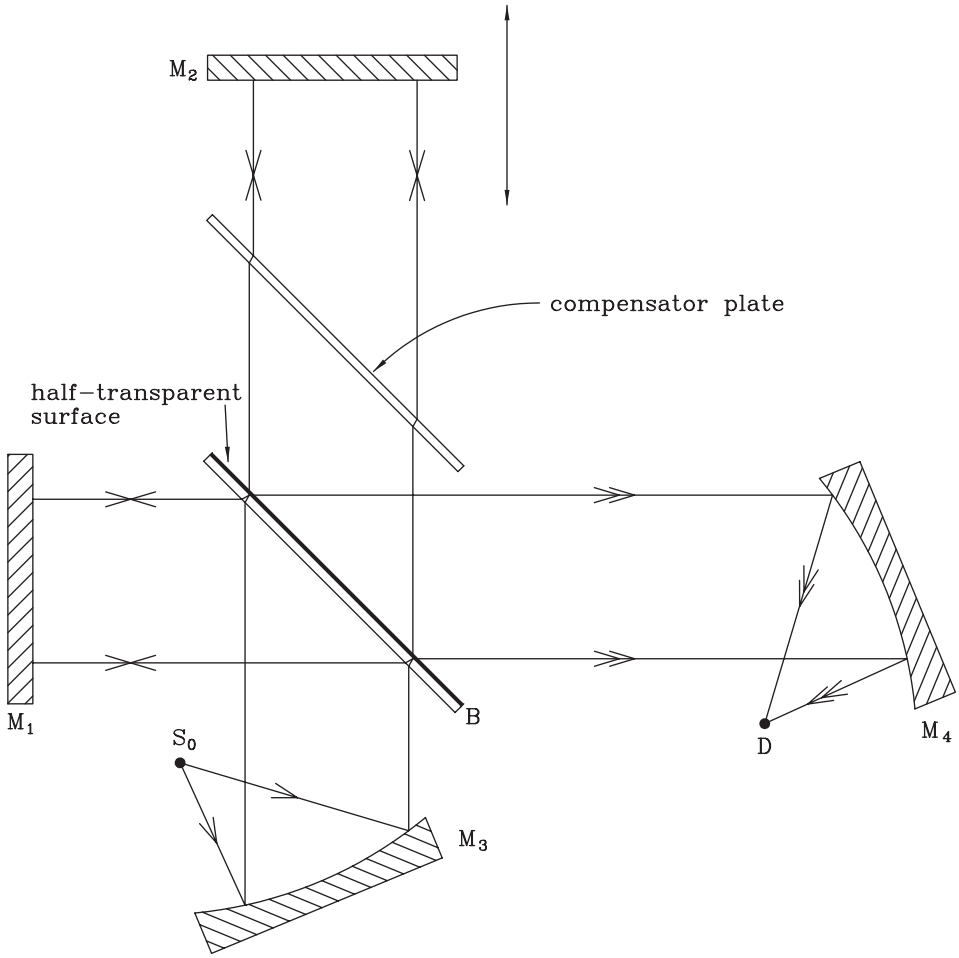


Figure 6.14: An interferometer with off-axis parabolic mirrors. S_0 is a light source, D is a detector, B is a beam splitter, M_1 is a fixed mirror, M_2 is a moving mirror, and M_3 and M_4 are fixed off-axis parabolic mirrors.

In practice, the light source of an interferometer is not a point source but an extended source, which has a finite area a, as shown in Figure 6.15. If A denotes the area of the collimator mirror or lens, f the focal length of the collimator, $\Omega_0 = A/f^2$ the solid angle of the collimator as seen from the source, and $\Omega = a/f^2$ the solid angle of the source as seen from the collimator, then the intensity of the signal entering the interferometer is proportional to

$$aA/f^2 = \Omega A = a\Omega_0. \tag{6.34}$$

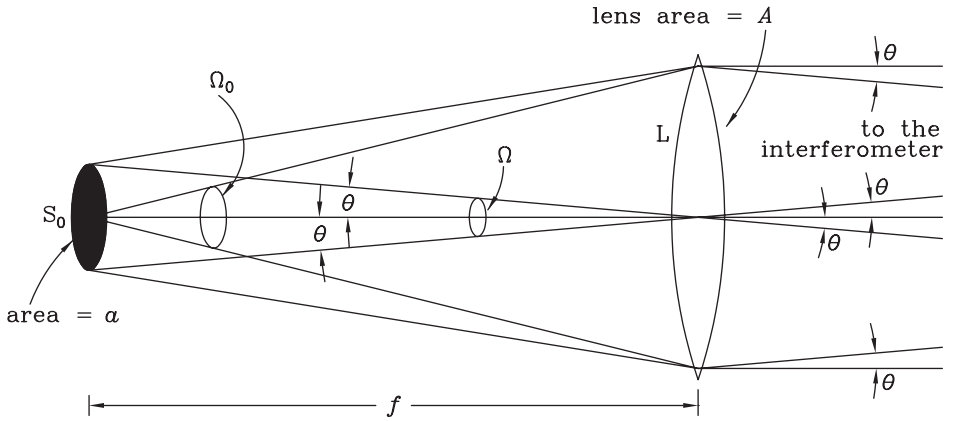


Figure 6.15: Extended radiation source S_0 at the focal point of the collimator lens L with the focal length f. The area of the source is a, and the area of the lens is A. Ω_0 is the solid angle of the lens, and $\Omega = \pi\theta^2$ is the solid angle of the source.

If the optical path difference of the two rays of the interferometer in the optical axis is $x = 2d$, then the optical path difference of the rays travelling in the direction α with respect to the optical axis is (see Examples 6.2 and 6.3)

$$\underbrace{x}_{2d}\cos\alpha \approx x\left(1 - \frac{\Omega'}{2\pi}\right), \tag{6.35}$$

where $\Omega' = \pi\alpha^2$. The direction α of the ray coming from the source may vary from 0 to the maximum θ which fulfills the condition $\Omega = \pi\theta^2$. As the direction α varies $0 \to \theta$, then the solid angle Ω' varies $0 \to \Omega$. The total interferogram is obtained by integrating over the source (see Example 6.4):

$$I_\Omega(x) = \int_0^\Omega I\left[x\left(1 - \frac{\Omega'}{2\pi}\right)\right] \mathrm{d}\Omega' = \Omega \int_{-\infty}^{\infty} E(\nu)\,\mathrm{sinc}\left(\frac{\nu x \Omega}{2}\right) e^{i2\pi\nu x\left(1-\frac{\Omega}{4\pi}\right)}\,\mathrm{d}\nu. \tag{6.36}$$

If we compare this interferogram to the interferogram of a point source in Equation 6.16,

$$I(x) = \int_{-\infty}^{\infty} E(\nu) e^{i2\pi\nu x}\,\mathrm{d}\nu, \tag{6.37}$$

we can see that one effect caused by the extended source is *a shift of the measured wavenumber*. The true wavenumber can be obtained from the measured wavenumber:

$$\nu_{\text{true}} = \frac{\nu_{\text{measured}}}{\left(1 - \frac{\Omega}{4\pi}\right)}. \tag{6.38}$$

Let us examine how a monochromatic spectrum behaves if the source is extended. If the true spectrum is $E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0)$, then Equation 6.36 gives

$$I_\Omega(x) = \Omega \left\{ \mathrm{sinc}\left(\frac{\nu_0 x \Omega}{2}\right) e^{i2\pi\nu_0 x\left(1-\frac{\Omega}{4\pi}\right)} + \mathrm{sinc}\left(\frac{\nu_0 x \Omega}{2}\right) e^{-i2\pi\nu_0 x\left(1-\frac{\Omega}{4\pi}\right)} \right\}. \quad (6.39)$$

Remembering that

$$\mathcal{F}^{-1}\{2L\,\mathrm{sinc}(2\pi\nu L)\} = \begin{cases} 1, & |x| \le L, \\ 0, & |x| > L, \end{cases}$$

the spectrum computed from the interferogram can be written as

$$\begin{aligned} E_\Omega(\nu) &= \mathcal{F}^{-1}\{I_\Omega(x)\} \\ &= \Pi_\Omega(\nu, \nu_0) * \left\{ \delta\left[\nu + \nu_0\left(1 - \frac{\Omega}{4\pi}\right)\right] + \delta\left[\nu - \nu_0\left(1 - \frac{\Omega}{4\pi}\right)\right] \right\}, \end{aligned} \quad (6.40)$$

where the boxcar function

$$\Pi_\Omega(\nu, \nu_0) = \mathcal{F}^{-1}\{\Omega\,\mathrm{sinc}(\nu_0 x \Omega/2)\} = \begin{cases} \dfrac{2\pi}{\nu_0}, & |\nu| \le \dfrac{\nu_0\Omega}{4\pi}, \\ 0, & |\nu| > \dfrac{\nu_0\Omega}{4\pi}. \end{cases}$$

This spectrum is shown in Figure 6.16. Due to the finite size of the source, the monochromatic peaks are seen as boxes. The width of the boxes is $\nu_0\Omega/(2\pi)$, and their height is $2\pi/\nu_0$.

We can see that another effect caused by the extended source, besides the shift of the measured wavenumber, is the broadening of the spectral lines. This is called the *aperture broadening*.

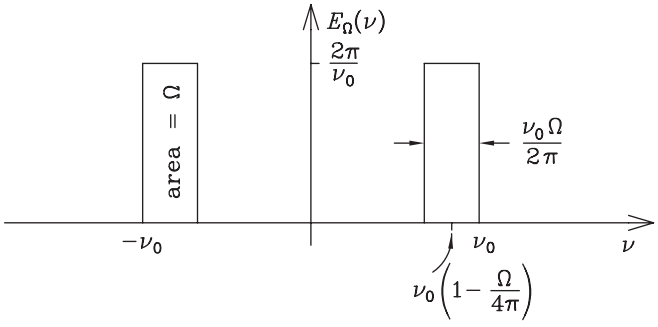


Figure 6.16: The observed spectrum, if the true spectrum is monochromatic $E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0)$, and the solid angle of the circular source is Ω. Due to the extended size of the source, monochromatic peaks are seen as boxes.

Further, we can write the monochromatic spectrum in the case of an extended source as

$$\begin{aligned} E_\Omega(\nu) &= \Pi_\Omega(\nu, \nu_0) * \left\{ \delta\left(\nu - \frac{\nu_0\Omega}{4\pi}\right) * \delta(\nu + \nu_0) + \delta\left(\nu + \frac{\nu_0\Omega}{4\pi}\right) * \delta(\nu - \nu_0) \right\} \\ &= W_\Omega(\mp\nu_0, \nu) * \underbrace{\delta(\nu \pm \nu_0)}_{E(\nu)}, \end{aligned} \tag{6.41}$$

where $W_\Omega(\mp\nu_0, \nu)$ is the instrumental function of an extended source

$$W_\Omega(\mp\nu_0, \nu) = \Pi_\Omega(\nu, \nu_0) * \delta\left(\nu \mp \frac{\nu_0\Omega}{4\pi}\right). \tag{6.42}$$

This function is shown in Figure 6.17. $W_\Omega(-\nu_0, \nu)$ operates in the negative wavenumber region and $W_\Omega(+\nu_0, \nu)$ in the positive wavenumber region. The Dirac's delta functions in Equation 6.41 can be replaced by a general narrow-band ($\nu_0 \approx$ constant) spectrum $E(\nu)$, and the equation is still approximately valid. Note, however, that $W_\Omega(\pm\nu_0, \nu)$ also depends on the line position ν_0, since aperture broadened lines are the narrower and the higher, the closer to the origin they are. Therefore, Equation 6.41 can only be applied to a spectral piece $E(\nu)$ which is very narrow compared with its distance from the origin.

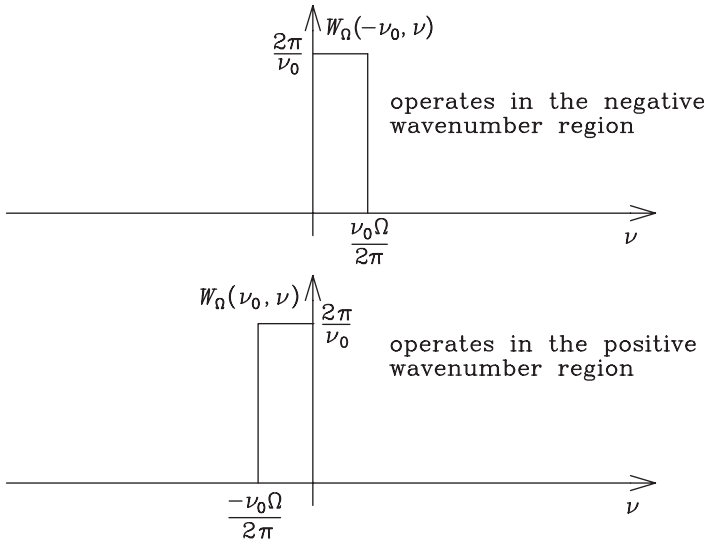


Figure 6.17: The instrumental function of an extended source is $W_\Omega(-\nu_0, \nu)$ in the negative wavenumber region and $W_\Omega(\nu_0, \nu)$ in the positive wavenumber region.

Let us now combine the effects of discrete sampling, truncation, and an extended source. The observed spectrum can be written as

$$\begin{aligned} E^{\Delta x}_{L,\Omega}(\nu) &= W_\Omega(\mp\nu_0, \nu) * E^{\Delta x}_L(\nu) \\ &= W_\Omega(\mp\nu_0, \nu) * W^{\Delta x}_L(\nu) * E(\nu) \\ &= W^{\Delta x}_{L,\Omega}(\nu) * E(\nu), \end{aligned} \tag{6.43}$$

where $W^{\Delta x}_{L,\Omega}(\nu)$ is the total instrumental function of a collimated interferometer,

$$\boxed{W^{\Delta x}_{L,\Omega}(\nu) = W_\Omega(\mp\nu_0, \nu) * \left\{ \sum_{k=-\infty}^{\infty} 2L \operatorname{sinc}\left[2\pi\left(\nu - \frac{k}{\Delta x}\right)L\right]\right\}.} \tag{6.44}$$

If the true spectrum is the monochromatic spectrum $E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0)$, then the spectrum given by a collimated interferometer is

$$\begin{aligned} E^{\Delta x}_{L,\Omega}(\nu) &= \sum_{k=-\infty}^{\infty} W_\Omega(-\nu_0, \nu) * 2L \operatorname{sinc}\left[2\pi\left(\nu + \nu_0 - \frac{k}{\Delta x}\right)L\right] \\ &\quad + \sum_{l=-\infty}^{\infty} W_\Omega(\nu_0, \nu) * 2L \operatorname{sinc}\left[2\pi\left(\nu - \nu_0 - \frac{l}{\Delta x}\right)L\right]. \end{aligned}$$

This spectrum is shown in Figure 6.18.

The *instrumental resolution* $\delta\nu$ is determined by the FWHM of the function

$$W_{L,\Omega}(\nu) = W_\Omega(\nu_0, \nu) * 2L \operatorname{sinc}(2\pi\nu L).$$

If the source is large and the interferogram is long, then the FWHM of $W_\Omega(\nu_0, \nu)$, which is $\nu_0\Omega/(2\pi)$, determines the instrumental resolution. If the source is small and the interferogram is short, then the FWHM of $2L\operatorname{sinc}(2\pi\nu L)$, which is $1.21/(2L)$, gives the instrumental resolution.

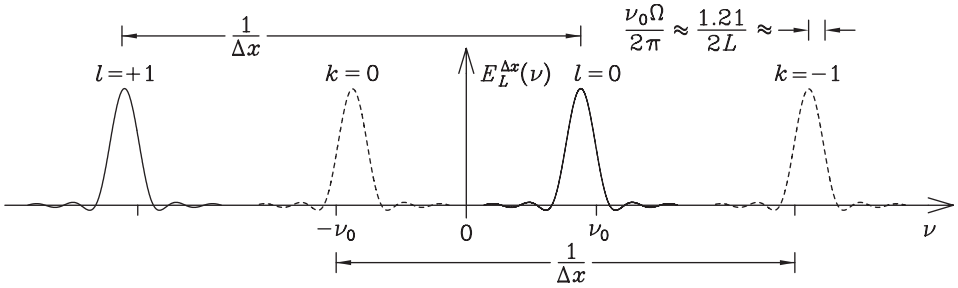


Figure 6.18: The spectrum $E^{\Delta x}_{L,\Omega}(\nu)$ given by a collimated interferometer if the true spectrum is $E(\nu) = \delta(\nu - \nu_0) + \delta(\nu + \nu_0)$.

The light intensity entering the detector is proportional to Ω (Equation 6.34), and it is naturally favorable to try to keep the signal-to-noise ratio large. An optimal situation is achieved when the aperture broadening and the truncation broadening have approximately equal widths, that is,

$$\boxed{\frac{\nu_0 \Omega}{2\pi} \approx \frac{1.21}{2L}}. \tag{6.45}$$

The *optimal truncation* of the signal can be found by calculating the truncation point L from this equation.

Figure 6.19 shows the instrumental line shape function $W_{L,\Omega}(\nu)$ with three different Ω values. The optimum situation is almost that in the second row.

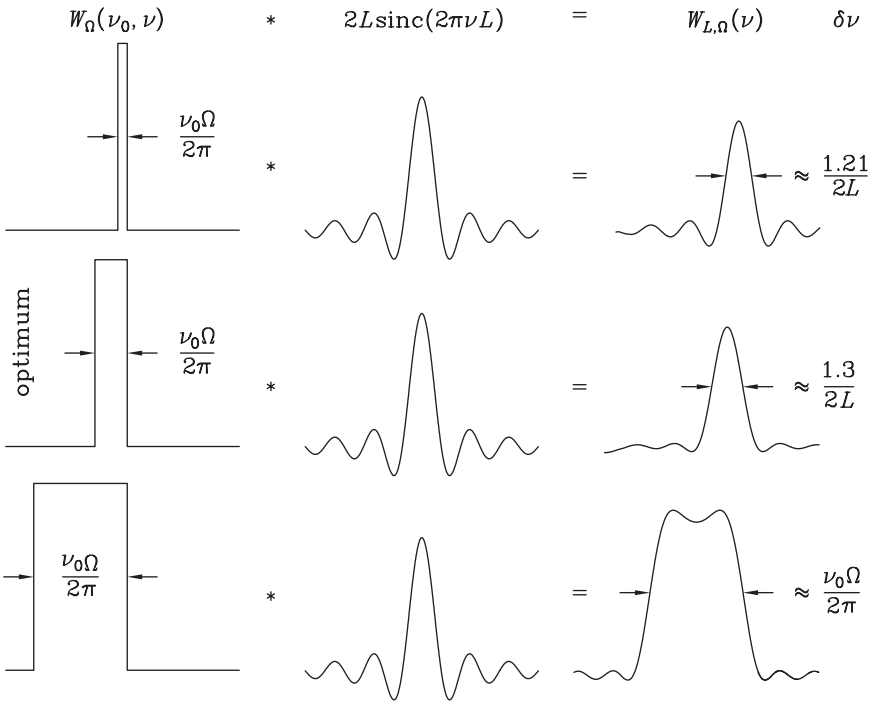


Figure 6.19: The total instrumental line shape function $W_{L,\Omega}(\nu)$ with three different Ω values. The instrumental resolution $\delta\nu$ is determined by the FWHM of the convolution $W_{L,\Omega}(\nu) = W_\Omega(\nu_0, \nu) * 2L\,\mathrm{sinc}(2\pi\nu L)$.

Example 6.2: The properties of a Michelson interferometer can be examined by finding the image of one of the mirrors formed by the beamsplitter in such a way that this image lies in the same direction as the other mirror. This was illustrated in Figures 6.2 and 6.3. We must imagine that one of the rays, ray 1, sees only the mirror M_1, and the second ray, ray 2, sees only the mirror M_2. (a) The rays arrive in the mirrors at the angle α with respect to the optical axis. What is the optical path difference between the two rays? (b) What is the optical path difference, if the rays arrive in the mirrors at the angle α, but the mirrors are cube-corner mirrors? A cube-corner mirror consists of three plane mirrors at direct angles with respect to each other. This kind of mirror is used in many interferometers.

Solution. (a) This situation is shown in Figure 6.20 (a). If the distance of the mirrors is d, then the difference of the two paths is $2d\cos\alpha$. (b) It is sufficient to examine the two-dimensional case of Figure 6.20 (b). If the distance of the mirrors is d, then the difference of the two paths is $2d\cos\alpha$, again.

This result means that if a ray inside the interferometer has a direction error α, then a wavenumber ν_0 is erroneously registered as the wavenumber $\nu_0\cos\alpha$.

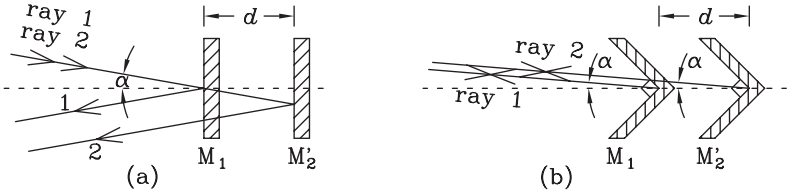


Figure 6.20: (a) The mirror M_1, and the image M_2' of the mirror M_2, of an interferometer. (b) The cube-corner mirror M_1, and the image M_2' of the cube-corner mirror M_2, of an interferometer. In both pictures, rays 1 and 2 arrive in the mirrors at the angle α. In a perfect cube-corner mirror, all rays are reflected back in exactly the same direction from which they came.

Example 6.3: Let us consider a small circle in a circular radiation source, shown in Figure 6.21. The center of the circle is the center of the radiation source. Let the radial angle of the circle be α, and the solid angle limited by the circle Ω'. Show that $\cos\alpha \approx 1 - \Omega'/(2\pi)$.

Solution. The radius of the circle is $r \approx f\alpha$. The area inside the circle is

$$a = \pi r^2 \approx \pi(f\alpha)^2.$$

The solid angle

$$\Omega' \approx \frac{a}{f^2} \approx \frac{\pi(f\alpha)^2}{f^2} = \pi\alpha^2.$$

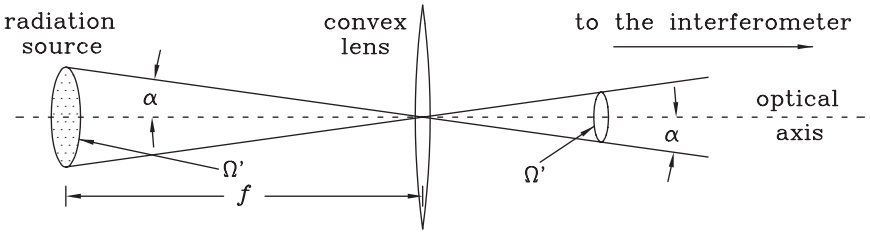


Figure 6.21: Part of a circular radiation source, limited by a circle of radial angle α as seen from the lens. The solid angle limited by the circle is Ω'.

Consequently,

$$\begin{aligned}\cos\alpha &\approx \frac{f}{\sqrt{(f\alpha)^2+f^2}} = \frac{1}{\sqrt{1+\alpha^2}} \\ &\approx \frac{1}{\sqrt{1+\Omega'/\pi}} \approx 1-\frac{\Omega'}{2\pi},\end{aligned}$$

since the serial expansion $(1+x)^{-1/2} = 1 - \frac{1}{2}x + \frac{1\cdot 3}{2\cdot 4}x^2 - \cdots$, if $-1 < x \leq 1$.

Example 6.4: Derive Equation 6.36,

$$I_\Omega(x) = \Omega \int_{-\infty}^{\infty} E(\nu)\,\mathrm{sinc}\left(\frac{\nu x\Omega}{2}\right) e^{i2\pi\nu x\left(1-\frac{\Omega}{4\pi}\right)}\,\mathrm{d}\nu,$$

where Ω is the solid angle of a circular source, $E(\nu)$ is the intensity of the source per unit solid angle and unit wavenumber, and $I_\Omega(x)$ is the intensity leaving the Michelson interferometer and arriving in the detector, as a function of the optical path difference x.

Solution. According to Equation 6.15, the intensity arriving in the detector if the source is a point source can be expressed as

$$I(x) = \int_{-\infty}^{\infty} E(\nu) e^{i2\pi\nu x}\,\mathrm{d}\nu.$$

Every point of an extended radiation source can be considered as a separate point source. If the radiation source is fully incoherent, we can expect that every source point follows the above equation separately. In addition, we must take into account the effect of the direction error (see the two previous examples): the optical path difference of radiation which is emitted

from a circle which makes an angle α with respect to the optical axis is

$$x_\alpha = x\cos\alpha \approx x(1-\frac{\Omega'}{2\pi}),$$

where Ω' is the solid angle limited by the circle. Consequently, the intensity which arrives in the detector from a circle at the angle α is

$$I_\alpha(x) = \int_{-\infty}^{\infty} E(\nu) e^{i2\pi\nu x_\alpha}\,\mathrm{d}\nu.$$

The whole intensity in the detector is obtained by integration over the whole solid angle of the radiation source:

$$\begin{aligned}
I_\Omega(x) &= \int_0^\Omega I_\alpha(x)\,\mathrm{d}\Omega' \\
&= \int_0^\Omega \int_{-\infty}^{\infty} E(\nu) e^{i2\pi\nu x(1-\frac{\Omega'}{2\pi})}\,\mathrm{d}\nu\mathrm{d}\Omega' \\
&= \int_{-\infty}^{\infty} \frac{E(\nu)}{-i\nu x}\left[e^{i2\pi\nu x(1-\frac{\Omega}{2\pi})} - e^{i2\pi\nu x}\right]\mathrm{d}\nu \\
&= \int_{-\infty}^{\infty} \frac{-E(\nu)}{i\nu x}(e^{-i\nu x\Omega/2} - e^{i\nu x\Omega/2})\, e^{i2\pi\nu x - i\nu x\Omega/2}\,\mathrm{d}\nu \\
&= \int_{-\infty}^{\infty} \frac{E(\nu)}{i\nu x} 2i\sin(\nu x\Omega/2)\, e^{i2\pi\nu x - i\nu x\Omega/2}\,\mathrm{d}\nu \\
&= \Omega\int_{-\infty}^{\infty} E(\nu)\,\mathrm{sinc}\left(\frac{\nu x\Omega}{2}\right) e^{i2\pi\nu x\left(1-\frac{\Omega}{4\pi}\right)}\,\mathrm{d}\nu.
\end{aligned}$$

6.5 Apodization

If the true spectral line is narrower than the aperture broadening $\nu_0\Omega/(2\pi)$, then the line shape which is given by an interferometer is $W_\Omega(\nu_0, \nu) * 2L\,\mathrm{sinc}(2\pi\nu L)$. The sinc function causes strong oscillation in the line shape, if $\nu_0\Omega/(2\pi) < 1.21/(2L)$. The sidelobes can be reduced by replacing the truncation function $\Pi(x)$ in Equation 6.20 by a window function (or weight function) $A(x)$, which equals unity at $x = 0$ and approaches zero smoothly at large values of $|x|$. This method is called *apodization*, and $A(x)$ is the *apodization function*. Instead of Equation 6.22, the spectrum is now

$$E_A(\nu) = \mathcal{F}^{-1}\{A(x)I(x)\} = \mathcal{F}^{-1}\{A(x)\} * E(\nu) = W_A(\nu) * E(\nu). \tag{6.46}$$

The simplest apodization function is the *triangular function*

$$\Lambda_L(x) = \begin{cases} 1 - \dfrac{|x|}{L}, & |x| \le L, \\ 0, & |x| > L, \end{cases} \tag{6.47}$$

the FWHM of which is L. It can be shown (see Problem 1.6 or 2.10) that the inverse Fourier transform of the triangular function is

$$W_A(\nu) = W_{\Lambda_L}(\nu) = L\,\mathrm{sinc}^2(\pi\nu L). \tag{6.48}$$

The spectral line shape in the case where apodization is made with the triangular apodization function is shown in Figure 6.22.

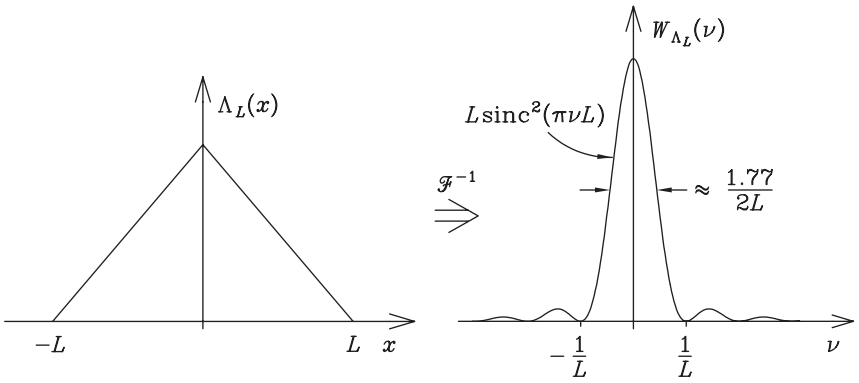


Figure 6.22: Triangular apodization function $A(x) = \Lambda_L(x)$, and its inverse Fourier transform, the line shape $W_{\Lambda_L}(\nu) = L\,\mathrm{sinc}^2(\pi\nu L)$.

6.6 Applications of FTS

Fourier transform spectroscopy is in frequent use in many various practical and scientific applications. Here, we shall review some of these applications.

The most important practical application of FT spectrometers is the *qualitative and quantitative analysis of unknown samples*. Conventionally, the analysis of a gas phase sample has taken place so that the user records the IR spectrum of the sample using some commercial instrument with its maximum resolution. Then, looking at the spectrum with the naked eye, or using some commercial search program based on the correlation between the observed and a library spectrum, the user tries to identify the unknown compound. The multicomponent analysis of gas mixtures has usually taken place by the so-called subtraction technique: the library spectrum multiplied by a constant is subtracted from the observed mixture spectrum. The constant, which gives the desired concentration of the component, is determined in such a way that the corresponding component just disappears after subtraction. This analysis method, however, requires high resolution (0.1 cm^{-1} – 0.5 cm^{-1}) and is slow and laborious to carry out. It requires an experienced user, and it is very difficult to automatize.

An example of recent development in gas analysis is the GASMET$^{\mathrm{TM}}$ FT-IR gas analyzer by Temet Instruments Inc. It is based on a new approach to perform gas analysis: it has relatively low resolution, and employs the multicomponent algorithm [1], where the resolution of spectra is not a limitation. The use of low resolution in gas analysis provides many important advantages. For example, the recording time is short, short enough for many process controls. Fewer data of the observed and the library spectra are needed, and calculation is fast and requires only a cheap and small computer with a reasonable-sized memory, typically a microcomputer. Thus, measurement and computations can be performed in one second.

An important scientific application of FTS is *molecular spectroscopy*. In this application, infrared FT spectrometers are used to study geometric structures of molecules, and the behavior of molecules in rotations and vibrations. IR spectra reveal quantum mechanical phenomena and interactions in an effective way. In this type of research, high resolution and accuracy are very important, and FT-IR spectrometry is the most commonly used method.

Let us, as a simple example of molecular spectroscopy, consider the IR spectrum of a linear, diatomic molecule. The energy levels of the molecules are, in gas phase,

$$E_{n,J} = E_{\mathrm{vibr}} + E_{\mathrm{rot}} = (n + \frac{1}{2})hc\nu_0 + hcBJ(J+1) - hcDJ^2(J+1)^2, \qquad (6.49)$$

where E_{vibr} is the vibrational energy and E_{rot} the rotational energy of the molecule. The last term in the latter expression is the effect of the centrifugal distortion of the molecule. n and J are the vibrational and the rotational quantum number, respectively. h is *Planck's constant* ($h = 6.626\,0755 \times 10^{-34}$ Js), and c is the velocity of light. B and D are rotational constants, which can be expressed as

$$B = \frac{h}{8\pi^2 I_0 c}\,, \qquad (6.50)$$

and

$$D = \frac{4B^2}{\nu_0^2}\,, \tag{6.51}$$

where I_0 is the moment of inertia, depending on the geometrical structure and mass of the molecule. ν_0 is a molecular constant which depends on the force between atoms;

$$\nu_0 = \frac{1}{2\pi c}\,\omega_0 = \frac{1}{2\pi c}\,\sqrt{\frac{k}{\mu}}, \tag{6.52}$$

where ω_0 is the classical oscillation angular frequency, k is string constant, and μ is reduced mass. When the molecules absorb infrared radiation of wavenumber ν, the energy $hc\nu$ excites the molecule to a higher energy level. From Equation (6.49),

$$\begin{aligned}\nu = \frac{\Delta E_{n,j}}{hc} \quad &= \quad \nu_0(n' - n'') + B'J'(J'+1) - D'J'^2(J'+1)^2 \\ &\quad - B''J''(J''+1) + D''J''^2(J''+1)^2,\end{aligned} \tag{6.53}$$

where $'$ and $''$ indicate higher and lower energy levels, respectively. The *selection rules* for absorption are $\Delta n = 1$ and $\Delta J = 0, \pm 1$. For the R-branch of the spectrum $J'' = J$ and $J' = J + 1$, so that $\Delta J = +1$. For the Q-branch the selection rule is $J' = J'' = J$, and $\Delta J = 0$. This branch does not exist for diatomic molecules. For the P-branch, $J'' = J$ and $J' = J - 1$, so that $\Delta J = -1$. The molecular constants ν_0, B', B'', D', and D'' can be derived very accurately by fitting the observed wavenumbers as functions of J.

Figure 6.23 shows an experimental example of a molecular spectrum. The spectrum contains the ν_1-region $10^00 \leftarrow 00^00$ of the $^{16}O^{12}C^{32}S$ molecule. The spectrum was recorded by the Oulu FT-IR spectrometer [2, 3]. The resolution of the spectrometer was 0.006 cm^{-1}. The spectrum was apodized. A comparison of the observed line positions with theoretically calculated wavenumbers showed that the accuracy of this measurement was very high [4].

Figure 6.24 shows another example of a high-resolution FT-IR measurement. The spectrum is the bending band ν_2 of CO_2 near 667 cm^{-1}. This spectrum was also recorded in Oulu. The instrumental resolution was 0.002 cm^{-1}. Table 6.1 shows how accurately one can fit the wavenumber according to model equations. The relative errors are typically 10^{-8}. The vibration energy levels of different normal modes are sometimes very close to each other. In this case these normal modes can be coupled so that the energy levels are shifted apart. Then the observed wavenumber cannot be expressed in a closed form (like Equation 6.53), and the rotational lines of the spectrum are shifted. With proper quantum mechanical calculations, it is possible to calculate these shifts. The high resolution and accuracy in the wavenumber scale of the FT-IR spectrometer make it a very effective way to study these kinds of interactions in molecular spectra.

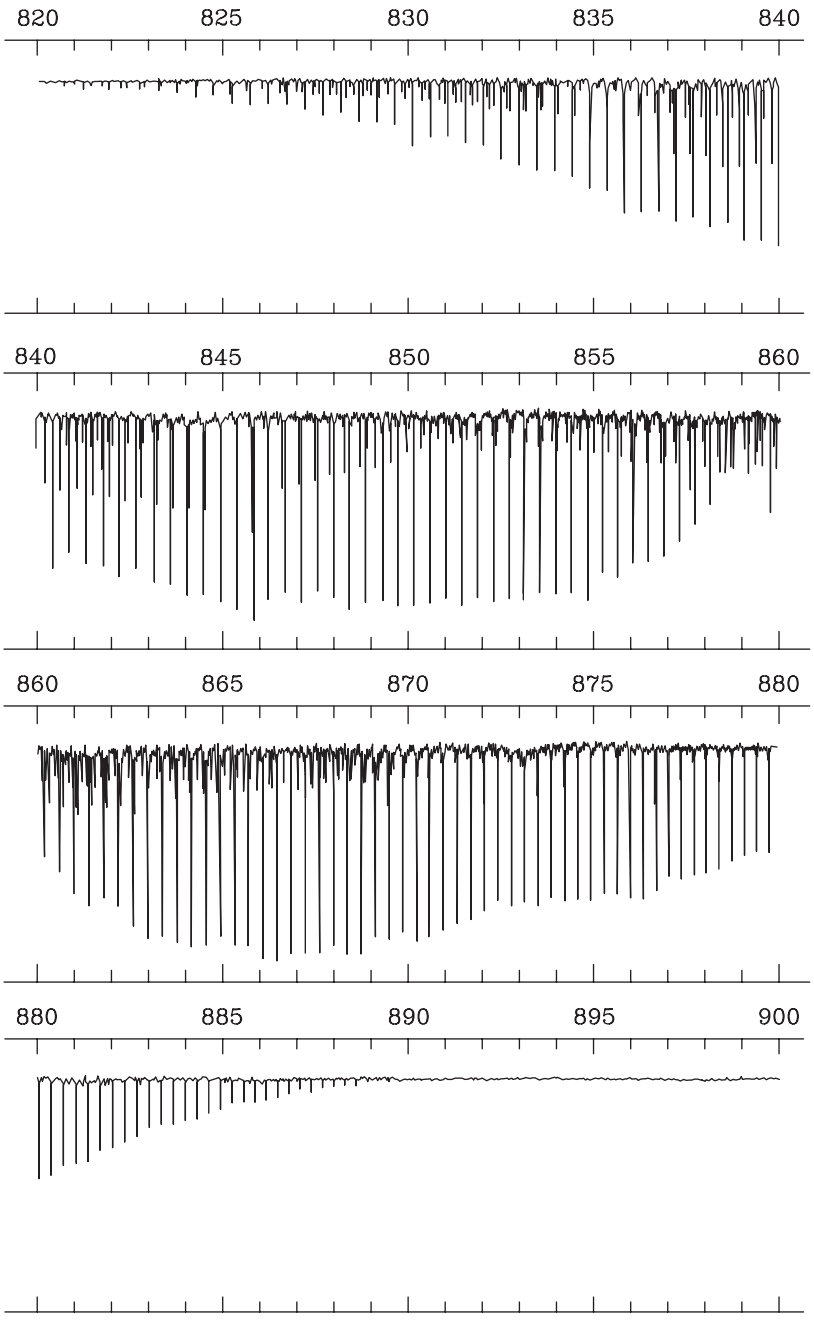


Figure 6.23: FTS measurement of the ν_1-region of the $^{16}O^{12}C^{32}S$ molecule. The wavenumber is expressed in the unit cm^{-1}. The spectrum is apodized. The resolution is 0.006 cm^{-1}, the absorption path length 1 m, and the gas pressure 40 Pa. This measurement was made at the University of Oulu [4].

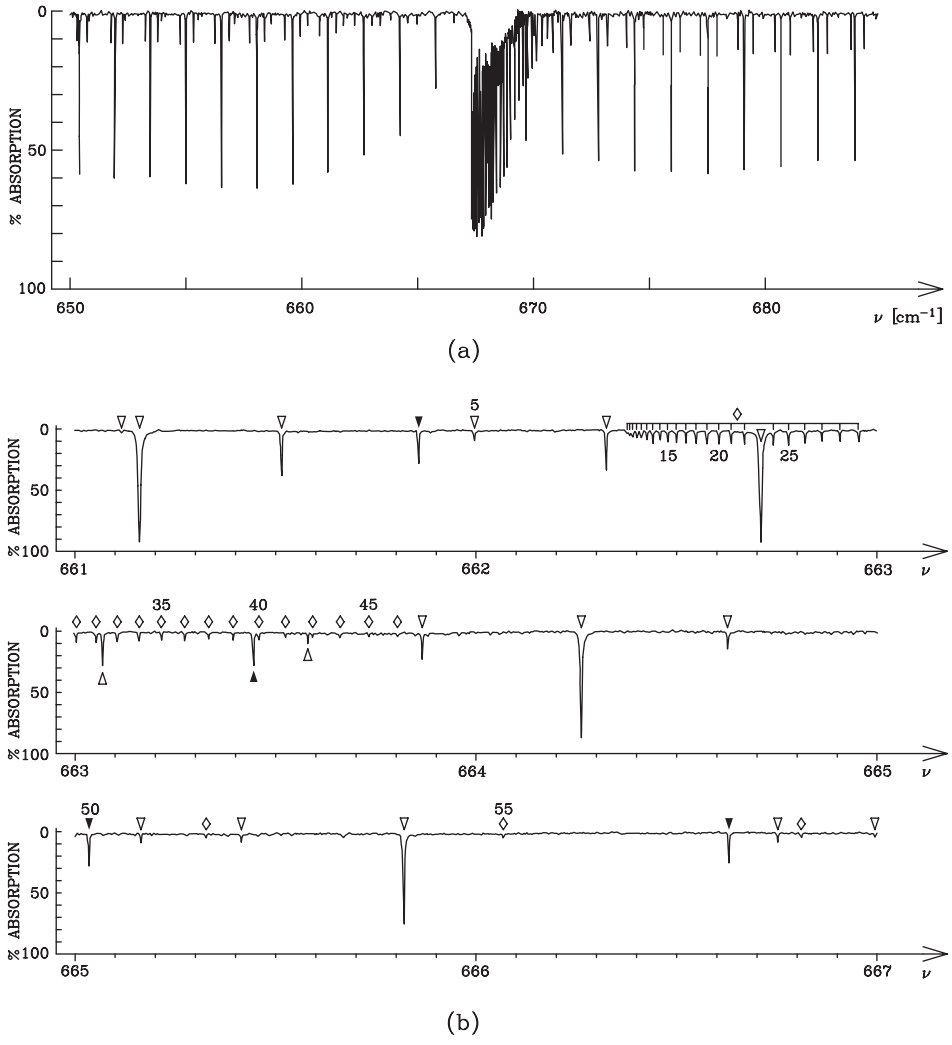


Figure 6.24: Part of the ν_2 bending band of CO_2 near 667 cm^{-1} measured by Oulu FT-IR spectrometer with resolution of 0.0045 cm^{-1} (a) and with resolution of 0.002 cm^{-1} (b). In (b), ▽ indicates $^{12}C^{16}O_2$, ▼ indicates $^{13}C^{16}O_2$ and ◊ indicates $^{16}O^{12}C^{18}O$. [3]

Table 6.1: Calculated wavenumbers, and difference between observed and calculated wavenumbers, of the fundamental of the ν_2 band $01^10 \leftarrow 00^00$ of CO_2, shown in Fig. 6.24. Standard deviation is about 1.3×10^{-5} cm^{-1}. The notation A means overlapping lines that are not in the fit.

J	$P(J)$ calc.	obs. − calc.	$R(J)$ calc.	obs. − calc.
0			668.161103	0.000014
2	665.819796	0.000010	669.726165	– 0.000020
4	664.263152	– 0.000005	671.294548	– 0.000018
6	662.709919	– 0.000022	672.866225	– 0.000008
8	661.160121	0.000006	674.441168	– 0.000021
10	659.613781	0.000001	676.019349	0.000018
12	658.070921	0.000013	677.600737	0.000027
14	656.531563	– 0.000003	679.185303	0.000021
16	654.995728	0.000009	680.773015	0.000030
18	653.463436	– 0.000009	682.363842	0.000020
20	651.934706	– 0.000019	683.957751	– 0.000004
22	650.409558	0.000002	685.554710	– 0.000008
24	648.888009	– 0.000035 A	687.154683	– 0.000006
26	647.370076	– 0.000004	688.757637	0.000003
28	645.855776	– 0.000021	690.363535	0.000004
30	644.345124	– 0.000017 A	691.972343	– 0.000008
32	642.838135	– 0.000003	693.584021	– 0.000019
34	641.334824	– 0.000016	695.198534	0.000000
36	639.835204	0.000002	696.815842	0.000004
38	638.339287	0.000003	698.435906	– 0.000014
40	636.847086	0.000009	700.058686	– 0.000008
42	635.358612	0.000008	701.684142	0.000007
44	633.873874	0.000020	703.312232	0.000001
46	632.392884	0.000005	704.942913	– 0.000008
48	630.915650	0.000019	706.576144	0.000001
50	629.442179	0.000015	708.211879	0.000006
52	627.972480	– 0.000009	709.850075	0.000003
54	626.506560	– 0.000019	711.490687	– 0.000026
56	625.044423	0.000012	713.133668	0.000011
58	623.586076	– 0.000015	714.778972	0.000015
60	622.131522	– 0.000005	716.426551	0.000081 A

In *dispersive FT-IR spectrometry* the sample is placed in one arm of a Michelson-type interferometer. The optical path length in that arm is increased by the amount $2d[n(\nu) - 1]$, where $n(\nu)$ is the refractive index of the sample of thickness d at wavenumber ν. This causes the centerburst of the interferogram to be shifted to the point $x = 2d[\overline{n(\nu)} - 1]$, where $\overline{n(\nu)}$ is the average refractive index. Thus the interferogram $I'(x)$ is asymmetric, and its inverse Fourier transform is a complex function of ν:

$$\mathcal{F}^{-1}\{I'(x)\} = \mathcal{F}_{\cos}\{I'(x)\} - i\mathcal{F}_{\sin}\{I'(x)\}, \tag{6.54}$$

where $\mathcal{F}_{\cos}$ is the cosine transform and $\mathcal{F}_{\sin}$ the sine transform (see Equation 2.3). The *dispersion spectrum* $n(\nu)$ of the sample can be computed from the phase spectrum

$$\phi(\nu) = -\arctan\frac{\mathcal{F}_{\sin}\{I'(x)\}}{\mathcal{F}_{\cos}\{I'(x)\}} = 2\pi\,\nu\,2d[n(\nu) - 1]. \tag{6.55}$$

Neglecting losses in reflections, the absorption spectrum of the sample can be derived from the amplitude spectrum

$$A(\nu) = \sqrt{[\mathcal{F}_{\cos}\{I'(x)\}]^2 + [\mathcal{F}_{\sin}\{I'(x)\}]^2}. \tag{6.56}$$

The FT technique is the only method which can give the broadband absorption and dispersion spectra of a sample simultaneously.

FT-IR spectrometers are very rapid-scanning instruments, but in some cases the lifetime of a species can be shorter than the scanning time of the interferometer. This problem can be solved by *time-resolved spectrometry*. There are two basic methods to record time-resolved spectra. Actually, both techniques are equivalent, but one was developed for step-scanning interferometers and the other for conventional rapid-scanning interferometers. Let us examine, for example, the step-scanning interferometers. In step-scanning interferometry, the moving mirror is stopped for a while at each sampling point x_j. The moving mirror is moved from one sample point to the next one as quickly as possible. In the time-resolved method the time-dependent phenomena are repeated at each sampling point x_j, and the interferogram signal values $I(x_j, t_0)$, $I(x_j, t_1)$, ..., $I(x_j, t_n)$ are collected at discrete time $t_0, t_1, \ldots, t_n$ with a short time interval (about 1 μs). Thus, the inverse Fourier transform of the sequence $I(x_0, t_j)$, $I(x_1, t_j)$, ..., $I(x_m, t_j)$ is the spectrum at the time t_j. The temporal resolution Δt of this technique is limited by the response of the detector and hence can be of the order of microseconds.

Photoacoustic spectroscopy is based on the photoacoustic phenomenon. If radiation goes through an absorbing material, the absorbed energy generates heat. If modulated radiation goes through the gas cell, then the absorption generates heat waves and further pressure pulses or sound waves, which can be detected, for example, by a microphone. If the sample is solid, it can be placed into a cell filled with a nonabsorbing noble gas like xenon, and the absorbed energy produces sound waves in the noble gas. A photoacoustic cell can be resonant or nonresonant. The resonant operation can be achieved by choosing the dimensions of the cell in such a way that a standing sound wave is produced inside the cell. There is also an alternative method of measuring temperature or pressure fluctuations of the gas around a solid sample, introduced by the absorption of modulated radiation: the thermal gradients and consequent refractive index gradients in the contact noble gas cause a probe laser beam to be deflected. This method is called *beam deflection technique*.

Earlier, the photoacoustic method was applied mainly in laser spectroscopy, but nowadays photoacoustic detection is used most successfully also in FT-IR spectrometers. An interference record $F(x)$ may have a strong constant base level $\frac{1}{2}F(0)$ due to a wide background spectrum. The area of the whole spectrum is $\frac{1}{2}F(0)$ (see Equation 6.13). However, the absorption signal due to spectral lines under examination is very small compared with $\frac{1}{2}F(0)$, typically between $10^{-3} \times F(0)$ and $10^{-6} \times F(0)$. This is the well-known dynamic range problem of the FTS method: the interesting small signals are superimposed on a very large constant level. The dynamic range problem can be cancelled by photoacoustic detection. In the photoacoustic method only absorption lines produce a signal, background does not. Consequently, a region where there are no absorption lines does not generate any signals.

The photoacoustic system is a cell and a selective detector at the same time. The photoacoustic FT-IR spectrometer is an especially effective method to study coatings on a non-

transparent material like metal. Of course the method has also some weak points. One of them is that the photoacoustic detection system is very sensitive to all kinds of external sounds and vibrations.

In addition to those mentioned above there are very many other applications of FT spectrometry. Useful practical applications of FT-IR spectrometers include quality control, surface control, semiconductor studies, biochemical or biomedical applications like phase transition of lipids, and infrared microscopy. Very often an FT-IR spectrometer is connected to gas or liquid chromatography. In scientific research, FTS may be applied also in polarization studies, matrix isolation, low-temperature experiments, surface analysis, attenuated total reflectance (ATR) studies, emission studies, biochemical and biomedical applications, and so forth.

Problems

1. A spectrometer which is based on a Michelson interferometer measures the optical path difference with a laser which emits radiation in three modes: $\nu_0 - \Delta\nu$, ν_0 and $\nu_0 + \Delta\nu$. The corresponding intensities of the modes are $\frac{1}{2}I$, I and $\frac{1}{2}I$, respectively. Compute the intensity which is obtained in the detector as a function of the optical path difference. At which value of the optical path difference does the interferogram vanish (*i.e.*, the intensity arriving in the detector does not oscillate) for the first time? You can assume that the line width of the modes is zero and that the spectrum is symmetric.

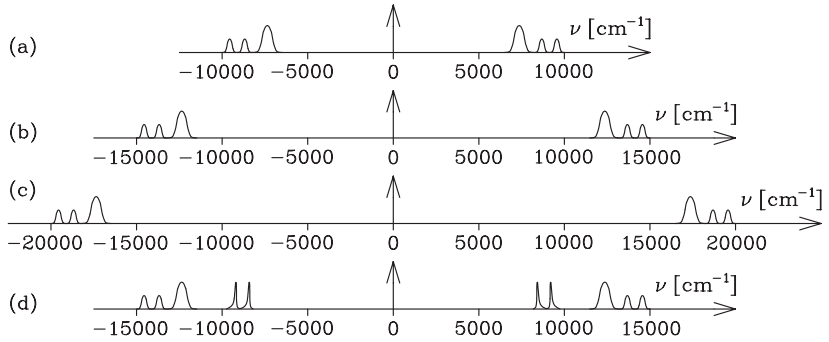
2. A Fourier transform spectrometer makes a constant error in the measurement of the optical path difference. Because of this, the peak of the interferogram seems to lie at the point $x = \varepsilon$, instead of $x = 0$. Show that if a two-sided signal is measured, then this error does not affect the power spectrum. You can neglect the change of the truncation points.

3. A Fourier transform spectrometer makes a periodic error in the measurement of the optical path difference: instead of the measured path difference x the true path difference is $x - \varepsilon \sin(2\pi x/K)$. Applying the serial expansion $\exp(iz\sin\theta) = \sum_{n=-\infty}^{\infty} J_n(z)\exp(in\theta)$, examine how the error distorts the spectrum of a monochromatic source. (The true interferogram is $I_0 \cos(2\pi\nu_0 x)$.)

4. When an interferogram, measured over the interval from $-L$ to L by a Fourier transform spectrometer, is transformed, a fully monochromatic signal (whose spectrum is Dirac's delta peak) is described as a sinc-shaped spectral line constant $\times$ $\mathrm{sinc}(2\pi L\nu)$. Since this curve determines the size of the smallest detail which can be observed in the spectrum, it is natural to call its FWHM $1.21/(2L)$ the theoretical resolution of the spectrum. Calculate the smallest number of data which enables the registration of a spectrum in the wavenumber region 0–4000 cm^{-1} at the resolution ≤ 0.5 cm^{-1}. Take into account the requirement of FFT that $2N$ must be an integer power of two.

5. Infrared spectra are measured by a Fourier transform spectrometer. The sampling interval Δx is such that $\nu_{\max} = 1/(2\Delta x) = 5000\,\mathrm{cm}^{-1}$. Outline how the following (true) spectra

look like after the computation of the inverse Fourier transform of the interferogram. (Since a single wavenumber is described in the interferometer as a cosine wave, the true spectra must be imagined to be symmetric; otherwise the spectrum and the interferogram would not constitute a Fourier transform pair.)



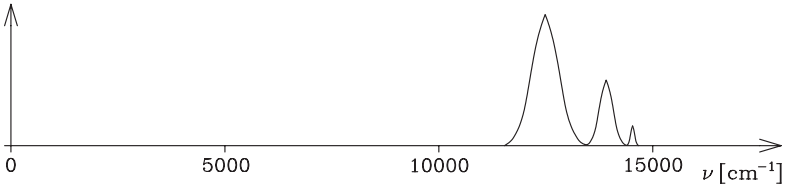
6. A Fourier transform spectrometer has an optical filter, which is transparent only in the wavenumber region 10 000 cm^{-1}–15 000 cm^{-1}. This region of a spectrum is shown in the following picture. Outline, how this spectrum will look like in a measurement where

   (a) $\nu_{\max} = 1/(2\Delta x) = 15\,000\ \mathrm{cm}^{-1}$,

   (b) $\nu_{\max} = 10\,000\ \mathrm{cm}^{-1}$,

   (c) $\nu_{\max} = 5000\ \mathrm{cm}^{-1}$.

   In each case, plot the whole wavenumber region $(-\nu_{\max}, \nu_{\max})$.



7. A spectrum under examination had spectral lines of the Loretzian shape. The FWHM of all the lines was 2σ. At which value of the optical path difference x has the amplitude a of the part of the interferogram corresponding to one line dropped down to one tenth of the constant component $\frac{1}{2}F(0)$ of the intensity in the detector?

8. If the angle of a beam propagating inside a Michelson interferometer is α, then a wavenumber ν_0 is observed as the wavenumber $\nu_0 \cos\alpha$. Let the source be circular. Thus, a monochromatic spectral line at wavenumber ν_0 spreads in the interval $[\nu_0 \cos\alpha_{\max}, \nu_0]$, where $\alpha_{\max}$ is the direction error of that beam which is emitted from the edge of the source. Compute the line shape which is observed in the case where the brightness of the source decreases linearly from a top value in the center of the source to zero in the edge.

9. The radiation source of a Michelson interferometer is a circular aperture of radius r. The focal length of the collimator lens is f. How far is it necessary to register a monochromatic signal of wavenumber ν_0, if we wish that the FWHM of the sinc-shaped distortion due to truncation would be as large as the boxcar distortion due to the finite size of the radiation source? You may assume that $f \gg r$.

10. According to the similarity theorem, the stretch a function by a constant leads to the contraction of its transform by the same constant. Consequently, the product of the FWHM of a spectral line shape and the FWHM of the corresponding interferogram amplitude is invariant. Find the value of this product for the following spectral line shapes:

    (a) Lorentzian line $\dfrac{\sigma/\pi}{\nu^2 + \sigma^2}$,

    (b) Gaussian line $e^{-\alpha\nu^2}$ $(\alpha > 0)$,

    (c) sinc line $\mathrm{sinc}(2\pi L\nu)$.

11. (a) Calculate a numerical estimate (as exact as possible) for the FWHM of the function $\mathrm{sinc}^2(\pi T f)$.

    (b) A signal which consists of cosine waves is known in the interval $-T \le t \le T$. The signal is apodized with the triangular function $(1 - \dfrac{|t|}{T})$. What is the minimum value of T, if we wish that the cosine waves $\cos(2\pi f_1 t)$ and $\cos(2\pi f_2 t)$ would be resolved in the spectrum, if $f_1 = 100.0$ MHz and $f_2 = 100.1$ MHz? Spectral lines are said to be resolved, if their distance is at least the FWHM of one line.

12. A signal consists of two cosine waves at the frequencies $f_1 = 199$ MHz and $f_2 = 200$ MHz. Samples of the signal are taken in the time intervals Δt over the interval from $-T$ to T. The samples are apodized with the triangular function

$$\Lambda_T(t) = \begin{cases} 1 - \dfrac{|t|}{T}, & |t| \le T, \\ 0, & |t| > T, \end{cases}$$

    before discrete Fourier transform.

    (a) What is the smallest value of T at which we are able to resolve the two frequencies in the spectrum? (Then the FWHM of one spectral line is $f_2 - f_1$.)

    (b) What is the largest possible value of Δt if we wish that the aliasing point $f_{\max}$ would be at least in the distance of ten FWHM from the larger frequency? Use the value of T which you obtained above.

    Hint: The FWHM of the function $\mathrm{sinc}^2(\pi T f)$ is approximately $1.7718/(2T)$.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

7 Nuclear magnetic resonance (NMR) spectroscopy

7.1 Nuclear magnetic moment in a magnetic field

Most atomic nuclei have a *magnetic moment* $\boldsymbol{\mu}$, which is proportional to the angular momentum $\boldsymbol{J}$ of the nucleus,

$$\boldsymbol{\mu} = \gamma \boldsymbol{J}. \tag{7.1}$$

The proportionality constant γ is called the *gyromagnetic ratio*, and it is a characteristic of the nucleus. According to quantum mechanics, the angular momentum is quantized, and can only possess values which satisfy the equation

$$|\boldsymbol{J}|^2 = J(J+1)\hbar^2, \qquad J = 0, 1, 2, 3, \ldots, \tag{7.2}$$

where J is the nuclear spin quantum number, and $\hbar = \frac{h}{2\pi} = 1.054\,572\,66 \times 10^{-34}$ Js. The z-component of $\boldsymbol{J}$ can have $2J + 1$ distinct values

$$J_z = m_J \hbar, \qquad m_J = J, J-1, \ldots, 0, -1, \ldots, -J. \tag{7.3}$$

Let us consider a nucleus in an external static magnetic field $\boldsymbol{B}$, which is assumed to lie in the z-direction. Applying Equations 7.1 and 7.3, the energy E of interaction of the nucleus and the magnetic field can be written as

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B} = -\gamma m_J \hbar B, \tag{7.4}$$

where B is the magnitude of $\boldsymbol{B}$. Since the z-component J_z of the angular momentum of the nucleus is quantized, also the interaction energies of nuclei in a static magnetic field are quantized, and they may only have $2J + 1$ distinct values. At thermal equilibrium, the relative probabilities of each energy state are given by the Boltzmann distribution.

The external magnetic field produces a torque $\boldsymbol{N}$, which tries to change the angular momentum and the magnetic moment of the nucleus:

$$\boldsymbol{N} = \frac{\mathrm{d}\boldsymbol{J}}{\mathrm{d}t} = \boldsymbol{\mu} \times \boldsymbol{B}, \text{ or } \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t} = \boldsymbol{\mu} \times (\gamma \boldsymbol{B}). \tag{7.5}$$

The latter equation is the equation of motion of precession: the magnetic moment vector precesses about the magnetic field vector. The angular frequency of this motion is

$$\boldsymbol{\omega}_L = -\gamma \boldsymbol{B}. \tag{7.6}$$

The size of an infinitesimal change $|\mathrm{d}\boldsymbol{\mu}|$ of the magnetic moment in time $\mathrm{d}t$, illustrated in Figure 7.1, can be expressed as

$$|\mathrm{d}\boldsymbol{\mu}| = |\boldsymbol{\mu}| \sin\theta \, |\mathrm{d}\phi| = \gamma |\mathrm{d}\boldsymbol{J}| = \gamma |\boldsymbol{N}| \, \mathrm{d}t = |\boldsymbol{\mu}| \, \gamma \, |\boldsymbol{B}| \sin\theta \, \mathrm{d}t, \tag{7.7}$$

where θ is the angle between $\boldsymbol{\mu}$ and $\boldsymbol{B}$, and $|\mathrm{d}\phi|$ is the change of the angle of $\boldsymbol{\mu}$ in the xy-plane. Consequently, we can see that the angular frequency of the precession is, indeed,

$$\boxed{\left|\frac{\mathrm{d}\phi}{\mathrm{d}t}\right| = \omega_L = \gamma B,} \tag{7.8}$$

where $\omega_L = |\boldsymbol{\omega}_L|$. The precession of the nuclear magnetic moment about the external magnetic field direction is called the *Larmor precession*, and its angular frequency $\boldsymbol{\omega}_L$ is the *Larmor angular frequency*.

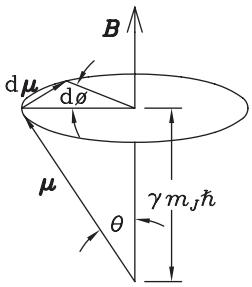


Figure 7.1: Larmor precession of a nucleus in an external static magnetic field $\boldsymbol{B}$. The magnetic moment of the nucleus is $\boldsymbol{\mu}$. The precession angle $\mathrm{d}\phi$ corresponds to the change $\mathrm{d}\boldsymbol{\mu}$ of the magnetic moment.

The total macroscopic magnetization $\boldsymbol{M}$ of all nuclei is the sum of all magnetic moments,

$$\boldsymbol{M} = \sum_i \boldsymbol{\mu}_i. \tag{7.9}$$

In a static magnetic field, the precessing magnetic moment vectors have arbitrary phases, and the angles ϕ_i which they make in the xy-plane are randomly distributed. Consequently, the x- and y-components of the magnetic moments cancel each other, and the macroscopic magnetization lies along the z-axis. Figure 7.2 shows an example of the summation of the magnetic moments of nuclei. In this example, the quantum number $J = \frac{1}{2}$, and there are only two possible energy states.

It is often useful to represent the Larmor precession in a coordinate system $x'y'z$ which rotates about the z-axis at an angular frequency $\boldsymbol{\omega}$. The rotating coordinates are illustrated in Figure 7.3. According to classical mechanics, for any vector $\boldsymbol{\mu}$ and angular frequency $\boldsymbol{\omega}$,

$$\left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t}\right)_{\mathrm{rot}} = \left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t}\right)_{\mathrm{fixed}} - \boldsymbol{\omega} \times \boldsymbol{\mu}, \tag{7.10}$$

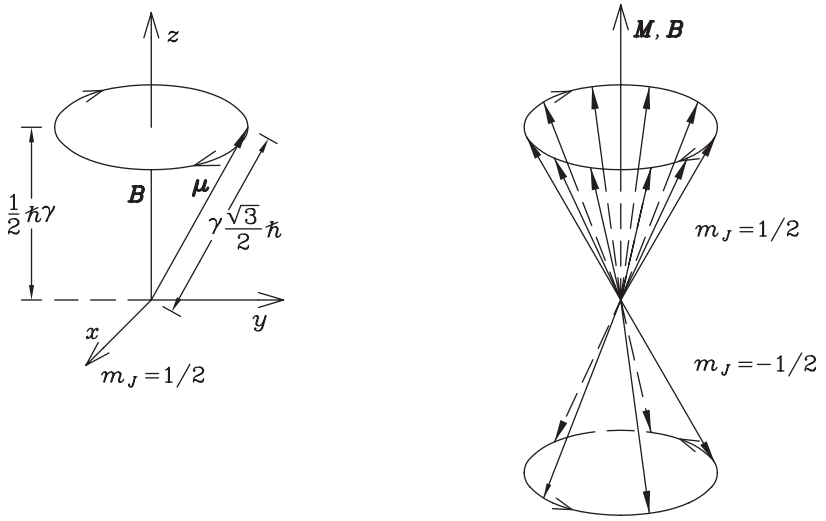


Figure 7.2: An example of the summation of the magnetic moments μ of nuclei. The spin quantum number of the nuclei is $J = \frac{1}{2}$, and there are only two possible energy states. The total magnetic moment $\boldsymbol{M}$ lies along the direction of the static magnetic field $\boldsymbol{B}$, in the z-direction.

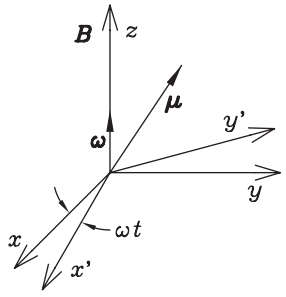


Figure 7.3: Magnetic moment μ in the coordinate frame $x'y'z$ which rotates about the z-axis of the fixed coordinate frame xyz at the angular frequency ω. $\boldsymbol{B}$ is a static magnetic field.

where the suffix $_{\text{fixed}}$ refers to the fixed frame and $_{\text{rot}}$ to the frame rotating at the angular frequency ω. Inserting this equation into Equation 7.5, we obtain the equation of motion of a magnetic moment in a static magnetic field in the rotating coordinates:

$$(\frac{\mathrm{d}\mu}{\mathrm{d}t})_{\text{rot}} = \gamma\mu \times \boldsymbol{B}_{\text{eff}}, \tag{7.11}$$

where $\boldsymbol{B}_{\text{eff}}$ is the effective magnetic field in the rotating coordinate system,

$$\boldsymbol{B}_{\text{eff}} = \boldsymbol{B} + \omega/\gamma. \tag{7.12}$$

We can see that in the rotating frame the magnetic moment precesses about the effective magnetic moment $\boldsymbol{B}_{\text{eff}}$ at the angular frequency $-\gamma \boldsymbol{B}_{\text{eff}}$.

7.2 Principles of NMR spectroscopy

The goal of nuclear magnetic resonance, NMR, is to study molecular structures, molecular motion, and various chemical characteristics. This kind of information can be found in the nuclear magnetic resonance spectrum. The NMR spectrum reveals the Larmor frequencies of the nuclei in a sample.

As explained in the previous chapter, the sum of the precessing magnetic moments of nuclei in a static magnetic field, that is, the macroscopic magnetic moment $\boldsymbol{M}$, lies along the z-axis (the direction of the magnetic field). In order to be able to measure $\boldsymbol{M}$ to obtain an NMR spectrum, it is necessary to rotate $\boldsymbol{M}$ in such a way that an observable component is achieved along, say, y-axis. In NMR, this is obtained by applying a weak oscillating magnetic field perpendicular to the static magnetic field, say, in x-direction, for a certain period of time.

Let us add an oscillating magnetic field

$$\boldsymbol{B}_x = 2B_0 \cos(\omega t)\boldsymbol{u}_x, \tag{7.13}$$

which lies along the x-direction. $\boldsymbol{u}_x$ is the unit vector in the positive x-direction. The oscillating magnetic field vector can be divided into two components, one rotating clockwise and one counterclockwise about the z-axis at the angular frequency ω. The component which rotates in the same direction as the magnetic moments is significant in NMR, but the component which rotates in the opposite direction can be neglected. In a rotating coordinate frame which rotates at the same angular frequency ω about the z-axis the significant component appears static, and can be written as

$$\boldsymbol{B}_0 = B_0 \boldsymbol{u}_{x'}. \tag{7.14}$$

The overall effective magnetic field experienced by a nuclear magnetic moment is, in the rotating frame,

$$\boldsymbol{B}_{\text{eff}} = \boldsymbol{B} + \omega/\gamma + \boldsymbol{B}_0. \tag{7.15}$$

In the rotating frame, the nuclear magnetic moment vector precesses about this effective magnetic field vector. This situation is illustrated in Figure 7.4.

The same situation in the laboratory frame is illustrated in Figure 7.5. If the angular frequency of the oscillating magnetic field (and the rotating frame) ω is far from the Larmor angular frequency ω_L of the nuclear magnetic moment, then the magnetic moment continues to precess around the z-axis almost as before, only its orbit is slightly disturbed by the additional oscillating field (nutation). However, if the angular frequency of the oscillating magnetic field ω is the same as the Larmor angular frequency ω_L, that is, they are in *resonance*,

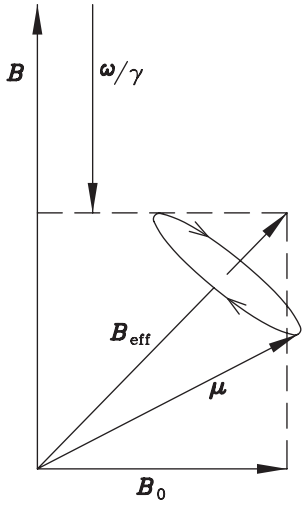


Figure 7.4: Precession of a nuclear magnetic moment vector μ about the effective magnetic field $\boldsymbol{B}_{\text{eff}} = \boldsymbol{B} + \boldsymbol{\omega}/\gamma + \boldsymbol{B}_0$ in a rotating coordinate frame. The frame as a whole rotates about the z-axis (direction of $\boldsymbol{B}$) at the angular frequency ω of the oscillating field $\boldsymbol{B}_0$. Thus $\boldsymbol{B}_0$ rotates together with the coordinate frame.

then the situation is very different. In this case, the nuclei are able to receive energy from the oscillating magnetic field, and they may be transferred to higher energy levels. As the energy level grows, the angle θ between the magnetic moment of an individual nucleus and the positive z-direction becomes larger: the magnetic moment "falls down".

Energy can be transferred only if the nucleus precesses in the same phase with the oscillating field, and this makes the precession of the magnetic moment vectors coherent. At the macroscopic level, this means that the x- and y-components of the moments no longer cancel each other. The macroscopic magnetization $\boldsymbol{M}$ is no longer directed along the z-axis, but acquires x- and y-components, and starts also to "fall down". In resonance,

$$\omega = \omega_L = -\gamma \boldsymbol{B}, \tag{7.16}$$

and the effective magnetic field in the rotating frame is

$$\boldsymbol{B}_{\text{eff}} = \boldsymbol{B}_0. \tag{7.17}$$

Consequently, in the rotating frame, the macroscopic magnetization precesses about the x'-axis at the angular frequency

$$\omega_0 = \gamma B_0. \tag{7.18}$$

This angular frequency indicates how rapidly the macroscopic magnetization falls down from the direction of the z-axis.

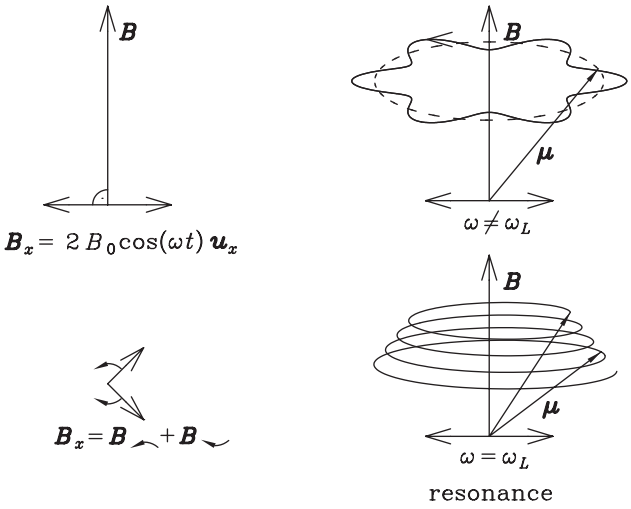


Figure 7.5: The movement of a nuclear magnetic moment $\boldsymbol{\mu}$ in a magnetic field which consists of a static magnetic field $\boldsymbol{B}$ and an oscillating magnetic field $\boldsymbol{B}_x = 2B_0 \cos(\omega t)\boldsymbol{u}_x$. The oscillating magnetic field can be divided into two components, one rotating clockwise and one counterclockwise. Only one of the components is significant in NMR, the one which rotates in the same direction where the magnetic moments are precessing. If the angular frequency ω of the oscillating magnetic field is far from the Larmor angular frequency ω_L of the nuclear magnetic moment, then the magnetic moment precession around $\boldsymbol{B}$ is only slightly distorted. If ω and ω_L are in resonance, then $\boldsymbol{\mu}$ will start to "fall down".

In NMR, the macroscopic magnetization is usually made to fall down by applying an oscillating single-frequency magnetic resonance field for a short duration Δt. This is called the *excitation* of NMR. The field strength and the duration of the pulse are chosen in such a way that the macroscopic magnetization will fall down a desired angle θ_M. θ_M may be chosen to be, for example, 90°, which is called the 90° pulse. After a 90° pulse the magnetization lies in the xy-plane. A common choice is the 30° pulse. If θ_M is chosen to be 180°, then the direction of the macroscopic magnetization after excitation is the negative z-direction.

After excitation, the macroscopic magnetization $\boldsymbol{M}$ rotates about the z-axis (in the laboratory frame). It is now easy to measure the y-component of $\boldsymbol{M}$, because it varies as

$$M_y = M_0 \cos(\omega_L t), \tag{7.19}$$

and induces a sinusoidal electric current in a coil of wire which is wrapped around the y-axis and used as the receiver.

After the excitation pulse is switched off, the nuclear magnetic moments tend to relax back towards the thermal equilibrium distribution. This occurs by several mechanisms. Energy may be transferred from the nucleus to surrounding molecules as thermal motion. This is known as spin-lattice or longitudinal relaxation. Energy may also be transferred to nuclei of a different type. This is called spin-spin or transverse relaxation. As a result of the relaxation of the nuclear magnetic moments, also the macroscopic magnetization $\boldsymbol{M}$ starts to relax back

towards the direction of the z-axis. The relaxation is exponential. The y-component of $\boldsymbol{M}$ is

$$M_y = M_0 e^{-t/\tau} \cos(\omega_L t), \tag{7.20}$$

where τ is the relaxation time. In NMR, this is the signal measured by the receiver. It is called the *free induction decay*, FID, signal. The FID-signal is a time-domain signal, which is sampled, stored, and Fourier transformed. At this stage, the properties of Fourier transforms should be considered. One should, for example, remember the Nyquist theorem in deciding the sampling interval and determining the maximal detectable frequency. The methods of processing the signal and the spectrum, presented in Chapters 11–13, can be applied. The final result of an NMR measurement is the NMR spectrum in the frequency-domain.

Figure 7.6 illustrates the NMR process.

7.3 Applications of NMR spectroscopy

The static magnetic field used in NMR measurements is typically a few Tesla in magnitude. Strong magnetic fields can be obtained by superconducting magnets. The *Larmor frequencies* $f_L = \omega_L/2\pi$ of nuclei are, then, of the order of tens or hundreds of MHz. The duration of the radio-frequency magnetic field pulse is typically a few μs, or even less, and the amplitude of the pulse is a few Gauss. The relaxation time after the pulse is switched off may be some seconds.

Because the excitation pulse, which makes the macroscopic magnetization fall down, is a truncated wave in t-domain, it covers a certain range of frequencies in the f-domain. If the duration of the pulse is Δt, it contains frequencies approximately from $\omega - \frac{1}{\Delta t}$ to $\omega + \frac{1}{\Delta t}$. A pulse of the order of 1 μs may give information over a frequency range of a few kHz.

Each set of distinct nuclei whose Larmor frequency falls in the frequency range of an NMR measurement will produce an exponentially decaying sine wave in the FID signal. The Larmor frequencies of different types of nuclei are easy to distinguish in a nuclear magnetic resonance spectrum. The width of one spectral peak may be as small as 1 Hz.

The Larmor frequency depends on the magnitude of the static magnetic field. The resolution of NMR spectroscopy can be very high, and effects of very small variations of the magnetic field strength on the Larmor frequency can be distinguished. The environment of each nucleus produces such variations. The external magnetic field may cause circulation of electrons in the surroundings of the nucleus, and this may reduce the true magnetic field. The electron circulation depends on the chemical bonds near the atom. Also the magnetic moments of other nuclei in the vicinity may affect the magnetic field. In practice, the true magnetic field is

$$\boldsymbol{B}_{\text{true}} = (1 - \sigma)\boldsymbol{B}, \tag{7.21}$$

where σ is a shielding factor. As a result, the Larmor frequencies of similar nuclei may differ if they are situated in different surroundings.

NMR spectroscopy can be employed to identify chemical bonds, to find the orientation of a bond, to measure distances of the nuclei, to determine molecular structures, to examine chemical dynamics, and so forth. In addition to high-resolution NMR spectroscopy, there is

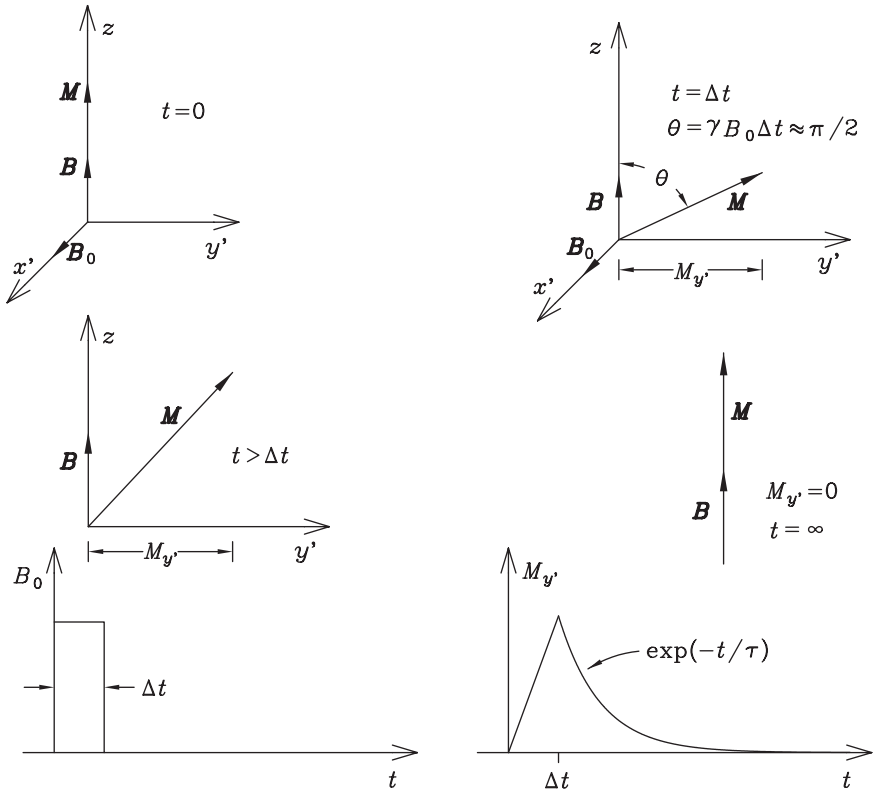


Figure 7.6: The principle of the NMR process. Initially ($t = 0$), the macroscopic magnetic moment $\boldsymbol{M}$ lies along the z-axis, which is the direction of the static magnetic field $\boldsymbol{B}$. An oscillating magnetic field, oscillating at the Larmor angular frequency ($\omega_L = \gamma B$), is switched on. In the rotating coordinate frame $x'y'z$ it is seen as the static magnetic field $\boldsymbol{B}_0$, which rotates about the z-axis together with the frame. $\boldsymbol{M}$ starts to fall down. At $t = \Delta t$ the oscillating field is switched off. At this moment, $\boldsymbol{M}$ makes an angle $\theta = \gamma B_0 \Delta t$ with respect to the z-axis, and its y'-component $M_{y'}$ reaches a maximum. After the switch-off ($t > \Delta t$), $\boldsymbol{M}$ relaxes back towards the direction of the z-axis, and $M_{y'}$ decreases exponentially. At $t = \infty$ $\boldsymbol{M}$ lies along the z-axis again.

a variety of ways to tailor NMR for different purposes. Modifications of NMR may be used to study highly relaxing nuclei which exhibit broad spectral lines, to examine solids, to make three-dimensional images, and so forth.

Figure 7.7 is an example of an experimental high-resolution NMR spectrum, measured at the University of Oulu. It is a spectrum of monodeuterobenzene C_6H_5D. The structure of the spectrum is determined by the proton chemical shifts and proton-proton spin-spin and dipole-dipole couplings. Figure 7.8 shows an example of an experimental solid state NMR measurement, made at the University of Turku. The signal is the proton solid echo. The frequency spectrum, calculated from the signal by FFT, reveals features of solid state molecular structure.

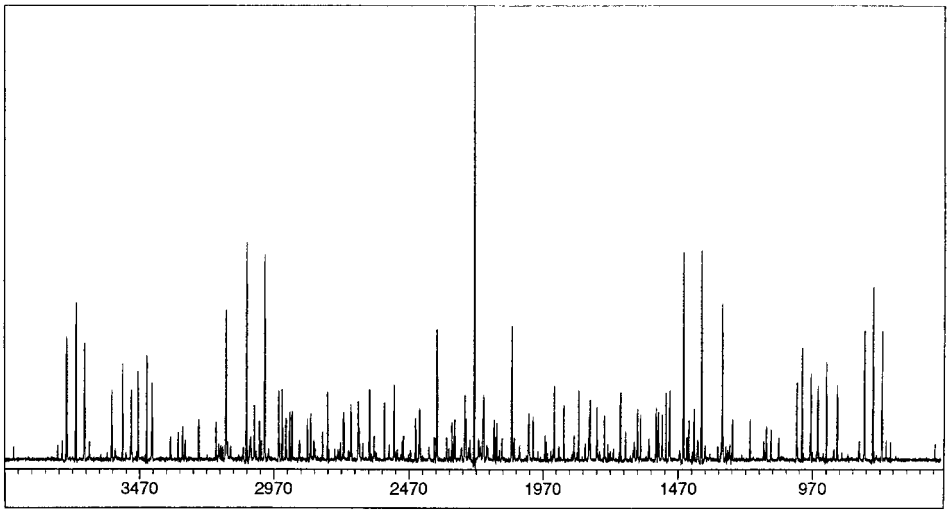


Figure 7.7: ^{1}H NMR spectrum of monodeuterobenzene C_6H_5D in liquid crystal Merck Phase 4 solvent. The spectrum is recorded in the nematic phase of the liquid crystal. The Larmor frequency was 300 MHz. The figures on the abscissa are in Hz relative to an arbitrary reference. *Courtesy of Susanna Ahola, Anu Kantola and Jani Saunavaara, NMR Research Group, Dept. of Physical Sciences, University of Oulu.*

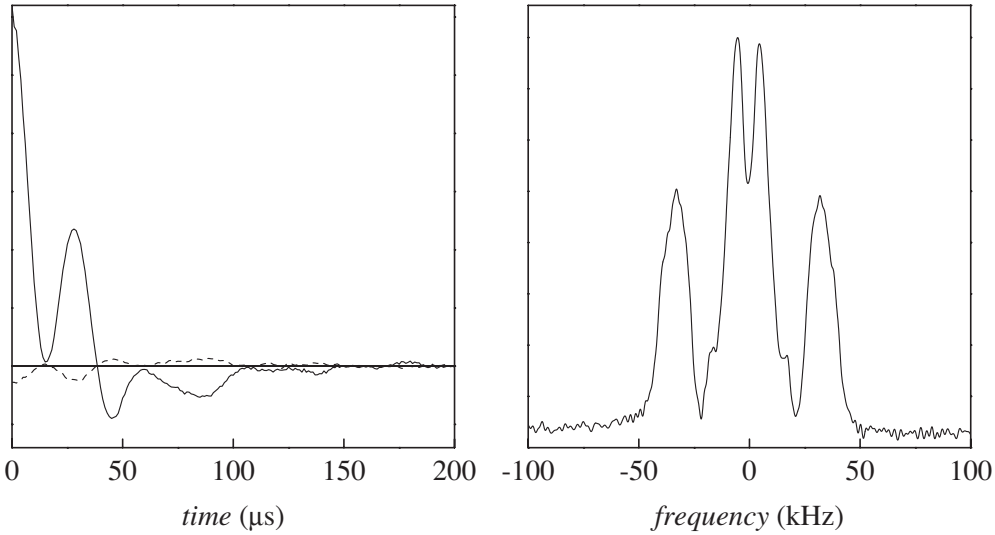


Figure 7.8: NMR measurement of $CH_3COONa{\cdot}3D_2O$ at 77 K. The purpose of the measurement was to study solid state molecular interactions and molecular dynamics. Left: The real part (the amplitude, solid line), and the imaginary part (the phase, dashed line) of the proton solid echo. Right: Frequency shifts from the proton Larmor frequency, calculated from the proton solid echo. These frequency shifts are caused by chemical interactions. *Courtesy of Wihuri Physical Laboratory, Dept. of Physics, University of Turku.*

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

8 Ion cyclotron resonance (ICR) mass spectrometry

8.1 Conventional mass spectrometry

When an ion of charge q moves at the velocity $\boldsymbol{v}$ in a region where there are an electric field $\boldsymbol{E}$ and a magnetic field $\boldsymbol{B}$, it is subjected to the Lorentz force

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{8.1}$$

In the absence of an electric field, only the magnetic component of the force is present. Because of the properties of the vector product, the magnetic force is perpendicular to the plane determined by the velocity and the magnetic field. If $\boldsymbol{v}$ would be parallel to $\boldsymbol{B}$, then the force would be zero.

Since the magnetic force is perpendicular to the velocity, the work done by the force is zero, and the kinetic energy of the ion moving in the magnetic field remains constant. The direction of the velocity changes, but the magnitude remains the same. If the magnetic field is uniform, then the ion moves in a uniform circular orbit. The equation of motion of an ion in a uniform magnetic field is

$$F = mv^2/R = qvB, \tag{8.2}$$

where m is the mass of the ion, and R is the radius of the circular orbit, and B is the magnitude of the magnetic field. Equation 8.2 yields

$$R = mv/qB. \tag{8.3}$$

A *mass spectrometer* is an instrument which separates ions of different mass-to-charge ratios m/q. In a uniform magnetic field, ions of different m/q but constant velocity $\boldsymbol{v}$ have a different radius of orbit R, and they can be separated. Consequently, a mass spectrum of the ions can be obtained.

Figure 8.1 shows a schematic illustration of a conventional mass spectrometer. Ions are produced by an ion source, and their direction is determined by the slits S_1, S_2 and S_3. The magnitude of the velocity of the ions is determined by the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$, perpendicular to each other and to the direction of the velocity. Ions are subjected to the Lorentz force (Equation 8.1), and only ions which satisfy the condition

$$qE = qvB, \tag{8.4}$$

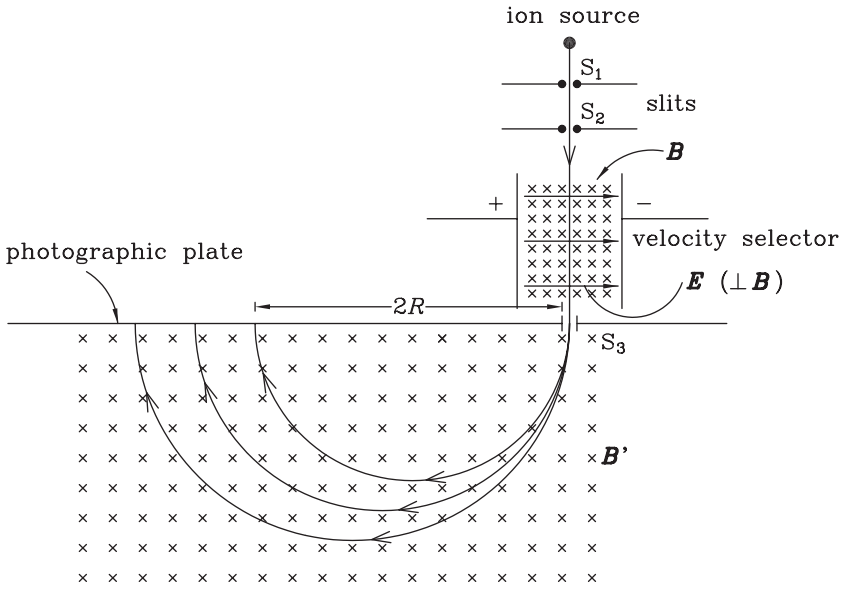


Figure 8.1: Conventional mass spectrometer. Ions are produced by an ion source, and their direction is determined by the slits S_1, S_2 and S_3. The magnitude of the velocity of the ions is determined by the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$. Another magnetic field $\boldsymbol{B}'$ forces the ions to circular orbits. The ions are registered by a photographic plate. The direction of the magnetic fields is perpendicular to the plane of the paper, inward.

that is, have the velocity

$$v = E/B, \tag{8.5}$$

continue moving straight forward and can enter the main chamber through the slit S_3. In the main chamber there is another uniform magnetic field $\boldsymbol{B}'$. In this field, the radii of the ions are different, depending on the m/q value of the ion. The radii are given by

$$R = mv/qB' = mE/qBB'. \tag{8.6}$$

Ions are registered by a photographic plate, where ions of different m/q arrive in different spots. The majority of the ions which are analyzed generally have a positive unit charge, and the mass m can be obtained directly from Equation 8.6. The result of the measurement with the spectrometer is the mass spectrum of the ions. A conventional mass spectrum is illustrated in Figure 8.2.

In practice, ions are often produced from sample molecules by bombarding them with a high-energy electron beam. The positive fragments of the molecules are then accelerated and directed to the main chamber of the mass spectrometer. Instead of a photographic plate, ions may be detected by a collector, where they generate an electric current which is amplified and

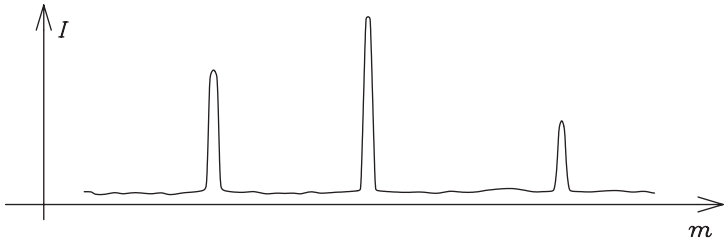


Figure 8.2: Conventional mass spectrum: intensity I of ions as a function of the mass m. The charges of the ions are assumed constant.

measured. The magnitude of the magnetic field in the main chamber may be varied, and in this way the mass-to-charge ratio of ions which hit the collector can be varied continuously.

8.2 ICR mass spectrometry

In *ion cyclotron resonance mass spectrometry*, ICR-MS, ions are stored in an ICR analyzer cell in a static magnetic field. Ions rotate in circular paths, trapped by the magnetic field and additional electric fields. The movement of ions is made coherent by applying in the trap an oscillating, resonant electric field. A signal which reveals the coherent movement of ions is obtained by detector plates.

As explained in Chapter 7, the path of an ion in a uniform magnetic field is a circular orbit. The angular frequency ω_c of the ion is

$$\omega_c = v/R = (q/m)B. \tag{8.7}$$

This angular frequency is called the *ion cyclotron frequency* . We can see that it is independent of the velocity v, and depends only on the ratio m/q and on the field B. An ion of a certain m/q value proceeds at the same cyclotron frequency, independent of the velocity of the ion. The magnitude of the static magnetic field of a mass spectrometer is typically of the order of a few Tesla. Thus, the cyclotron frequencies range typically from a few kHz to a few MHz.

Equation 8.6 gives the ion cyclotron orbital radius. If an ion is in thermal equilibrium with its surroundings at a temperature T, then its kinetic energy $\frac{1}{2}mv^2$ is, in the average, equal to kT, where k is the Boltzmann constant. Solving v and inserting it in Equation 8.6 yields the orbital radius of an ion in equilibrium in a uniform magnetic field,

$$R = \frac{1}{qB}\sqrt{2mkT}. \tag{8.8}$$

In a magnetic field of a few Tesla, the orbital radius of an ion in equilibrium at the room temperature is, generally, much less than 1 mm. At higher temperatures and higher kinetic energies the ion cyclotron orbital radii may be of the order of centimeters.

Taking the first derivative of ω in Equation 8.7 with respect to m, we obtain that

$$\frac{d\omega}{\omega} = \frac{-dm}{m}. \tag{8.9}$$

This means that the mass resolution of ICR is equal to the ion cyclotron frequency resolution. The mass resolution can be written as

$$\frac{\Delta m}{m} = \frac{-m\Delta\omega}{qB}. \tag{8.10}$$

Let us choose the coordinates in such a way that the magnetic field is directed along the z-direction: $\boldsymbol{B} = B\boldsymbol{u}_z$. The magnetic field traps an ion effectively in a circular orbit in the xy-plane, and the ion cannot escape from the trap in the x- or y-directions. The ion may, however, move freely in the z-direction. In an ICR mass spectrometer, the escape of an ion in the z-direction is prevented by applying a small electrostatic potential to electrodes in both sides of the ion trap, at $z = \pm\frac{a}{2}$, where a is the height of the trap. These trapping electrodes expel ions towards $z = 0$ in the center of the trap.

If the trap is cubic, and the trapping electrodes are connected to a voltage V_T, whereas the plates at the other four walls of the trap are grounded, then the electrostatic potential near the center of the trap can be shown to be, approximately,

$$V(x, y, z) = \frac{V_T}{3} - \frac{\alpha V_T x^2}{a^2} - \frac{\alpha V_T y^2}{a^2} + \frac{2\alpha V_T z^2}{a^2}, \tag{8.11}$$

where $\alpha = 1.386$. In the z-direction, the electric field is

$$E_z(z) = -\frac{\mathrm{d}V}{\mathrm{d}z} = -\frac{4\alpha V_T}{a^2}z = -k_z z, \tag{8.12}$$

where k_z is a constant. The electric force is

$$F_z(z) = m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -qk_z z. \tag{8.13}$$

This force makes the ion oscillate back and forth around $z = 0$ at the frequency $\omega_t = \sqrt{k_z q/m}$ (where $q > 0$). This *trapping frequency* of oscillation in z-direction is, generally, much smaller than the cyclotron frequency. The circular motion of ions in the ICR-MS cell is, hence, modulated by a slight oscillation in the z-direction.

The electrostatic potential of Equation 8.11 produces also a radial electric field

$$E(r) = \frac{2\alpha V_T}{a^2}r = E_r r \tag{8.14}$$

in the xy-plane. Consequently, the radial force in the trap consists of both the magnetic force, pulling the ion towards the center, and an electric force, pushing the ion away from the center. The equation of motion of an ion in the xy-plane inside a cubic ICR mass spectrometer is thus

$$m\omega^2 r = qB\omega r - qE_r r. \tag{8.15}$$

This equation has two solutions:

$$\omega_0 = \frac{qB + \sqrt{q^2B^2 - 4mqE_r}}{2m}, \tag{8.16}$$

and

$$\omega_m = \frac{qB - \sqrt{q^2B^2 - 4mqE_r}}{2m}. \tag{8.17}$$

ω_0 is the frequency of the orbital motion of the ion. We can see that the cyclotron frequency (Equation 8.7) is slightly reduced by the electric trapping potential. In exact calculations, the cyclotron frequency ω_c given by Equation 8.7 must be replaced by ω_0. ω_m is called the *magnetron frequency*. It is much smaller than the cyclotron frequency. The existence of ω_m means that an ion moving in the ICR cell is also making a small precession around its orbit.

In an ICR mass spectrometer, the movement of ions is detected by placing detector plates at the sides of the trap and measuring the charge which is induced in the detector plates by the moving ions. In a cubic trap, the plates may be placed at $x = \pm a/2$, where a is the width of the trap.

In a static electromagnetic ion trap, the ion cyclotron orbital motion is incoherent: the phase of the orbital motion of ions is random. No overall charge is induced in the detector plates, because the average of movements of all ions is zero. In order to obtain a signal in the detector plates, the movement of ions should be spatially coherent: ions should move in the same phase. A coherent movement of ions can be created by applying in the trap an oscillating, resonant electric field, produced by excitation plates placed at the sides of the trap. This electric field is made to oscillate at the cyclotron frequency of the ions of interest:

$$E(t) = \cos(\omega_c t). \tag{8.18}$$

The power absorption of the ions is proportional to the dot product $\boldsymbol{E}(t) \cdot \boldsymbol{v}$, and the magnitude of the velocity of resonant ions and the radius of their orbit will grow. Due to the oscillating electric field, ions are forced to move in their orbits in the same phase. Resonant ions form ion packets, which rotate at growing orbits. The oscillating excitation voltage is switched off after a suitable period of time, before the orbits of the ions grow too large for the trap.

However, the acceleration of ions whose cyclotron frequency differs clearly from the frequency of the excitation electric field experience only a limited acceleration. The radius of orbit of off-resonant ions grows only minimally. If we wish to study a larger frequency region in one measurement, it is necessary to apply a more complicated excitation pulse, as discussed in the next section.

Figure 8.3 shows the electrodes around a cubic electromagnetic ion trap, used in ICR-MS. The ion trap may also be designed in several alternative ways.

As a coherent ion packet rotates in the cell, it induces a signal in the detector plates. The signal obtained is of the form

$$I(t) \propto N(t)\cos(\omega_c t), \tag{8.19}$$

where $N(t)$ is the number of ions rotating in the same phase at the frequency ω_c. After the excitation voltage is switched off, the number of the rotating ions is reduced by collisions:

$$N(t) = N_0 e^{-t/\tau}, \tag{8.20}$$

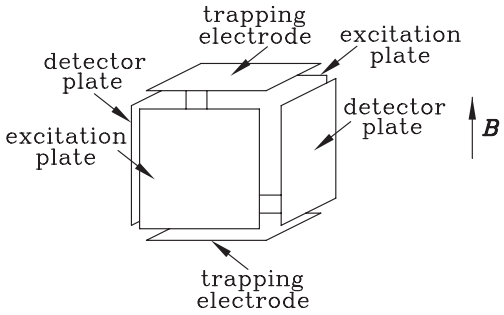


Figure 8.3: Electrodes around a cubic electromagnetic ion trap, used in ion cyclotron resonance mass spectrometry.

where τ is a damping constant. In addition to this, the signal recorded in the detector plates is reduced because ions whose natural cyclotron frequencies differ from the exact excitation frequency begin to differ in phase. Consequently, the whole signal, which is obtained in the detector plates is of the form

$$I(t) \propto \sum_i N_i e^{-t/\tau_i} \cos(\omega_i t + \phi_i), \tag{8.21}$$

where N_i, τ_i, ω_i, and ϕ_i are the number, the damping constant, the cyclotron frequency, and the initial phase of different types of ions, respectively.

8.3 Fourier transforms in ICR mass spectrometry

In principle, in an ICR mass spectrometer, an infinitely long excitation at the frequency ω_c would accelerate only ions in exact resonance. In practice, however, the excitation must be accomplished by a short pulse of duration T. We remember that truncation of a signal in the time domain corresponds to broadening of the spectrum by a sinc function in the frequency domain. The power spectrum of a harmonic excitation pulse is a sinc^2 function, and the ICR orbital radius obtained by ions of certain ω_c is proportional to the value of the power spectrum at that frequency ω_c. Consequently, also off-resonant ions are accelerated, but less than exactly resonant ions. The FWHM of the main peak of the power spectrum is $1.7718/2T$.

Since the ICR signal strength is proportional to the ICR orbital radius of ions, the ICR signal induced after excitation by a short harmonic pulse varies very strongly with frequency: off-resonant ions have smaller orbits than resonant ions. In order to measure the amount of ions of different m/q values, it would be ideal if ions of a wide frequency range would have the same orbital radius. This situation can be produced, if the excitation signal is tailored in such a way, that its power spectrum is flat over a selected frequency band. The correct excitation waveform can be found by first choosing the desired power spectrum, and then applying Fourier transforms. This method is generally called the *stored-waveform inverse Fourier transform*, SWIFT, excitation.

The time-domain signal given by an ICR measurement is Fourier transformed, and a frequency spectrum is obtained:

$$\mathcal{F}^{-1}\{I(t)\} = N(\omega). \tag{8.22}$$

This spectrum reveals the cyclotron frequencies of the accelerated ions. All the properties of Fourier transforms are available, and the signal and the spectrum may be processed by the various methods described in Chapters 11–13. The mass spectrum of ions is obtained with the help of Equation 8.7.

Figure 8.4 is an example of a high-resolution ICR measurement, made at the University of Joensuu. The figure shows the mass number m/z of a protein molecule. m/z is the mass in atomic mass units divided by the charge as the multiple of the elementary charge.

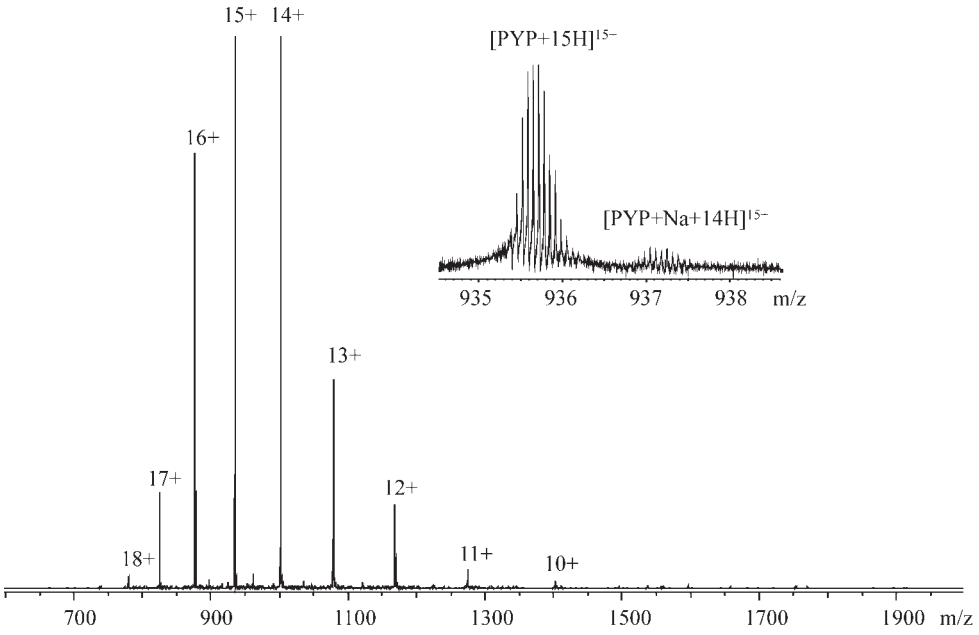


Figure 8.4: Electrospray ionization Fourier transform mass spectrum of photoactive yellow protein PYP (10 μM in MeCN/H_2O (1/1, v/v) + 1 % HAc). Inset shows the expansion of the isotopically resolved pattern of the 15+ charge state. The goal of the measurement was to determine the molecular mass of PYP at high accuracy. The mass of the protein was averaged over the observed charge states and determined to be 14018.85 Da, that is in good agreement with the theoretical average molecular mass 14018.78 Da of PYP containing a thioester-linked 4-hydroxycinnamyl chromophore. *Courtesy of Marjaana Nousiainen, Department of Chemistry, University of Joensuu.*

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

9 Diffraction and Fourier transform

9.1 Fraunhofer and Fresnel diffraction

Fraunhofer diffraction takes place when a plane wave is distorted by an obstacle and the rays which leave the obstacle in the same direction interfere. The dimensions of the obstacle must be comparable to the wavelength of the wave. The incident rays are considered parallel, and the rays received by the observer are also, effectively, parallel. The incident and the diffracted beam are plane waves if the source and the screen are at an infinite distance from the obstacle. Plane waves may also be achieved by lenses or mirrors. Figure 9.1 illustrates Fraunhofer diffraction by a narrow slit.

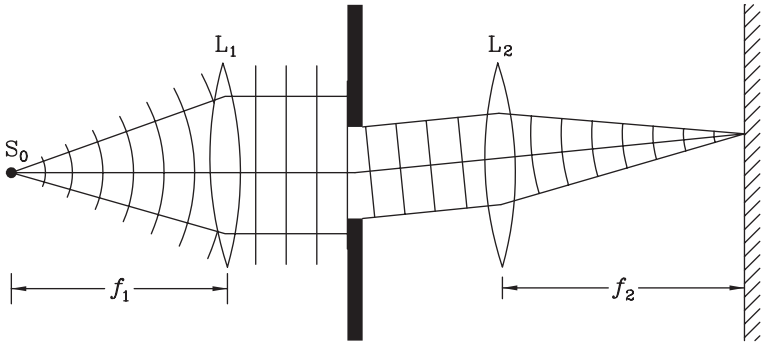


Figure 9.1: Fraunhofer diffraction by a narrow slit. The spherical wave from the point source S_0 is turned into a plane wave by the collimator lens L_1 of focal length f_1. Behind the slit, the plane wave is focused on the screen by the lens L_2 of focal length f_2.

Diffraction is called *Fresnel diffraction* if the conditions of Fraunhofer diffraction are not satisfied. Fresnel diffraction takes place when a spherical wave is distorted by an obstacle which has dimensions comparable to the wavelength of the wave. The incident wave is spherical if it originates from a point source at a finite distance from the obstacle. Diffraction is Fresnel diffraction also if the wave front between the obstacle and the screen consists of spherical waves. This is the case if a particular observation point is at a finite distance from the obstacle. Fresnel diffraction by a narrow slit is illustrated in Figure 9.2.

Mathematical calculation of Fresnel diffraction is complicated, because the intensity of

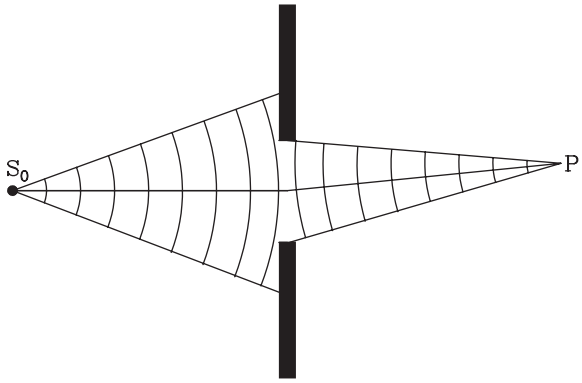


Figure 9.2: Fresnel diffraction by a narrow slit. A spherical wave originates from the point source S_0 and is observed at the observation point P.

spherical waves changes with distance, and this makes their treatment laborious. In the following sections we shall concentrate on Fraunhofer diffraction, because the intensity of plane waves stays constant, and calculations are simpler than for Fresnel diffraction. By examining Fraunhofer diffraction, the principles of diffraction can be better clarified.

9.2 Diffraction through a narrow slit

An obstacle may be mathematically described by a *transmission function*. The transmission function tells how large part of incident radiation can pass a point of an obstacle.

If the origin of the coordinates is fixed at the center of the slit, the transmission function $B(x)$ of a slit of width D can be written as

$$B(x) = \begin{cases} 1, & |x| \leq D/2, \\ 0, & |x| > D/2, \end{cases} \tag{9.1}$$

which is a boxcar function. This function is shown in Figure 9.3.

Let us examine the situation in Figure 9.4, where a plane wave E hits a narrow slit perpendicularly. Let us denote $\mathrm{d}E_1$ and $\mathrm{d}E_2$ two rays of infinitesimal width $\mathrm{d}x$ leaving the slit at the angle θ with respect to the direction of incidence. θ is assumed to be very small, so that $\mathrm{d}x \approx \mathrm{d}x \cos\theta$. We shall choose $\mathrm{d}E_1$ to leave the slit at the center, and $\mathrm{d}E_2$ at a distance x from the center. A ray $\mathrm{d}E$ of the width $\mathrm{d}x$ may be written as

$$\mathrm{d}E = E_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\mathrm{d}x = E_0 e^{i2\pi r/\lambda} e^{-i\omega t}\mathrm{d}x, \tag{9.2}$$

where E_0 is the amplitude of the plane wave per unit length. With the help of the transmission function $B(x)$, we can write

$$\begin{aligned} \mathrm{d}E_1 &= B(0) E_0 e^{i2\pi r_0/\lambda} e^{-i\omega t}\mathrm{d}x, \text{ and} \\ \mathrm{d}E_2 &= B(x) E_0 e^{i2\pi (r_0 + x\sin\theta)/\lambda} e^{-i\omega t}\mathrm{d}x. \end{aligned}$$

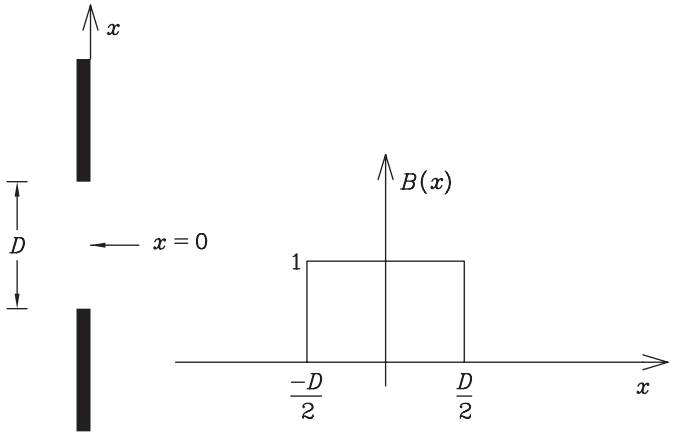


Figure 9.3: A slit of the width D (left), and its transmission function $B(x)$ (right).

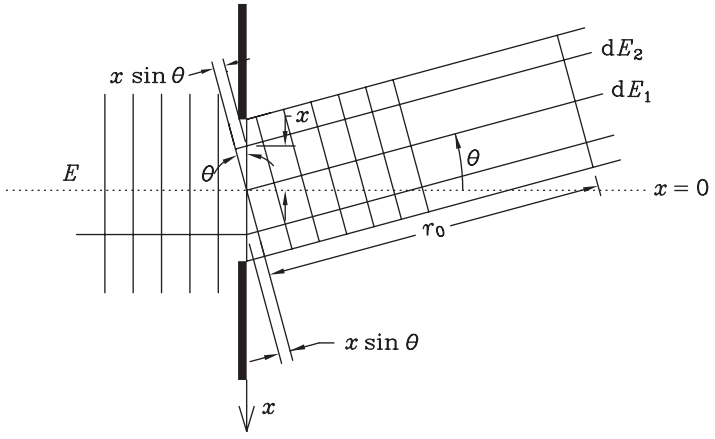


Figure 9.4: A slit, an incident plane wave E, and two rays $\mathrm{d}E_1$ and $\mathrm{d}E_2$ leaving the slit at different places, propagating at the angle θ. (In the figure, $\theta > 0$.)

All rays diffracting at the same angle θ interfere, and the total wave in that direction is

$$E_{\text{out}} = \sum_i \mathrm{d}E_i = e^{-i\omega t} \int_{-\infty}^{\infty} B(x) E_0 e^{i2\pi(r_0 + x\sin\theta)/\lambda}\,\mathrm{d}x. \tag{9.3}$$

The intensity $I_{\text{out}} = E^*_{\text{out}} E_{\text{out}}$. Since $e^{-i\omega t} e^{i\omega t} = 1$, we can neglect the time dependency. The term $E_0 e^{i2\pi r_0/\lambda}$ is constant, and can also be neglected. Consequently, the plane wave in the

direction θ is of the form

$$E_{\text{out}} = \int_{-\infty}^{\infty} B(x) e^{i2\pi \nu x \sin\theta} \, \mathrm{d}x. \tag{9.4}$$

Let us change coordinates, and use the *spatial coordinate*

$$s = \nu \sin\theta. \tag{9.5}$$

The wave can then be written as

$$E_{\text{out}}(s) = \int_{-\infty}^{\infty} B(x) e^{i2\pi s x} \, \mathrm{d}x. \tag{9.6}$$

This is the Fourier transform of $B(x)$. We can see, that *the spatial amplitude distribution of the plane wave leaving the obstacle is the Fourier transform of the transmission function of the obstacle*, $\mathcal{F}\{B(x)\}$. This is a very useful result, which can be generalized also for other obstacles than one narrow slit.

The Fourier transform of a boxcar function is a sinc function, and we obtain that the plane wave leaving one narrow slit is

$$E_{\text{out}}(s) = \mathcal{F}\{B(x)\} = D \operatorname{sinc}(D\pi s). \tag{9.7}$$

Consequently, we can obtain the intensity as a function of the spatial coordinate s or the direction θ:

$$I_{\text{out}}(s) = E^*_{\text{out}}(s) E_{\text{out}}(s) = D^2 \operatorname{sinc}^2(D\pi s) = I_0 \frac{\sin^2\left(\frac{\pi D \sin\theta}{\lambda}\right)}{\left(\frac{\pi D \sin\theta}{\lambda}\right)^2}. \tag{9.8}$$

9.3 Diffraction through two slits

Let us examine two slits, both of the width D, which are situated at a distance d from each other. The transmission function of two slits consists of two boxes. As shown in Figure 9.5, it can be expressed as the convolution of a transmission function of one slit $B_D(x)$ and a function describing the places of the slits $B_d(x) = \delta(x - d/2) + \delta(x + d/2)$,

$$B(x) = B_D(x) * B_d(x). \tag{9.9}$$

The plane wave leaving the slits at the spatial coordinate s is the Fourier transform of the transmission function

$$E_{\text{out}}(s) = \mathcal{F}\{B(x)\} = \mathcal{F}\{B_D(x)\}\,\mathcal{F}\{B_d(x)\} = D \operatorname{sinc}(\pi D s)\, 2\cos\left(2\pi \frac{d}{2} s\right). \tag{9.10}$$

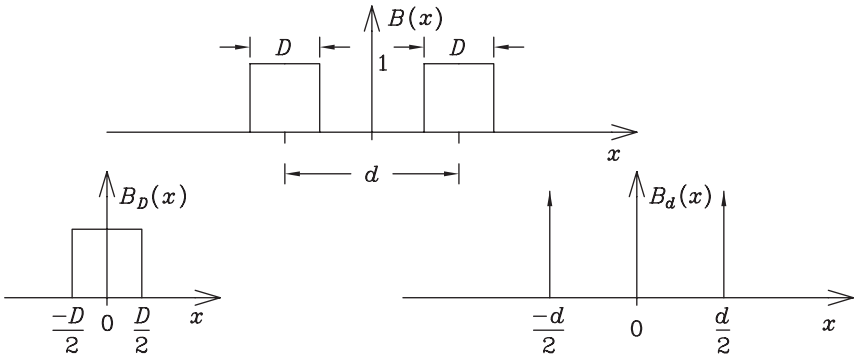


Figure 9.5: Transmission function $B(x)$ of two slits is obtained as the convolution the transmission function $B_D(x)$ of one slit and the function $B_d(x) = \delta(x - d/2) + \delta(x + d/2)$ describing the locations of the slits.

The intensity of the wave is

$$\boxed{I_{\text{out}} = I_0 \frac{\sin^2 \alpha}{\alpha^2} \cos^2 \beta \,,} \tag{9.11}$$

where

$$\begin{cases} \alpha &= \pi D s &= \pi D \nu \sin\theta, \\ \beta &= \pi d s &= \pi d \nu \sin\theta. \end{cases} \tag{9.12}$$

Figure 9.6 illustrates the intensity on the screen behind two slits in the case $D{:}d = 2{:}5$.

Instead of using the transmission function of two slits, Equation 9.10 could also be obtained using the transmission function of one slit and the shift theorem of the Fourier transforms:

$$\begin{aligned} E_{\text{out}}(s) &= \mathcal{F}\{B_D(x + d/2)\} + \mathcal{F}\{B_D(x - d/2)\} \\ &= D\,\text{sinc}(\pi D s) e^{-i2\pi s d/2} + D\,\text{sinc}(\pi D s) e^{i2\pi s d/2} \\ &= 2D\,\text{sinc}(\pi D s)\,\cos(\pi d s). \end{aligned} \tag{9.13}$$

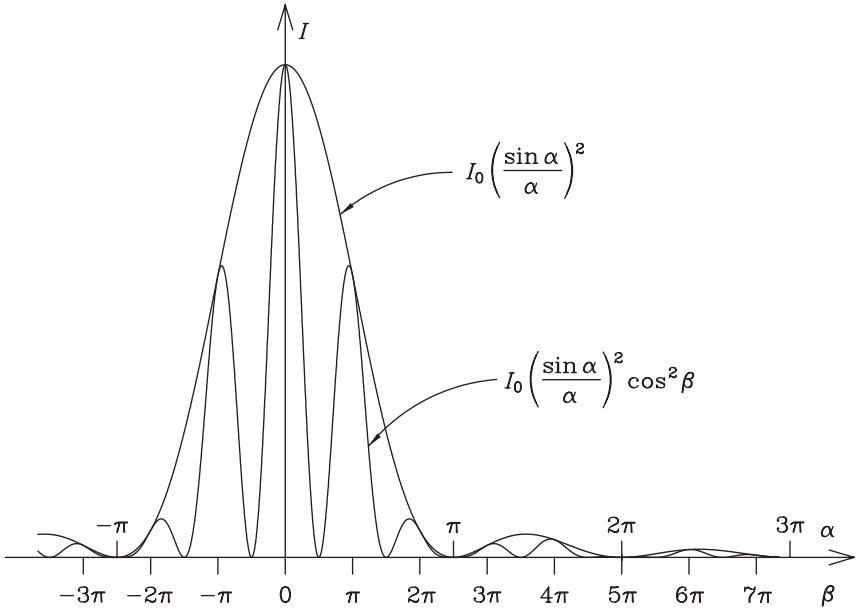


Figure 9.6: Intensity $I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2 \cos^2\beta$ of a plane wave diffracted by two slits, if the ratio of the width D of the slits and the distance d of the slits is $D{:}d = 2{:}5$.

9.4 Transmission grating

A system of a large number of equally spaced parallel slits of equal width is called a *transmission grating*. Figure 9.7 illustrates the transmission function $B(x)$ of a grating with an odd number of slits. If a grating consists of N slits of width D, equally spaced with the distance d,

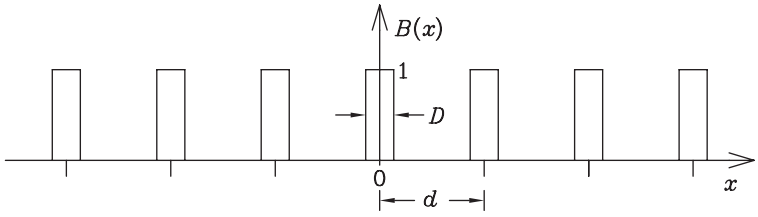


Figure 9.7: The transmission function $B(x)$ of a grating with an odd number N of slits. The width of each slit is D and the distance of successive slits is d ($d > D$).

then the transmission function of the grating is (see also Problem 3)

$$B(x) = \Pi_D(x) * \left[\Pi_{Nd}(x) \sum_{j=-\infty}^{\infty} \delta(x - jd) \right] = \Pi_D(x) * \left[\Pi_{Nd}(x) \frac{\text{Ш}(x, d)}{d} \right], \quad (9.14)$$

where $\Pi_D(x)$ is a boxcar function of width D, which is the transmission function of one slit, and $\Pi_{Nd}(x)$ is a boxcar function of width Nd, which determines the overall length of the grating. The *comb function*

$$\text{Ш}(x, d) = \sum_{j=-\infty}^{\infty} \delta\left(\frac{x}{d} - j\right) \overset{(1.13)}{=} \sum_{j=-\infty}^{\infty} d\,\delta(x - jd) \quad (9.15)$$

determines the locations of the slits. These three functions are shown in Figure 9.8.

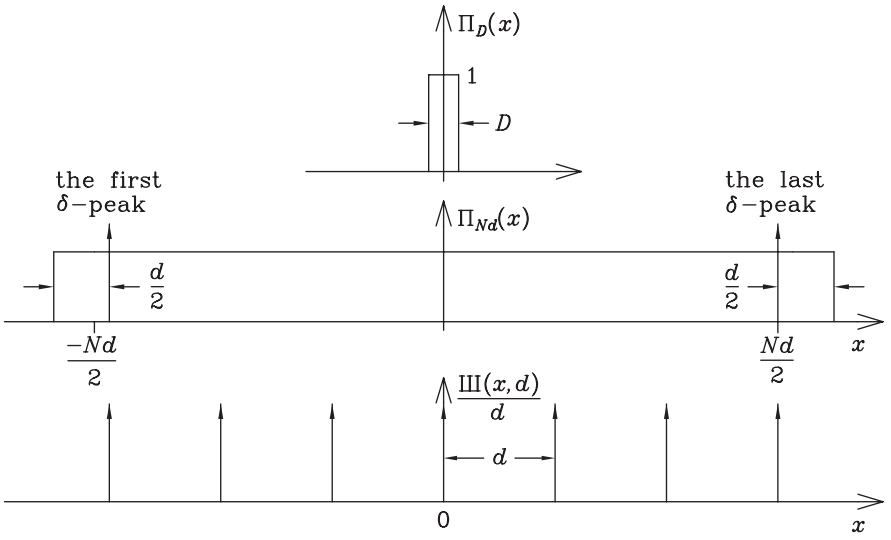


Figure 9.8: The transmission function of one slit $\Pi_D(x)$, the boxcar function $\Pi_{Nd}(x)$ which determines the total length of a grating, and the comb function $\text{Ш}(x, d)/d$ which determines the locations of the slits.

The wave diffracted through a grating and propagating at the angle θ with respect to the normal of the grating, which is also the direction of the incident beam, is shown in Figure 9.9.

In the same way as in the case of one slit or two slits, the amplitude distribution of the diffracted wave is obtained as the Fourier transform of the transmission function, that is, $\mathcal{F}\{B(x)\} = E_{\text{out}}(s)$, where the argument $s = \sin\theta/\lambda = \nu \sin\theta$. It can be shown that the Fourier transform of the comb function

$$\mathcal{F}\{\text{Ш}(x, d)\} = d\,\text{Ш}(s, 1/d). \quad (9.16)$$

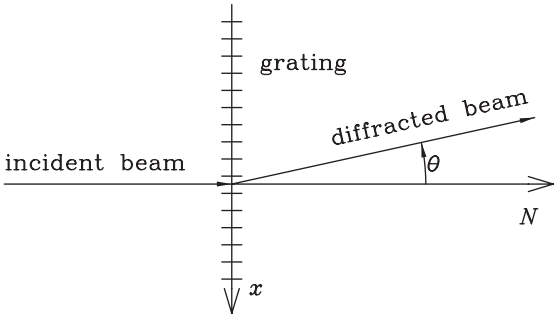


Figure 9.9: The direction of the incoming wave, which is the direction ***N*** normal to the plane of the grating, and the direction of the diffracted wave, which makes an angle θ with respect to the normal ***N***.

Consequently, we can obtain the amplitude distribution of the wave diffracted through a grating:

$$\begin{aligned} E_{\text{out}}(s) &= \mathcal{F}\{B(x)\} = \mathcal{F}\{\Pi_D(x)\}\left[\mathcal{F}\{\Pi_{Nd}(x)\} * \mathcal{F}\{\text{Ш}(x,d)/d\}\right] \\ &= D\,\text{sinc}(\pi Ds)[Nd\,\text{sinc}(\pi Nds) * \text{Ш}(s,1/d)] \\ &\overset{(1.13)}{=} D\,\text{sinc}(\pi Ds)\left[N\,\text{sinc}(\pi Nds) * \sum_{k=-\infty}^{\infty}\delta\left(s-\frac{k}{d}\right)\right] \\ &\overset{(3.11)}{=} D\,\text{sinc}(\pi Ds)\sum_{k=-\infty}^{\infty} N\,\text{sinc}\left[\pi Nd(s-k/d)\right]. \end{aligned} \tag{9.17}$$

This distribution pattern is illustrated in Figure 9.10.

Once again, the intensity of the wave may be obtained from the equation

$$I_{\text{out}}(s) = E^*_{\text{out}}(s)E_{\text{out}}(s). \tag{9.18}$$

We can see that the principal intensity maxima are obtained at the spatial coordinates

$$s = \frac{\sin\theta}{\lambda} = \frac{k}{d}. \tag{9.19}$$

The condition of the intensity maxima can be written as

$$\boxed{d\sin\theta = k\lambda,} \tag{9.20}$$

where k is an integer. The maxima with different values of k are the different *diffraction orders*.

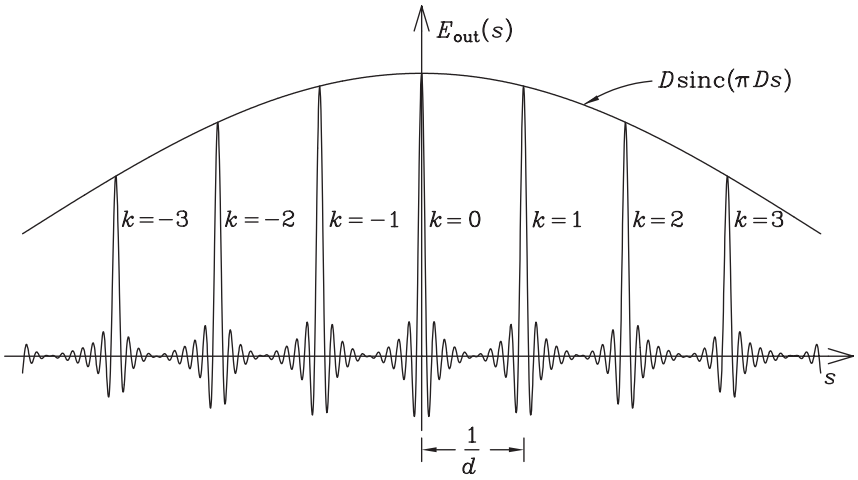


Figure 9.10: The wave diffracted through a grating consisting of N slits of width D and distance d has the amplitude distribution $E_{\text{out}}(s) = D\,\text{sinc}(\pi Ds) \sum_{k=-\infty}^{\infty} N\,\text{sinc}\,[\pi Nd(s - k/d)]$ as the function of the spatial coordinate s.

From Equations 9.17 and 9.18 we can calculate the intensity

$$I_{\text{out}} = I_0\,\text{sinc}^2(\pi Ds)\left\{\sum_{k=-\infty}^{\infty}\text{sinc}\left[\pi Nd\left(s-\frac{k}{d}\right)\right]\right\}^2 = I_0\,\text{sinc}^2(\pi Ds)\,\frac{\sin^2(\pi Nds)}{N^2\sin^2(\pi ds)}, \tag{9.21}$$

because, if N is an odd integer,

$$\sum_{k=-\infty}^{\infty}\text{sinc}\left[\pi Nd\left(s-\frac{k}{d}\right)\right] = \frac{\sin(\pi Nds)}{N\sin(\pi ds)}, \tag{9.22}$$

which can be verified through the serial expansion

$$\frac{1}{\sin x} = \sum_{j=-\infty}^{\infty}\frac{(-1)^j}{x-\pi j}. \tag{9.23}$$

The intensity can also be written as

$$\boxed{I_{\text{out}} = I_0\,\frac{\sin^2\alpha}{\alpha^2}\,\frac{\sin^2 N\beta}{N^2\sin^2\beta}}, \tag{9.24}$$

where

$$\begin{cases}\alpha &= \pi Dv\sin\theta,\\ \beta &= \pi dv\sin\theta.\end{cases}$$

The intensity distribution of the grating is, of course, valid also for an even integer N, which is easy to show (see Problem 2).

In spectroscopy, it is common to use *diffraction gratings* instead of transmission gratings. Diffraction gratings are based on interference of reflected beams, instead of transmitted, and they give a larger intensity. The diffraction pattern of a diffraction grating is rather similar to the diffraction pattern of the transmission grating.

Example 9.1: Two identical transmission gratings are placed behind an illuminated aperture, as shown in Figure 9.11. The width of the slits in the gratings is a, the slits are positioned in the intervals of $2a$, and the entire number of slits which fit into the length of the illuminated aperture is N. Initially, the slits of the two gratings lie on the top of each other so that the transmission function of the system is the same as the transmission function of one grating. Which diffraction orders are observed? The second grating is then moved perpendicularly to the direction of the slits. How will the interference pattern change?

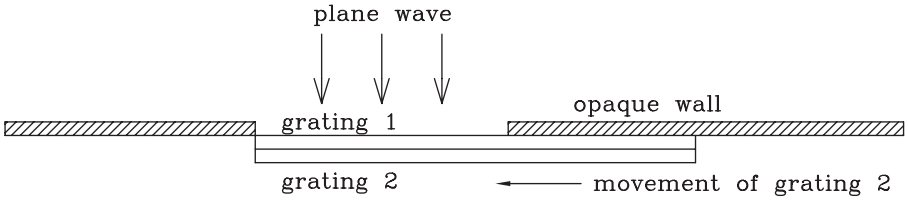


Figure 9.11: The gratings of Example 9.1.

Solution. Combination of the two gratings is one grating, consisting of N slits in the intervals of $2a$. Only the width of the slits of the combined system changes as the second grating is moved. Initially, the width D of the slits of the combined system is a, but as the second grating is moved a distance l the width decreases as $D = a - l$ until $D = 0$ at $l = a$. After that, the width of the slits grows again, and reaches the maximum $D = a$ at $l = 2a$. In this way, D changes periodically as the second grating is moved, with the period $2a$.

The intensity of the diffracted beam is

$$I_{\text{out}}(s) = I_0 \operatorname{sinc}^2(\pi D s)\, \frac{\sin^2(\pi N d s)}{N^2 \sin^2(\pi d s)},$$

where $d = 2a$. The last part of the formula,

$$I_G(s) = \frac{\sin^2(\pi N d s)}{N^2 \sin^2(\pi d s)},$$

describes a rapidly changing interference pattern (compare this to Figure 9.10), which has the intensity maxima, or the diffraction orders, at

$$\sin(\pi 2 a s) = 0, \text{ or } s = 0, \pm\frac{1}{2a}, \pm\frac{1}{a}, \pm\frac{3}{2a}, \ldots$$

This part of the formula stays constant as the second grating is moved.

The part

$$I_D(s) = \operatorname{sinc}^2(\pi D s)$$

describes a slowly changing diffraction pattern. The positions of its zeroes and maxima change as the second grating is moved and the width of the slits of the combined system changes.

If $l = 0, 2a, 4a, \ldots$, then $D = a$, and the zeroes of $I_D(s)$ lie at $\pi D s = k\pi$, $k \neq 0$, or $s = \pm\frac{1}{a}, \pm\frac{2}{a}, \pm\frac{3}{a}, \ldots$ This means that even diffraction orders (excluding zeroth order) vanish.

If l grows (at $l < a$), then D diminishes, and the minima of $I_D(s)$ move farther. This means that intensity is transferred from low orders of magnitude to higher orders of magnitude. At the same time, the total intensity decreases, as the slits become narrower.

In the limit $l \to a$, $D \to 0$, the function $I_D(s)$ approaches a constant function, and all orders of magnitude are seen as strong. Of course, the intensity of the diffraction pattern also tends to zero.

9.5 Grating with only three orders

Let us consider if it is possible to make a grating which would give, instead of an infinite number of diffraction orders, only exactly three orders -1, 0 and 1, as shown in Figure 9.12. If such a grating would exist, the amplitude distribution of the diffracted wave would be something like

$$E_{\text{out}}(s) = \delta(s) + \delta\left(s - \frac{1}{d}\right) + \delta\left(s + \frac{1}{d}\right). \tag{9.25}$$

A transmission function which would give this distribution would be

$$B(x) = \mathcal{F}^{-1}\{E_{\text{out}}(s)\} = 1 + 2\cos(2\pi x/d), \tag{9.26}$$

which is a function whose values vary between -1 and 3. However, a true transmission function varies in the range $[0, 1]$. Thus, the amplitude distribution of Equation 9.25 is impossible.

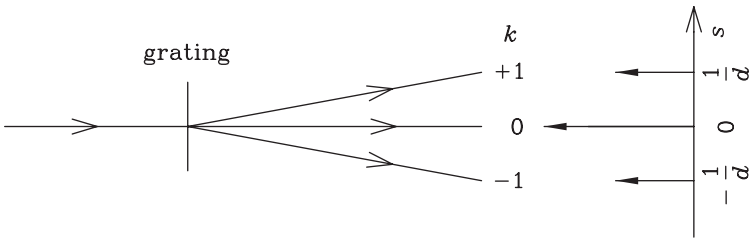


Figure 9.12: The amplitude distribution behind a grating which gives exactly three diffraction orders, which are $k = -1$, 0, and 1.

If we choose the amplitude distribution to be

$$E_{\text{out}}(s) = \frac{1}{2}\,\delta(s) + \frac{1}{4}\,\delta\left(s - \frac{1}{d}\right) + \frac{1}{4}\,\delta\left(s + \frac{1}{d}\right), \tag{9.27}$$

then the transmission function is

$$\boxed{B(x) = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi x}{d}\right).} \tag{9.28}$$

This function is shown in Figure 9.13. We can see that it is a sensible transmission function. A grating which has this transmission function gives a diffracted wave which has exactly three diffraction orders. However, the central order has larger intensity than the other two orders.

In reality, however, a grating has a finite length l, and the truncation of the transmission function causes that the true $E_{\text{out}}(s)$ does not consist of Dirac's delta functions, but is convolved with the function

$$\frac{l\sin\alpha}{\alpha}, \qquad \alpha = \pi l \nu \sin\theta.$$

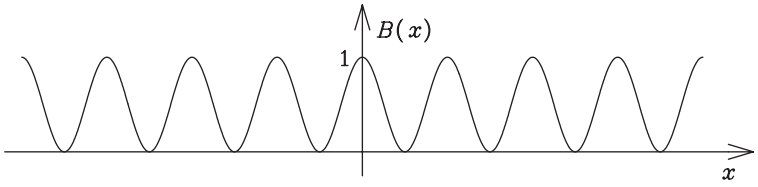


Figure 9.13: The function $B(x) = \frac{1}{2} + \frac{1}{2}\cos(2\pi x/d)$. A grating with this transmission function gives exactly three diffraction orders.

9.6 Diffraction through a rectangular aperture

Let us next study a two-dimensional case, the diffraction pattern produced by a rectangular aperture. We can set the coordinates in such a way that the aperture lies in the xy-plane and the origin is in the center of the aperture. Let us denote X the width of the aperture in the x-direction and Y the width of the aperture in the y-direction. This rectangular aperture is illustrated in Figure 9.14. Let us denote $\boldsymbol{n}$ the normal vector of the aperture plane. We shall examine a diffracted wave which propagates in the direction $\boldsymbol{n}'$ which is determined by angles θ_y and θ_x with respect to the direction $\boldsymbol{n}$, as depicted in Figure 9.14.

The path difference of the rays leaving an aperture in the origin and at the point (x, y) is $x\sin\theta_x + y\sin\theta_y$. We can write the wave diffracted by any two-dimensional aperture by using the two-dimensional transmission function $B(x, y)$ as

$$E_{\text{out}}(\theta_x, \theta_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x, y)e^{i2\pi\nu(x\sin\theta_x + y\sin\theta_y)}\,\mathrm{d}x\,\mathrm{d}y. \tag{9.29}$$

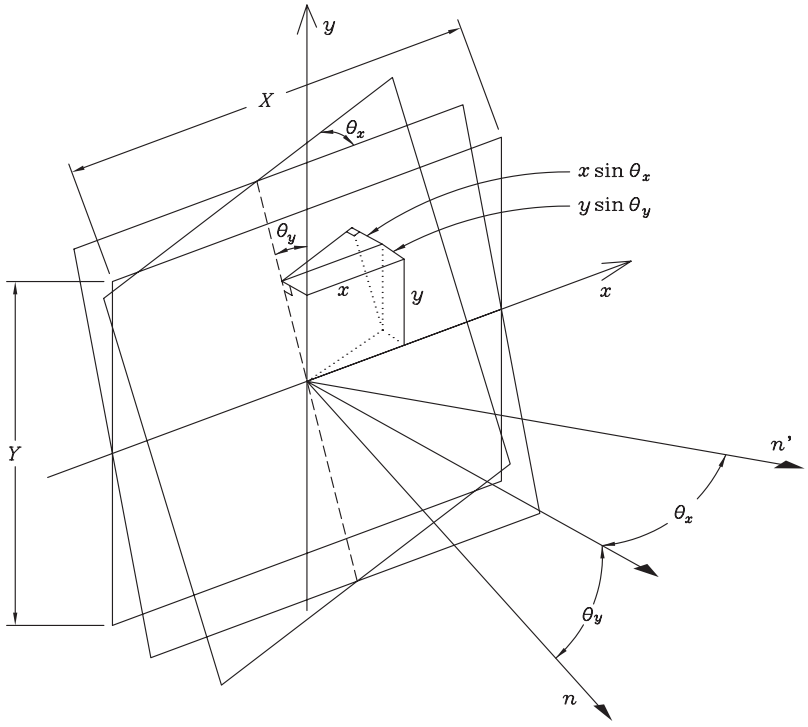


Figure 9.14: A rectangular aperture of size $X \times Y$ in the xy-plane. The plane normal to the direction $\boldsymbol{n}'$ of the diffracted wave is obtained by first turning the plane of the aperture, of normal $\boldsymbol{n}$, by angle θ_y around x-axis and then turning this new plane by angle θ_x around the normal projection of y-axis on the new plane. The path difference of the rays leaving the aperture in the origin and at the point (x, y) is $x \sin\theta_x + y \sin\theta_y$. In the figure $\theta_x < 0$ and $\theta_y < 0$.

If we use the spatial coordinates

$$\begin{cases} s &= \nu \sin\theta_x, \\ p &= \nu \sin\theta_y, \end{cases} \tag{9.30}$$

we can see that the diffracted wave is obtained by the two-dimensional Fourier transform

$$E_{\text{out}}(s, p) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, y) e^{i2\pi(sx+py)} \, \mathrm{d}x \, \mathrm{d}y = \mathcal{F}\{B(x, y)\}. \tag{9.31}$$

If we can separate $B(x, y) = B(x)B(y)$, we can write

$$E_{\text{out}}(s, p) = \int_{-\infty}^{\infty} B(x) e^{i2\pi sx} \, \mathrm{d}x \int_{-\infty}^{\infty} B(y) e^{i2\pi py} \, \mathrm{d}y, \tag{9.32}$$

and, consequently,

$$E_{\text{out}}(s, p) = E_{\text{out}}(s)E_{\text{out}}(p) = \mathcal{F}\{B(x)\}\mathcal{F}\{B(y)\}. \tag{9.33}$$

The transmission function of the rectangular aperture of size $X \times Y$ can be written as $B(x, y) = B_x(x)B_y(y)$, where $B_x(x)$ is a boxcar function of length X and $B_y(y)$ a boxcar function of length Y, as shown in Figure 9.15. Consequently, we can find the amplitude distribution of the wave diffracted by a rectangular aperture:

$$E_{\text{out}}(s, p) = E_{\text{out}}(s)E_{\text{out}}(p) = X\,\text{sinc}(X\pi s)Y\,\text{sinc}(Y\pi p). \tag{9.34}$$

The intensity of the diffracted wave is

$$I_{\text{out}}(s, p) = I_0\,\text{sinc}^2(X\pi s)\,\text{sinc}^2(Y\pi p). \tag{9.35}$$

If we return to the variables θ_x and θ_y, we obtain

$$I_{\text{out}}(\theta_x, \theta_y) = I_0\,\frac{\sin^2(X\pi\,\sin\theta_x/\lambda)}{(X\pi\,\sin\theta_x/\lambda)^2}\,\frac{\sin^2(Y\pi\,\sin\theta_y/\lambda)}{(Y\pi\,\sin\theta_y/\lambda)^2}. \tag{9.36}$$

Let us examine a quadratic aperture of size $D \times D$, and the special case where the direction of the diffracted wave is such that $\theta_x = \theta_y$. We shall use new coordinates (z_1, z_2), which are obtained by rotating the coordinates (x, y) by an angle $-\pi/4$, as shown in Figure 9.16. If $\theta_x = \theta_y$, then, in the new coordinates, $\theta_{z_2} = 0$, and $\theta_{z_1} \approx \sqrt{2}\theta_x$. The spatial coordinate $q = \nu\sin\theta_{z_1} \approx \nu\sqrt{2}\,\sin\theta_x$. We know that the Fourier transform of a triangle is $\mathcal{F}\{\Lambda_L(x)\} = L^2\,\text{sinc}^2(\pi\nu L)$, where the triangle function

$$\Lambda_L(x) = \begin{cases} \dfrac{L - |x|}{L}, & |x| \le L, \\ 0, & |x| > L. \end{cases} \tag{9.37}$$

We notice that the transmission function of the quadratic aperture can be expressed in one dimension with the help of the triangle function as $B(z_1) = 2\Lambda_L(z_1)$, where $L = \sqrt{2}D/2$.

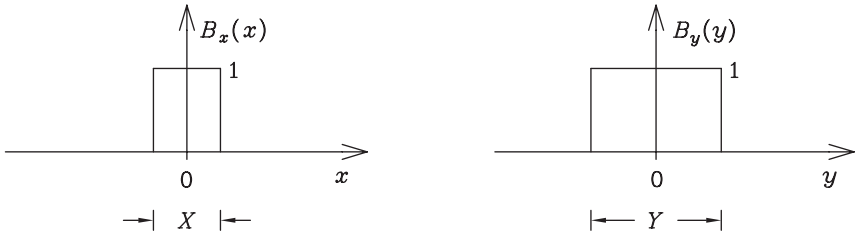


Figure 9.15: The x-component $B_x(x)$ and the y-component $B_y(y)$ of the transmission function $B(x, y) = B_x(x)B_y(y)$ of a rectangular aperture of size $X \times Y$.

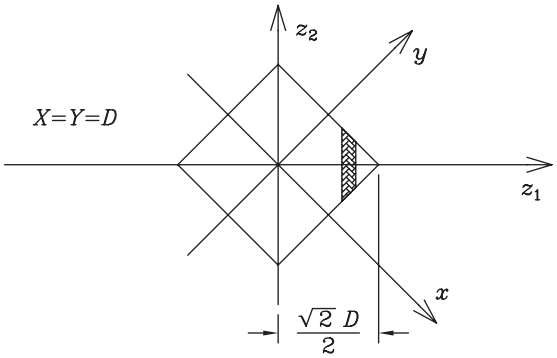


Figure 9.16: A quadratic aperture of size $X \times Y = D \times D$, and the coordinates (z_1, z_2) obtained by rotating the coordinates (x, y) by an angle $-\pi/4$.

Consequently, we can write the wave diffracted in the direction determined by the spatial coordinate q as

$$E_{\text{out}}(q) = \mathcal{F}\{B(z_1)\} = D^2 \operatorname{sinc}^2\left(\pi q \frac{\sqrt{2}}{2} D\right). \tag{9.38}$$

The intensity of this wave is

$$I_{\text{out}} = I_0 \operatorname{sinc}^4(\pi D \sin\theta_x/\lambda). \tag{9.39}$$

This is consistent with our earlier result, Equation 9.35, when $X = Y = D$ and $\theta_x = \theta_y$.

Figure 9.17 is a simulation of the diffraction pattern of a rectangular aperture.

Example 9.2: Determine the diffraction pattern of the symmetric transparent cross in Figure 9.18. (It is sufficient to find the wave $E_{\text{out}}(s, p)$.)

Solution. Let us choose the origin of coordinates in the center of the cross. The transmission function $B(x)$ of the cross is $B(x, y) = B_1(x, y) + B_2(x, y) - B_3(x, y)$, where $B_1(x, y)$ is the transmission function of the rectangular $\frac{-L}{2} \le x \le \frac{L}{2}$, $\frac{-l}{2} \le y \le \frac{l}{2}$, $B_2(x, y)$ is the transmission function of the rectangular $\frac{-l}{2} \le x \le \frac{l}{2}$, $\frac{-L}{2} \le y \le \frac{L}{2}$, and $B_3(x, y)$ is the transmission function of the rectangular $\frac{-l}{2} \le x \le \frac{l}{2}$, $\frac{-l}{2} \le y \le \frac{l}{2}$. Consequently,

$$\begin{aligned} E_{\text{out}}(s, p) &= \mathcal{F}\{B(x, y)\} = \mathcal{F}\{B_1(x, y)\} + \mathcal{F}\{B_2(x, y)\} - \mathcal{F}\{B_3(x, y)\} \\ &= Ll \operatorname{sinc}(\pi Ls) \operatorname{sinc}(\pi lp) + lL \operatorname{sinc}(\pi ls) \operatorname{sinc}(\pi Lp) \\ &\quad -l^2 \operatorname{sinc}(\pi ls) \operatorname{sinc}(\pi lp). \end{aligned}$$

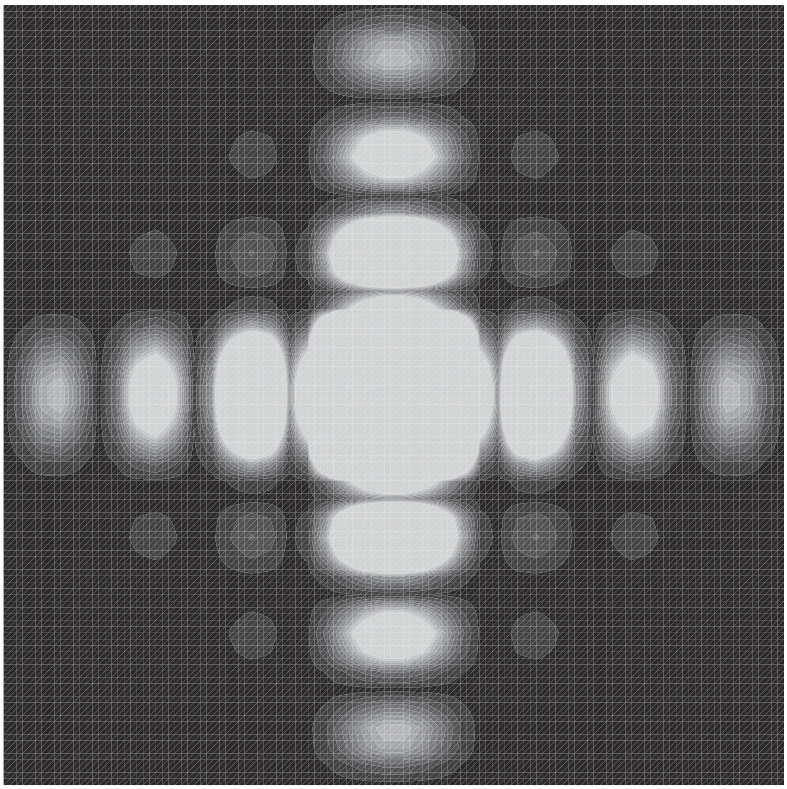


Figure 9.17: A simulation of the diffraction pattern of a rectangular aperture. The diffraction orders $-3, \ldots, 3$ are shown in both directions x and y.

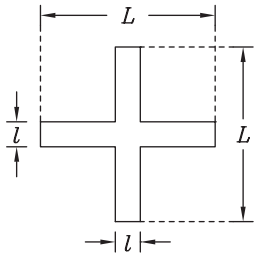


Figure 9.18: The transparent cross of Example 9.2.

9.7 Diffraction through a circular aperture

If the aperture is circular, as shown in Figure 9.19, coordinates can always be chosen in such a way that the phase difference of the rays leaving the aperture depends only on one coordinate x, instead of two coordinates x and y. In the general two-dimensional case, as in Figure 9.14, two angular variables θ_x and θ_y are needed to determine the angular deviation of the diffracted wave, whereas in the case of a circular aperture only one angular variable θ is sufficient. The coordinates are chosen so that $\theta = \theta_x$. The only necessary spatial coordinate is then $s = \nu \sin\theta$.

If the radius of the circle is R, then the integration in the direction y equals multiplication by $2\sqrt{R^2 - x^2}$. The one-dimensional transmission function is therefore

$$B(x) = 2\sqrt{R^2 - x^2}. \tag{9.40}$$

The amplitude distribution of the diffracted wave is

$$\begin{aligned} E_{\text{out}}(s) &= \int_{-\infty}^{\infty} B(x) e^{i2\pi sx}\,\mathrm{d}x = \int_{-R}^{R} 2\sqrt{R^2 - x^2}\, e^{i2\pi sx}\,\mathrm{d}x \\ &\overset{u=\frac{x}{R} \Rightarrow \mathrm{d}u=\frac{\mathrm{d}x}{R}}{=} \int_{-1}^{1} 2R^2\sqrt{1-u^2}\, e^{i2\pi Rsu}\,\mathrm{d}u = 4R^2 \int_{0}^{1} \sqrt{1-u^2}\cos(2\pi Rsu)\,\mathrm{d}u \\ &= \pi R^2 \times 2\,\frac{J_1(2\pi Rs)}{2\pi Rs}, \end{aligned} \tag{9.41}$$

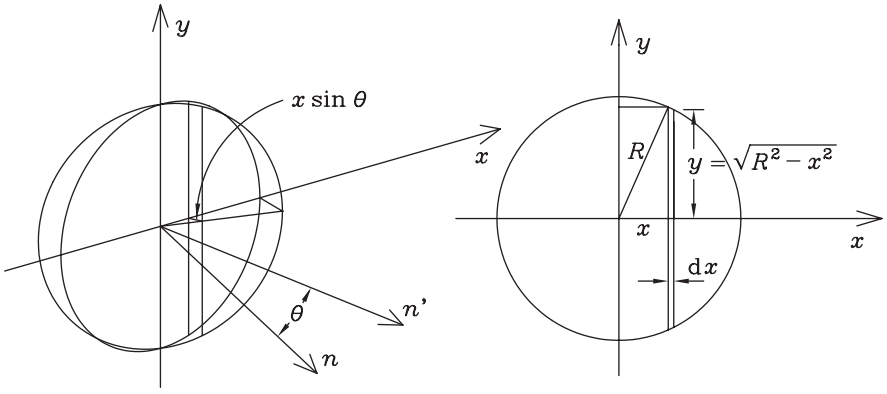


Figure 9.19: Circular aperture of radius R in the xy-plane with normal vector $\boldsymbol{n}$. The angle between $\boldsymbol{n}$ and the propagation direction $\boldsymbol{n}'$ of the diffracted wave is θ. The coordinates (x, y) have been chosen in such a way that the plane normal to $\boldsymbol{n}'$ is obtained by only rotating the aperture plane by the angle θ around the y-axis. The phase shift of the rays leaving from an infinitesimally narrow strip $\mathrm{d}x$ in the direction $\boldsymbol{n}'$ is approximately constant.

and the intensity of the wave is

$$I_{\text{out}}(s) = I_0 \left[\frac{2J_1(2\pi Rs)}{2\pi Rs} \right]^2 = I_0 \left[\frac{2J_1(2\pi R \sin\theta/\lambda)}{2\pi R \sin\theta/\lambda} \right]^2. \tag{9.42}$$

In these equations $J_1(x)$ is the *Bessel function* of the first order. Generally, the Bessel function of order n is

$$J_n(x) = \frac{1}{\pi} \int_0^{\pi} \cos(n\theta - x \sin\theta)\, \mathrm{d}\theta, \tag{9.43}$$

where n, the order, is an integer.

9.8 Diffraction through a lattice

A *lattice* of rectangular apertures is a two-dimensional diffraction obstacle, which consists of two gratings crossing each other, as shown in Figure 9.20. If the size of each aperture of the lattice is $D \times D$, the number of the apertures is $N_x \times N_y$, and the distance of successive apertures in both directions x and y is d, then the transmission function of the lattice is (see Equation 9.14)

$$\begin{cases} B(x) &= \Pi_D(x) * \left[\Pi_{N_x d}(x) \displaystyle\sum_{j=-\infty}^{\infty} \delta(x - jd) \right], \\ B(y) &= \Pi_D(y) * \left[\Pi_{N_y d}(y) \displaystyle\sum_{j=-\infty}^{\infty} \delta(y - jd) \right], \end{cases} \tag{9.44}$$

where $\Pi_D(x)$ and $\Pi_D(y)$ are boxcar functions of length D, which are the transmission functions of individual slits, $\Pi_{N_x d}$ and $\Pi_{N_y d}$ are boxcar functions of length $N_x d$ and $N_y d$, respectively, determining the size of the lattice in directions x and y, and Dirac's delta functions determine the positions of the apertures.

The amplitude distribution of the diffracted wave is, again, the two-dimensional Fourier transform of $B(x, y) = B(x)B(y)$, and the obtained intensity of the wave is (compare with Equation 9.24)

$$I_{\text{out}} = I_0 \operatorname{sinc}^2 \alpha_x \left[\frac{\sin(N_x \beta_x)}{N_x \sin\beta_x} \right]^2 \operatorname{sinc}^2 \alpha_y \left[\frac{\sin(N_y \beta_y)}{N_y \sin\beta_y} \right]^2, \tag{9.45}$$

where

$$\begin{cases} \alpha_x &= \pi D \nu \sin\theta_x, \\ \beta_x &= \pi d \nu \sin\theta_x, \\ \alpha_y &= \pi D \nu \sin\theta_y, \\ \beta_y &= \pi d \nu \sin\theta_y. \end{cases}$$

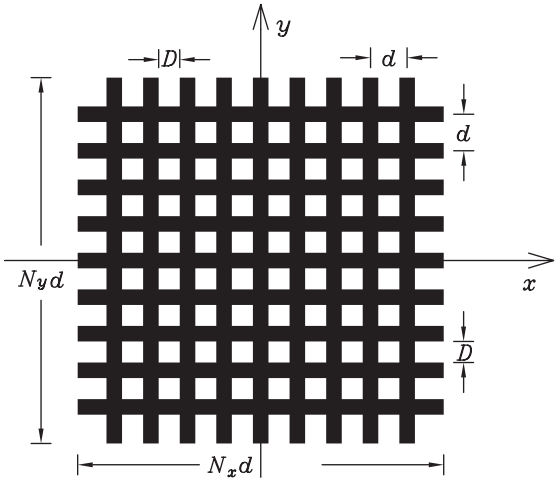


Figure 9.20: A diffraction lattice of rectangular apertures. The size of each aperture is $D \times D$, the number of the apertures is $N_x \times N_y$, and the distance of the apertures is d.

9.9 Lens and Fourier transform

A convex lens focuses parallel rays to one point in the focal plane. Rays of a certain direction are focused to a certain corresponding point.

Let us examine a system, where a narrow slit has been placed in front of a convex lens, as shown in Figure 9.21. If the slit is illuminated by a monochromatic plane wave, then the amplitude of the wave leaving the slit in the direction determined by the spatial coordinate s is the Fourier transform of the transmission function of the slit, $\mathcal{F}\{B(x)\}$. Since the lens focuses parallel rays to the same point, the amplitude distribution obtained on the screen in the focal plane of the lens is

$$E_{\text{out}}(s) = \mathcal{F}\{B(x)\} = \int_{-\infty}^{\infty} B(x) e^{i2\pi s x}\, \mathrm{d}x, \tag{9.46}$$

where

$$s = \nu \sin\theta = \frac{\sin\theta}{\lambda} \approx \frac{x'}{f\lambda}, \tag{9.47}$$

x' being the coordinate of the point in the screen corresponding the spatial coordinate s. The lens thus converts the function E_{out} from *direction (spatial) domain* to *space domain*. The focal plane (x', y') behind the lens, where the amplitude distribution is the Fourier transform of the transmission function of the slit, is called the *transform plane.* The lens which converts the Fourier transform from direction domain to the transform plane is called the *transform lens.*

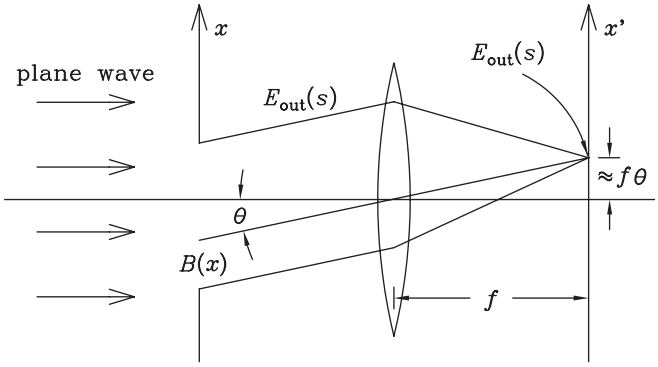


Figure 9.21: A narrow slit is illuminated by a plane wave. Parallel rays leaving the slit are focused by a convex lens to a screen in the focal plane of the lens. The transmission function of the slit is $B(x)$ and the amplitude distribution on the screen is E_{out}.

Figure 9.22 illustrates a two-dimensional system, which consists of a light source, a collimator which converts a spherical wave to a plane wave, a diffraction obstacle, an objective, and a screen. The focal length of the objective is f. The screen is at the distance f behind the objective. If $B(x, y)$ is the transmission function of the obstacle, then the amplitude distribution behind the obstacle in the direction domain is

$$E_{\text{out}}(s, p) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x, y)e^{i2\pi(sx+py)}\,\mathrm{d}x\,\mathrm{d}y = \mathcal{F}\{B(x, y)\} = b(s, p). \tag{9.48}$$

The lens converts the spatial coordinates (s, p) to the point (x', y') on the screen:

$$s = \frac{x'}{\lambda f}, \qquad p = \frac{y'}{\lambda f}. \tag{9.49}$$

In other words, the lens in Figure 9.22 converts the Fourier transform of the transmission function from direction domain to space domain to the focal plane. The intensity distribution seen on the screen in the focal plane is the Fraunhofer diffraction pattern. The convex lens does, of course, form an image of the obstacle somewhere, in some other plane than the transform plane. The position of the image is given by the conventional lens equation. If the distance of the obstacle and the lens is f, then the image is formed at infinity.

An image may be analogically and optically processed by the lens system in Figure 9.23. The system contains two lenses L_1 and L_2. Their focal lengths are f_1 and f_2, respectively. Lens L_1 is placed at the distance f_1 behind the obstacle. Lens L_1 converts the Fourier transform $b(s, p)$ of the transmission function $B(x, y)$ of the obstacle from the direction domain to the $x'y'$-plane, according to Equations 9.48 and 9.49. This distribution pattern $b(s, p)$ is then filtered in a desired way by a suitable filter in the $x'y'$-plane. The second lens L_2 is

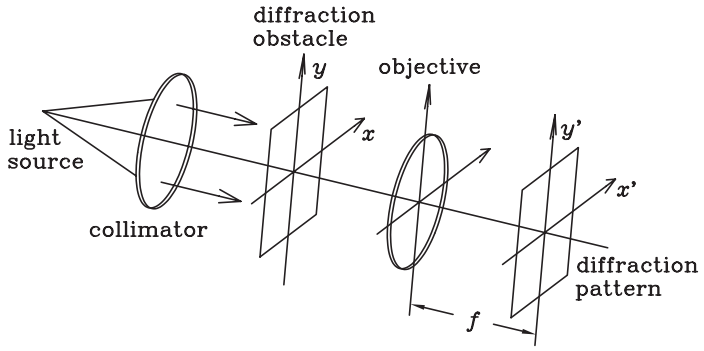


Figure 9.22: A system which takes the Fourier transform of the transmission function of an obstacle in the plane (x, y), and converts the Fourier transform from direction domain to space domain to the transform plane (x', y'), which is the focal plane of the objective. A diffraction pattern is obtained on the screen in the transform plane. The focal length of the objective is f.

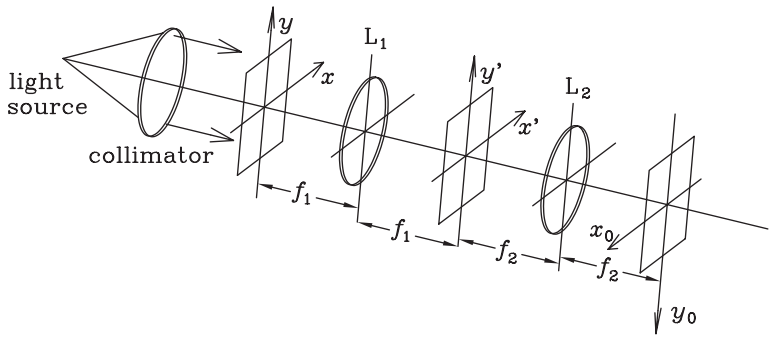


Figure 9.23: A lens system for image processing. The lens L_1 converts the Fourier transform of the transmission function $B(x, y)$ of the obstacle to the $x'y'$-plane, where this amplitude distribution may be processed by a desired filter. The lens L_2 forms a manipulated image of the obstacle in the x_0y_0-plane. The new image is given by an inverse Fourier transform of the manipulated amplitude distribution leaving the transform plane. This image is inverted, and the directions of the coordinates (x_0, y_0) have been chosen respectively.

used to form a new image. The new image is given by an inverse Fourier transform of the manipulated amplitude distribution leaving the transform plane. Since the distance between the obstacle and L_1 was chosen to be f_1, the final image is obtained in the focal plane (x_0, y_0) of the second lens. The second lens may be called an *imaging lens.* The new image is now a manipulated image.

Figure 9.24 illustrates how a plane wave propagates in the direction θ with respect to the optical axis, arrives in the first lens, is focused to a point in the common focal plane of the two lenses, propagates to the second lens, and continues from the second lens as a plane wave in the direction θ' with respect to the optical axis. The direction of the wave leaving the second lens is

$$\theta' \approx -\frac{f_1}{f_2}\,\theta, \tag{9.50}$$

where f_1 is the focal length of L_1 and f_2 is the focal length of L_2. The magnification of the lens system is, according to basic geometrical optics, $-f_1/f_2$.

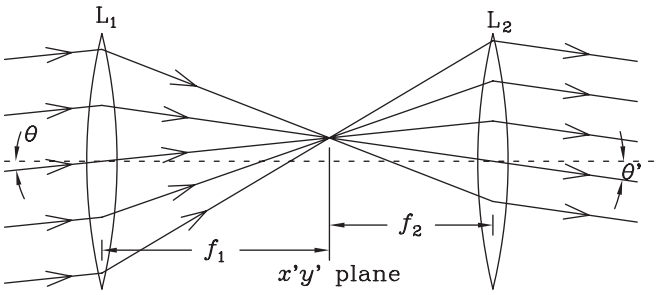


Figure 9.24: Two successive lenses L_1 and L_2 of focal lengths f_1 and f_2, respectively. Prior to lens L_1, the transform $\mathcal{F}\{B(x, y)\}$ is in the direction domain. In the $x'y'$-plane it is in the space domain and can therefore be easily filtered.

If the plane (x', y') is empty, an inverted image of $B(x, y)$, magnified by $-f_1/f_2$, is obtained in the plane (x_0, y_0).

Figure 9.25 shows some *binary filters,* which may be placed in the $x'y'$-plane. Binary filters consist of fully transparent and fully opaque areas. Each of the filters in Figure 9.25 has its own purpose of use. The filters remove rays with certain spatial coordinates; this process is called *spatial filtering*.

Figure 9.26 shows the principle of a lens and a filter system which works as a *low-pass filter*. The first lens converts the Fourier transform from direction domain to space domain, and the second lens forms a new image. A beam whose intensity oscillates with high frequency in the xy-plane has components with a large spatial coordinate $s = \nu \sin\theta$ and, consequently, large coordinates in the $x'y'$-plane. These components are removed in the $x'y'$-plane by a filter, which in this case is a circular aperture. Only low-frequency spectral components reach the image plane.

In this section, we have discussed a method which can be used in filtering or smoothing of an image. These same operations could be achieved also by mathematical treatment of a signal or a spectrum. Mathematical filtering and smoothing will be discussed in Chapter 11.

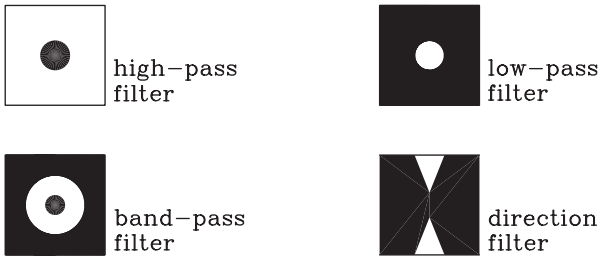


Figure 9.25: Four different binary filters.

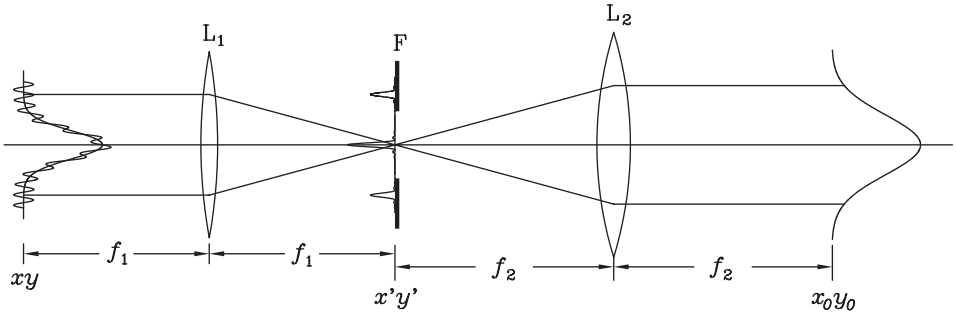


Figure 9.26: The principle of low-pass filtering by lenses. Lens L_1 converts the Fourier transform of the original wave to the $x'y'$-plane. The transformed intensity distribution has tails, which come from the high-frequency spectral components. They are removed by a spatial low-pass filter. Only the central maximum continues to the lens L_2, which forms a new, smoothed image.

Example 9.3: The two-dimensional Dirac's delta function can be defined as the product of two one-dimensional Dirac's delta functions: $\delta(x, y) = \delta(x)\delta(y)$.

(a) Show that for an arbitrary two-dimensional function $f(x, y)$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\delta(x - a, y - b)\,\mathrm{d}x\,\mathrm{d}y = f(a, b).$$

(b) Show that the Fourier transform of the function $\frac{1}{2}\,[\delta(x+a, y+b)+\delta(x-a, y-b)]$ is the plane wave $\cos(2\pi\,\boldsymbol{\rho}\cdot\boldsymbol{\nu})$, where $\boldsymbol{\rho} = (s, p)$ and $\boldsymbol{\nu} = (a, b)$. (In the transform, xy-coordinates change into sp-coordinates.)

Solution. (a) A direct calculation gives

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x-a,y-b)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)\delta(x-a)\,\mathrm{d}x\right]\delta(y-b)\,\mathrm{d}y$$
$$= \int_{-\infty}^{\infty} f(a,y)\delta(y-b)\,\mathrm{d}y = f(a,b).$$

(b) Applying the result of (a), we obtain that

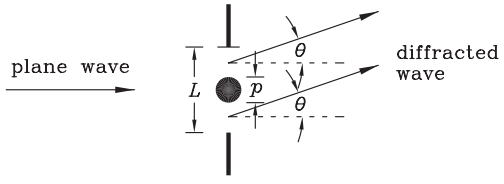
$$\begin{aligned}
&\mathcal{F}\{\frac{1}{2}\left[\delta(x+a,y+b)+\delta(x-a,y-b)\right]\}\\
=\ &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2}e^{i2\pi(sx+py)}\,\delta(x+a,y+b)\,\mathrm{d}x\mathrm{d}y\\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2}e^{i2\pi(sx+py)}\,\delta(x-a,y-b)\,\mathrm{d}x\mathrm{d}y\\
=\ &\frac{1}{2}\left[e^{-i2\pi(sa+pb)}+e^{i2\pi(sa+pb)}\right]\\
=\ &\cos\left[2\pi(sa+pb)\right]=\cos(2\pi\,\boldsymbol{\rho}\cdot\boldsymbol{\nu}).
\end{aligned}$$

Problems

1. To determine its thickness p, a thin wire is positioned in the middle of a slit of width L, in the center of the slit, as shown in the following picture. Show that if the slit is illuminated by a monochromatic plane wave of wavenumber ν, then the intensity in the direction θ behind the slit is

$$I(\theta) = I_0(L-p)^2\cos^2\left(\pi\,\frac{L+p}{2}\,\nu\sin\theta\right)\mathrm{sinc}^2\left(\pi\,\frac{L-p}{2}\,\nu\sin\theta\right).$$

Also calculate the direction at which the intensity of the interference pattern vanishes for the first time, if $p = 20.0\ \mu$m, $L = 100.0\ \mu$m and $\lambda = 632.8$ nm.



2. Deduce the Fraunhofer diffraction pattern of a grating from its transmission function written in the form

$$B(x) = \Pi_D(x) * \sum_{j=-(N-1)/2}^{(N-1)/2} \delta(x - jd),$$

where $\Pi_D(x) = \begin{cases} 1, & |x| \le D/2, \\ 0, & |x| > D/2, \end{cases}$ $\delta(x)$ is Dirac's delta function and N is the number of slits of the grating.

3. Show that convolution with the function $d(t) = \delta(t - t_0)$ shifts a function the distance t_0 "to the right", *i.e.*, $g(t) * d(t) = g(t - t_0)$. Applying this result, compute the convolution $\frac{1}{\Delta t}\, g(t) * \text{Ш}_{\Delta t}(t)$, where $\text{Ш}_{\Delta t}(t) = \sum_{k=-\infty}^{\infty} \Delta t\, \delta(t - k\Delta t)$.

4. A transmission grating consists of groups of five slits, each slit of width a. Between two slits of the group there is always an opaque area of the same width a. Every group of five slits is followed by an opaque area of width $9a$. These kinds of periods are repeated N times. Find the Fraunhofer diffraction pattern of the grating.

   Hint: The transmission function of the grating can be presented as

$$B(x) = \Pi_a(x) * \sum_{n=0}^{N-1} \sum_{m=0}^{4} \delta(x - n18a - m2a).$$

5. Show that the interference pattern I_{out} of a grating reduces to

   (a) the interference pattern $I_0 \operatorname{sinc}^2(\pi Ds) \cos^2(\pi ds)$ of two slits, if the value $N = 2$ is inserted, and

   (b) the interference pattern $I_0 \operatorname{sinc}^2(\pi Ds)$ of one slit, if the value $N = 1$ is inserted.

6. A grating, placed in the yz-plane, has a transmission function of the form

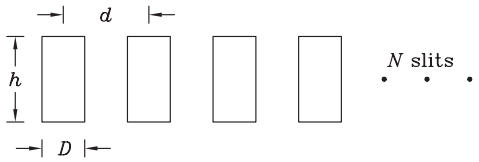
$$B(y) = \frac{1}{3}\left[2 + \cos(2\pi y/d)\right].$$

   (a) What percentage of incident light intensity passes the grating straight forward, if the incident plane wave propagates in the direction x. The width of the wave in the y-direction is assumed to be infinite.

   (b) Calculate the intensity distribution $I(\theta)$ of the diffracted wave as a function of the direction θ of the diffracted wave with respect to the x-axis, if the incident light is monochromatic (of wavelength λ), the plane wave hits the grating perpendicularly, $d = 1\ \mu$m, and the width of the wave in the y-direction is 10 mm.

7. (a) Find the Fraunhofer diffraction pattern of a diffraction obstacle consisting of N slits of width D and height h, at the distance d from each other, as shown in the following picture.

   (b) Check that insertion of $d = D$ in the result of (a) gives the diffraction pattern of one slit of width Nd and height h.



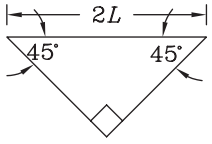
8. An opaque rectangular obstacle of width X and height Y is illuminated by a plane wave. Find the Fraunhofer diffraction pattern.

9. A square aperture of the size $2R \times 2R$ is covered by a concentric disc of radius R. Find the Fraunhofer diffraction pattern.

10. Let $G(\boldsymbol{\rho}) = \mathcal{F}\{g(\boldsymbol{r})\}$, where $\boldsymbol{\rho} = (s, p)$ and $\boldsymbol{r} = (x, y)$. Show that if $g(\boldsymbol{r})$ is a function of circular symmetry, *i.e.*, $g(\boldsymbol{r}) = g_r(r) = g_r\left(\sqrt{x^2 + y^2}\right)$, then also its two-dimensional Fourier transform is of circular symmetry, *i.e.*, $G(\boldsymbol{\rho}) = G_\rho(\rho) = G_\rho\left(\sqrt{s^2 + p^2}\right)$, and $G_\rho(\rho) = 2\pi \int_0^\infty r g_r(r) J_0(2\pi\rho r)\,\mathrm{d}r$, *i.e.*, $G_\rho(\rho)$ is the so called Hankel transform of $g_r(r)$.

    Hint: Bessel function $J_0(a) = \dfrac{1}{\pi} \int_0^\pi \cos(a \sin\theta)\,\mathrm{d}\theta$.

11. Find the Fraunhofer diffraction pattern of the triangular aperture shown in the following picture.



12. A black and white slide picture contains an annoying raster pattern, which has a period d in both x- and y-directions. The raster pattern shall be filtered away by a spatial filter. The slide is illuminated by a monochromatic plane wave of wavenumber ν, and the diffraction pattern is collected by a lens of focal length f and focused on the spatial filter. The filter is a screen which has a square aperture of the size $D \times D$. What is the largest value of D at which the raster pattern will vanish? Calculate the largest value of D, if $d = 0.20$ mm, $f = 40.0$ cm and $\lambda = 1/\nu = 633$ nm.

13. A laser beam has interference fringes which become increasingly dense the further they are from the center of the beam towards the edges. The amplitude of the beam is of the form $A\left[1+\cos(Cr^2)\right]$, where A and C are constants. The laser beam is Fourier transformed by a lens of focal length f into the plane $x'y'$. Find the intensity distribution in the $x'y'$-plane. (The angles θ_x and θ_y are small.)

    Hint: From mathematical tables, $\int_0^\infty J_0(\beta x)\cos(\alpha x^2)x\,\mathrm{d}x = \frac{1}{2\alpha}\sin\left(\frac{\beta^2}{4\alpha}\right)$. See also Problem 10 for Fourier transforming circularly symmetric functions.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

10 Uncertainty principle

10.1 Equivalent width

The *equivalent width* Δt of a real function $h(t)$ is defined as

$$\Delta t = \frac{\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t}{h(0)}, \tag{10.1}$$

where the coordinates have been chosen so that the maximum value of the function is obtained at $t = 0$: $h_{\max} = h(0)$. The equivalent width of a function is illustrated in Figure 10.1.

Let $H(f)$ be the inverse Fourier transform of $h(t)$:

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{i2\pi ft}\,\mathrm{d}f = \mathcal{F}\{H(f)\}, \tag{10.2}$$

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,\mathrm{d}t = \mathcal{F}^{-1}\{h(t)\}. \tag{10.3}$$

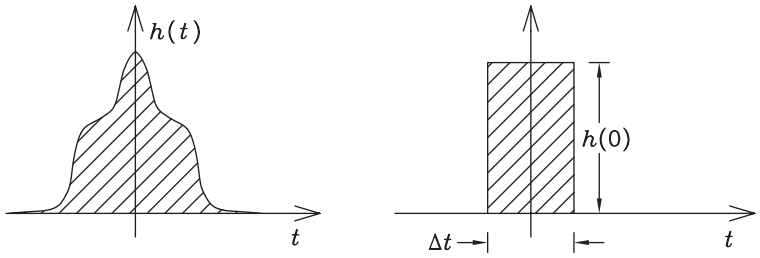


Figure 10.1: The equivalent width $\Delta t = \int_{-\infty}^{\infty} h(t)\,\mathrm{d}t \big/ h(0)$ of a function $h(t)$. The area of the box of width Δt and height $h(0) = h_{\max}$ is equal to the area under the function $h(t)$.

Clearly, the value of $h(t)$ at $t = 0$ is

$$h(0) = \int_{-\infty}^{\infty} H(f)\,\mathrm{d}f, \tag{10.4}$$

and the value of $H(f)$ at $f = 0$ is

$$H(0) = \int_{-\infty}^{\infty} h(t)\,\mathrm{d}t. \tag{10.5}$$

The equivalent width Δf of $H(f)$ is

$$\Delta f = \frac{\int_{-\infty}^{\infty} H(f)\,\mathrm{d}f}{H(0)} = \frac{h(0)}{\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t} = \frac{1}{\Delta t} = \frac{h(0)}{H(0)}. \tag{10.6}$$

Consequently, the distributions of a Fourier transform pair in the frequency domain and the time domain obey

$$\boxed{\Delta f\,\Delta t = 1.} \tag{10.7}$$

Let us consider plane waves $e^{-ikx+i\omega t}$, whose angular frequency is ω, and wavenumber is $k = 2\pi/\lambda = 2\pi\nu$, where λ is the wavelength. A wave packet at time $t = 0$ can be expressed as

$$\psi(x) = \int_{-\infty}^{\infty} A(k)e^{ikx}\,\mathrm{d}k = \int_{-\infty}^{\infty} A(\nu)e^{i2\pi\nu x}\,\mathrm{d}\nu = \mathcal{F}\{A(\nu)\}, \tag{10.8}$$

where the amplitude $A(\nu) = 2\pi A(k)$. We can see that $\psi(x)$ is the Fourier transform of $A(\nu)$. Consequently, we can apply Equation 10.7, which is valid for any Fourier transform pair $h(t)$ and $H(f)$. By substituting t by x and f by ν we obtain

$$\boxed{\Delta x\,\Delta\nu = 1.} \tag{10.9}$$

This is the *uncertainty principle* of a wave packet. It is impossible to make Δx and $\Delta\nu$ very small at the same time. A wave packet which obeys these equations may be, for example, an electromagnetic wave packet, and the uncertainty principle is valid, for example, for a photon.

In quantum mechanics a particle is described by a wave whose wavelength is *de Broglie wavelength*

$$\boxed{\lambda = \frac{h}{p}.} \tag{10.10}$$

In this equation, h is Planck's constant, and p is the momentum of the particle $p = mv$, where m is mass and v is velocity. In practice, a wave cannot be infinitely long, which means that the wavelength of the particle is spread in the close vicinity of de Broglie wavelength. Let us choose the coordinates in such a way that the particle moves in the x-direction. The particle is represented by the wave packet

$$\psi(x) = \int_{-\infty}^{\infty} A(\nu) e^{i2\pi\nu x}\,\mathrm{d}\nu = \int_{-\infty}^{\infty} \phi(p) e^{i2\pi px/h}\,\mathrm{d}p, \tag{10.11}$$

where we have applied Equation 10.10, which gives

$$\nu = 1/\lambda = p/h. \tag{10.12}$$

The amplitude

$$\phi(p) = \frac{A\left(\frac{p}{h}\right)}{h}. \tag{10.13}$$

We can now rewrite the uncertainty principle of Equation 10.9 as

$$\Delta x\,\Delta\nu = \Delta x\,\frac{\Delta p}{h} = 1, \tag{10.14}$$

or

$$\boxed{\Delta x\,\Delta p = h.} \tag{10.15}$$

This means that if a wave packet is highly localized, then it is impossible to associate it with a well-defined momentum. On the other hand, a wave packet whose momentum is well defined must be spatially broad.

The energy of a particle is $E = hf$. If we substitute $\Delta E/h$ for Δf in Equation 10.7, we obtain the uncertainty principle of energy and time,

$$\boxed{\Delta E\,\Delta t = h.} \tag{10.16}$$

In quantum mechanics, the wave packet $\psi(x)$, which describes a particle, is given a *probability interpretation*: if $\psi(x)$ is normalized, then

$$\begin{aligned} |\psi(x)|^2\,\mathrm{d}x &= \psi^*(x)\psi(x)\,\mathrm{d}x \\ &= \text{probability to find the particle in the interval } (x, x+\mathrm{d}x). \end{aligned} \tag{10.17}$$

In other words, the intensity $|\psi(x)|^2$ of the wave is the *probability density*.

If an operator $\hat{\alpha}$ corresponds to a physically measurable quantity, the expectation value of such a measurement, given the state $\psi(x)$, is

$$\langle\hat{\alpha}\rangle = \frac{\displaystyle\int_{-\infty}^{\infty} \psi^*(x)\hat{\alpha}\psi(x)\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} |\psi(x)|^2\,\mathrm{d}x}. \tag{10.18}$$

If the wave packet ψ is the *eigenfunction* of the operator $\hat{\alpha}$, that is, it satisfies the equation

$$\hat{\alpha}\psi(x) = a\psi(x), \tag{10.19}$$

then the average value of the operator in the state ψ is $\langle\hat{\alpha}\rangle = a$.

Since quantum mechanics uses the probability interpretation, it is customary to express the width of a quantity as the statistical standard deviation σ (see Section 10.3), instead of using the equivalent width of the function.

In the following sections, we shall examine how statistical quantities are expressed with Fourier transform, and what the uncertainty principle looks like, if statistical standard deviations are used instead of equivalent widths.

Example 10.1: Determine the equivalent width of the Lorentz curve $H(f) = \dfrac{C}{\sigma^2 + f^2}$.

Solution. The Lorentz curve can be expressed as

$$H(f) = \frac{C}{\sigma^2 + f^2} = \frac{C\pi}{\sigma}\,\frac{\sigma/\pi}{\sigma^2 + f^2}.$$

Its Fourier transform is

$$h(t) = \mathcal{F}\{H(f)\} = \frac{C\pi}{\sigma}\,e^{-2\pi\sigma|t|}.$$

The equivalent width is

$$\Delta f = \frac{\int_{-\infty}^{\infty} H(f)\,\mathrm{d}f}{H(0)} = \frac{h(0)}{H(0)} = \frac{C\pi/\sigma}{C/\sigma^2} = \pi\sigma.$$

10.2 Moments of a function

Let us, once again, assume that $h(t)$ and $H(f)$ are a Fourier transform pair. The derivative theorem of Fourier transforms states that

$$H^{(k)}(f) = \frac{\mathrm{d}^k H(f)}{\mathrm{d}f^k} = \int_{-\infty}^{\infty} (-i2\pi t)^k h(t) e^{-i2\pi f t}\,\mathrm{d}t. \tag{10.20}$$

Applying this, we can express the *moment of order k* of a function $h(t)$ as

$$\boxed{\int_{-\infty}^{\infty} t^k h(t)\,\mathrm{d}t = \frac{H^{(k)}(0)}{(-i2\pi)^k}\,.} \tag{10.21}$$

The zeroth moment of a function is the same as the area of the function:

$$\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t = H(0). \tag{10.22}$$

This means that the area of a function equals the value of its inverse Fourier transform at $f = 0$. This is illustrated in Figure 10.2.

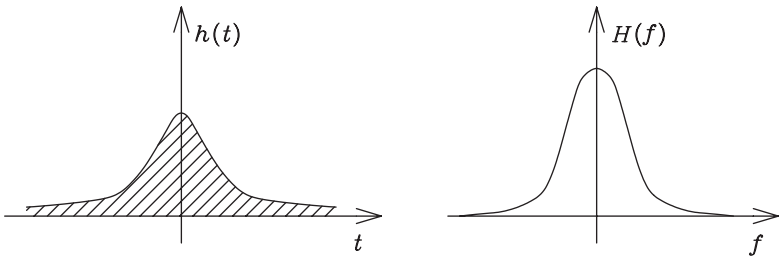


Figure 10.2: The area of a function $h(t)$, which is shaded in the figure, is equal to $H(0)$, which is the value of $H(f)$ at $f = 0$.

The first moment of a function $h(t)$ can be expressed with the help of the first derivative $H^{(1)}(f) = \mathrm{d}H(f)/\mathrm{d}f$ of the corresponding spectrum:

$$\int_{-\infty}^{\infty} t h(t)\,\mathrm{d}t = \frac{H^{(1)}(0)}{-i2\pi}. \tag{10.23}$$

The *center of mass* of a function is the weighted average, or the weighted mean value, of the function. It can be written as

$$\boxed{\langle t\rangle_h = \frac{\displaystyle\int_{-\infty}^{\infty} t h(t)\,\mathrm{d}t}{\displaystyle\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t} = \frac{H^{(1)}(0)}{-i2\pi H(0)}.} \tag{10.24}$$

The weighted average can be obtained from the zeroth and the first moments as

$$\langle t\rangle = \frac{\text{1. moment}}{\text{0. moment}}. \tag{10.25}$$

Equation 10.24 is, actually, a special case of Equation 10.18, which gives the mean value of operators in quantum mechanics.

10.3 Second moment

The second moment of a function $h(t)$ can be obtained from the second derivative $H^{(2)}(f) = \mathrm{d}^2 H(f)/\mathrm{d}f^2$ of the corresponding spectrum:

$$\boxed{\int_{-\infty}^{\infty} t^2 h(t)\,\mathrm{d}t = -\frac{H^{(2)}(0)}{4\pi^2}.} \tag{10.26}$$

The second moment is, physically, the *moment of inertia*.

The *mean square* of a function is the ratio of the second moment and the zeroth moment:

$$\boxed{\left\langle t^2 \right\rangle = \frac{\displaystyle\int_{-\infty}^{\infty} t^2 h(t)\,\mathrm{d}t}{\displaystyle\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t} = -\frac{H^{(2)}(0)}{4\pi^2 H(0)}.} \tag{10.27}$$

The square root of the mean square of a function, $\sqrt{\langle t^2 \rangle}$, is the *root mean square* or the *rms value* of the function.

The *variance* or the *mean square deviation* of a function $h(t)$ is:

$$\sigma^2 = \left\langle (t - \langle t \rangle)^2 \right\rangle = \frac{\displaystyle\int_{-\infty}^{\infty} (t - \langle t \rangle)^2 h(t)\,\mathrm{d}t}{\displaystyle\int_{-\infty}^{\infty} h(t)\,\mathrm{d}t}. \tag{10.28}$$

On the other hand, the mean square deviation may be written as

$$\sigma^2 = \left\langle t^2 \right\rangle - 2\langle t \rangle \langle t \rangle + \langle t \rangle^2 = \left\langle t^2 \right\rangle - \langle t \rangle^2, \tag{10.29}$$

which is obtained by

$$\boxed{\sigma^2 = -\frac{H^{(2)}(0)}{4\pi^2 H(0)} + \frac{1}{4\pi^2}\left[\frac{H^{(1)}(0)}{H(0)}\right]^2.} \tag{10.30}$$

The square root σ of the mean square deviation is called the *standard deviation*.

Let us next examine the uncertainty principle again, but use standard deviations instead of equivalent widths.

The derivative theorem states that any Fourier transform pair $h(t)$ and $H(f)$ obeys

$$\mathcal{F}^{-1}\left\{h^{(1)}(t)\right\} = \mathcal{F}^{-1}\left\{\frac{\mathrm{d}h(t)}{\mathrm{d}t}\right\} = i2\pi f H(f). \tag{10.31}$$

Applying this, and the power theorem, we obtain that

$$\int_{-\infty}^{\infty} h^{(1)}(t)[h^{(1)}(t)]^* \,\mathrm{d}t = 4\pi^2 \int_{-\infty}^{\infty} f^2 H(f) H^*(f) \,\mathrm{d}f. \tag{10.32}$$

The *Cauchy–Schwarz inequality*, well-known in mathematics, can be generalized to complex functions:

$$4 \int_{-\infty}^{\infty} hh^* \,\mathrm{d}t \int_{-\infty}^{\infty} gg^* \,\mathrm{d}t \geq \left| \int_{-\infty}^{\infty} (h^* g + hg^*) \,\mathrm{d}t \right|^2. \tag{10.33}$$

Integrating by parts, it is possible to show that

$$\left| \int_{-\infty}^{\infty} t h^{(1)} \,\mathrm{d}t \right| = \left| \int_{-\infty}^{\infty} h \,\mathrm{d}t \right|. \tag{10.34}$$

Applying these three results, let us examine the product of the mean square deviations σ_h^2 of $h(x)$ and σ_H^2 of $H(\nu)$. The quantum mechanical probability densities in the x- and ν-domains can be written as $h(x) = \psi(x)^* \psi(x)$ and $H(\nu) = A(\nu)^* A(\nu)$, respectively. We shall denote $\mathrm{d}\psi(x)/\mathrm{d}x = \psi^{(1)}$. We shall assume that the averages $\langle x \rangle = 0$ and $\langle \nu \rangle = 0$. We can show the following inequality:

$$\sigma_h^2 \sigma_H^2 = \frac{\displaystyle\int_{-\infty}^{\infty} x^2 \psi\psi^* \,\mathrm{d}x \int_{-\infty}^{\infty} \nu^2 AA^* \,\mathrm{d}\nu}{\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x \int_{-\infty}^{\infty} AA^* \,\mathrm{d}\nu} \overset{(10.32)}{=} \frac{\displaystyle\int_{-\infty}^{\infty} x\psi x\psi^* \,\mathrm{d}x \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \psi^{(1)}\psi^{(1)*} \,\mathrm{d}x}{\left[\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x\right]^2}$$

$$\overset{(10.33)}{\geq} \frac{\left|\displaystyle\int_{-\infty}^{\infty} (x\psi^*\psi^{(1)} + x\psi\psi^{(1)*}) \,\mathrm{d}x\right|^2}{16\pi^2 \left[\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x\right]^2} = \frac{\left|\displaystyle\int_{-\infty}^{\infty} x \frac{\mathrm{d}}{\mathrm{d}x}(\psi\psi^*) \,\mathrm{d}t\right|^2}{16\pi^2 \left[\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x\right]^2}$$

$$\overset{(10.34)}{=} \frac{\left|\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x\right|^2}{16\pi^2 \left[\displaystyle\int_{-\infty}^{\infty} \psi\psi^* \,\mathrm{d}x\right]^2} = \frac{1}{16\pi^2}. \tag{10.35}$$

This means that if we use the statistical standard deviations to express the widths of the functions, that is, $\Delta x = \sigma_h$ and $\Delta \nu = \sigma_H$, then the uncertainty principle of Equation 10.9 achieves the form

$$\boxed{\Delta x \, \Delta \nu \geq \frac{1}{4\pi}.} \tag{10.36}$$

Consequently, if statistical standard deviations are used, Equation 10.15, which is the uncertainty principle of the momentum and the place of a particle, is written as

$$\boxed{\Delta x \, \Delta p \geq \frac{h}{4\pi}.} \tag{10.37}$$

The exact equivalence is obtained, if the function $\psi(x)$ and its Fourier transform $A(\nu)$ are of the same functional shape (for an example, see Problem 6).

Example 10.2: The probability density function of a random variable is

$$f(x) = Ae^{-\alpha|x-x_0|},$$

where A, α and x_0 are constants (A is the normalization constant). Applying Fourier transforms, determine the second moment or the mean value $\langle x^2 \rangle$.

Solution. We can rearrange the exponent of the function:

$$f(x) = Ae^{-\alpha|x-x_0|} = Ae^{-2\pi \frac{\alpha}{2\pi}|x-x_0|}.$$

Applying the shift theorem and Table 1.1, we can find the inverse Fourier transform of $f(x)$:

$$F(\nu) = A \, \frac{\frac{\alpha}{2\pi^2}}{(\frac{\alpha}{2\pi})^2 + \nu^2} \, e^{-i2\pi x_0 \nu} = 2\alpha A \, \frac{e^{-i2\pi x_0 \nu}}{\alpha^2 + 4\pi^2 \nu^2}.$$

Consequently,

$$F(0) = \frac{2A}{\alpha}.$$

The first derivative of $F(\nu)$ is

$$\begin{aligned} F^{(1)}(\nu) &= 2\alpha A \, \frac{-i2\pi x_0(\alpha^2 + 4\pi^2\nu^2) - 8\pi^2\nu}{(\alpha^2 + 4\pi^2\nu^2)^2} \, e^{-i2\pi x_0 \nu} \\ &= \frac{-i4\alpha A \pi x_0}{\alpha^2 + 4\pi^2\nu^2} \, e^{-i2\pi x_0 \nu} - \frac{16\alpha A \pi^2 \nu}{(\alpha^2 + 4\pi^2\nu^2)^2} \, e^{-i2\pi x_0 \nu}. \end{aligned}$$

The second derivative of $F(\nu)$ is

$$\begin{aligned} F^{(2)}(\nu) &= i4\alpha A \pi x_0 \, \frac{i2\pi x_0(\alpha^2 + 4\pi^2\nu^2) + 8\pi^2\nu}{(\alpha^2 + 4\pi^2\nu^2)^2} \, e^{-i2\pi x_0 \nu} \\ &- 16\alpha A \pi^2 \, \frac{(1 - i2\pi x_0 \nu)(\alpha^2 + 4\pi^2\nu^2)^2 - 2(\alpha^2 + 4\pi^2\nu^2)8\pi^2\nu^2}{(\alpha^2 + 4\pi^2\nu^2)^4} \, e^{-i2\pi x_0 \nu}. \end{aligned}$$

Consequently,

$$\begin{aligned} F^{(2)}(0) &= i4\alpha A\pi x_0 \frac{i2\pi x_0\alpha^2}{\alpha^4} - 16\alpha A\pi^2 \frac{\alpha^4}{\alpha^8} \\ &= \frac{-8A\pi^2 x_0^2}{\alpha} - \frac{16A\pi^2}{\alpha^3}. \end{aligned}$$

Now we can calculate the requested mean value, which is

$$\langle x^2 \rangle = \frac{-F^{(2)}(0)}{4\pi^2 F(0)} = \frac{-\alpha}{8A\pi^2} F^{(2)}(0) = x_0^2 + \frac{2}{\alpha^2}.$$

Problems

1. What is the equivalent width of the function $\mathrm{sinc}^2(Cx)$?

2. Show that $W_{f*g} = \dfrac{W_f W_g}{W_{fg}}$, where we denote W_h the equivalent width of a function $h(x)$. $f(x)$ and $g(x)$ are real and even functions.

3. Determine the kth derivative of the function $\mathrm{sinc}(Cx)$ at the point $x = 0$. k is an arbitrary, non-negative integer.

4. The probability density of a random variable is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

   where x may have a value from $-\infty$ to $+\infty$. Applying Fourier transforms, compute the mean value and the standard deviation of the variable.
   Hint: $\mathcal{F}\left\{\sqrt{\alpha/\pi}\, e^{-\alpha x^2}\right\} = \mathcal{F}^{-1}\left\{\sqrt{\alpha/\pi}\, e^{-\alpha x^2}\right\} = e^{-\frac{\pi^2\nu^2}{\alpha}}$. Also apply the shift theorem.

5. f and g are probability density functions. σ^2 is variance.
   (a) Show that $\langle x \rangle_{f*g} = \langle x \rangle_f + \langle x \rangle_g$.
   (b) Show that $\sigma^2_{f*g} = \sigma^2_f + \sigma^2_g$.

6. The wave functions of a particle in the x-domain and in the ν-domain constitute a Fourier transform pair. The wave function (or the probability amplitude) $\psi(x)$ in the x-domain is of Gaussian shape. Show that if the uncertainties Δx and $\Delta\nu$ are expressed as the standard deviations of x and ν, respectively, then the optimal uncertainty relation $\Delta x\, \Delta p = \dfrac{h}{4\pi}$ is obtained. The coordinates can be chosen in such a way that the mean value $\langle x \rangle = 0$.

7. Determine the standard deviation in the cases where the probability density is, neglecting the normalization constant,

   (a) $f(x) = e^{-|x|}$,

   (b) $f(x) = e^{-\pi^2 a^2 x^2}$.

8. Applying the formula of the kth moment of a function, determine the value of the integral $\int_0^\infty x e^{-px} \sin(qx)\,\mathrm{d}x$, where p and q are real constants, and $p > 0$.

   Hint: Use also the shift theorem.

9. The wave function of a particle in the (one-dimensional) p-domain is

   $$\phi(p) = A\,\mathrm{sinc}^2\left[\pi(p - p_0)/B\right],$$

   where A, B and p_0 are constants. Compute the probability that the particle is found in the interval $\left[-\frac{h}{2B}, \frac{h}{2B}\right]$.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

11 Processing of signal and spectrum

11.1 Interpolation

The process of finding estimates for the values of a function at new points between known data by computational methods is called *interpolation*. Let us assume that we have measured the data H_k of a spectrum $H(f)$. In interpolation, we try to find estimates for the values of the spectrum $H(f)$ at new points between the known data H_k. We can, of course, apply interpolation to other functions than spectra as well.

A fast and simple interpolation method is obtained by the use of Fourier transforms. In what follows, we consider only the positive half of the signal with $t > 0$, and assume that symmetric operations are always performed on the negative half. Note that when using FFT, the negative half is aliased behind the positive half, so that adding new data after the known ones actually means adding in the middle of the data sequence (see Problem 1). Firstly, the Fourier transform of the spectrum is computed from the known values of the spectrum:

$$\mathcal{F}\{H_k\} = h_j, \qquad j = 0, 1, 2, \ldots, N-1.$$

This can be done by FFT. Secondly, this computed sequence of data h_j, consisting of N data in t-domain, is extended by adding a desired number M of new data behind the old N data. The new M data are all given the value zero. This is called zero-padding. The new data sequence h'_j consists of $N + M$ data, in such a way that $h'_j = h_j$, if $0 \leq j \leq N-1$, and $h'_j = 0$, if $N - 1 < j \leq N + M - 1$. Thirdly, inverse Fourier transform is computed from the new, enlargened data sequence:

$$\mathcal{F}^{-1}\{h'_j\} = H'_k, \qquad k = 0, 1, 2, \ldots, M + N - 1.$$

In this way, $M + N$ data of the spectrum are obtained, instead of the original N.

If we wish to use FFT in the calculation of the Fourier transforms, then we must choose $N = 2^{m_1}$ and $M + N = 2^{m_2}$, where m_1 and m_2 are integers. Figure 11.1 shows an example of a spectrum $H(f)$ of which $N = 2^m$ data H_k have been measured. Figure 11.2 shows the calculated Fourier transform of the spectrum, $\mathcal{F}\{H_k\} = h_j$, as well as the new values zero, which are added behind the calculated data for interpolation. In this figure, the number of the new zero data is chosen to be equal to the number of the old calculated data, that is, $M = N$. Figure 11.3 shows the interpolated spectrum $\mathcal{F}^{-1}\{h'_j\} = H'_k$, calculated from the extended data sequence h'_j.

We can see that the original data remained unchanged in this interpolation. Interpolation does not distort the original information. The new, additional data do not, however, contain additional information of the true spectrum.

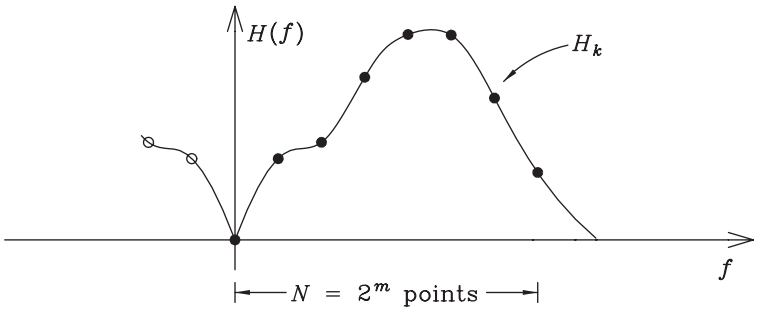


Figure 11.1: An example of a spectrum $H(f)$ of which $N = 2^m$ data H_k have been measured.

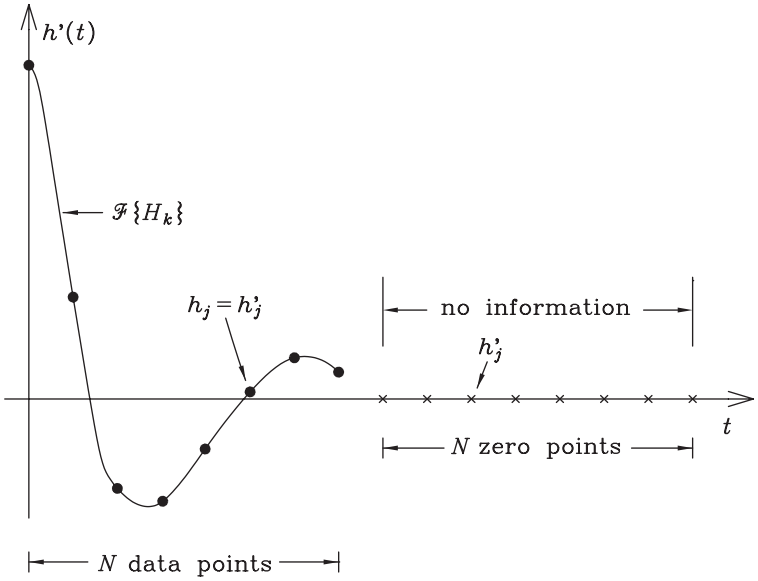


Figure 11.2: The calculated Fourier transform $\mathcal{F}\{H_k\} = h_j$ of a measured spectrum H_k with N data, and N new zero data, which are added behind the calculated data for interpolation.

The *degree of interpolation* is the higher the larger number of zero data we add behind the calculated Fourier transform data. If we wish to achieve an interpolation degree K, which is an integer ≥ 1, then we add $(K-1)N$ zero data behind the N calculated Fourier transform data h_j, and use the obtained KN data h'_j. The interpolated spectrum $\mathcal{F}^{-1}\{h'_j\} = H'_k$ consists of the original spectral data H_k and new $K-1$ interpolated data between each two original data at constant intervals, as shown in Figure 11.4.

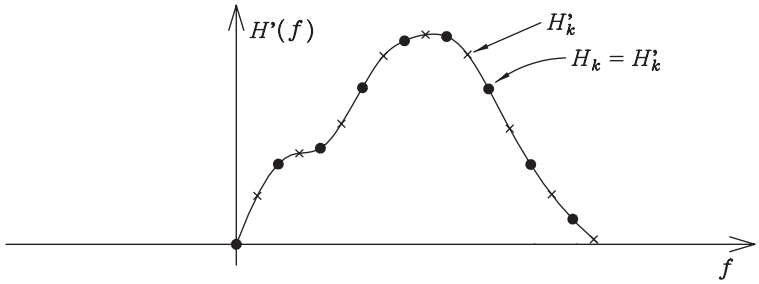


Figure 11.3: Interpolated spectrum H'_k, which has been calculated by adding to the Fourier transform $\mathcal{F}\{H_k\}$ of a measured spectrum H_k as many new zero data as the measured spectrum has measured data. • is an original, measured datum, and × is a new, interpolated datum.

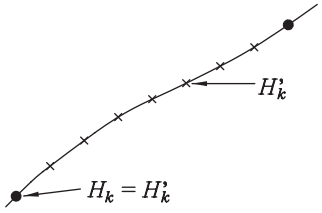


Figure 11.4: A piece of an interpolated spectrum with the degree of interpolation $K = 8$. • is an original, measured datum, and × is a new, interpolated datum.

Generally, an interpolation $H'(f)$ of a function $H(f)$ can be expressed in the f-domain as a convolution integral

$$H'(f) = W(f) * H(f) = \int_{-\infty}^{\infty} W(u)H(f-u)\,\mathrm{d}u = \int_{-\infty}^{\infty} H(u)W(f-u)\,\mathrm{d}u. \quad (11.1)$$

$W(u)$ tells the weight of the values $H(f-u)$ in calculation of $H'(f)$. In discrete form, an interpolated spectral datum $H'_{k+\lambda}$ between two successive original spectral data H_k and H_{k+1} separated by a distance Δf can be written as

$$H'_{k+\lambda} = \sum_{j=-\infty}^{\infty} H_j W\left[(k+\lambda)\Delta f - j\Delta f\right]\Delta f, \qquad 0 \le \lambda \le 1, \qquad (k+\lambda)\Delta f = f. \quad (11.2)$$

The interpolation of a spectrum $H(f)$ by convolution with a function $W(f)$ is illustrated in Figure 11.5.

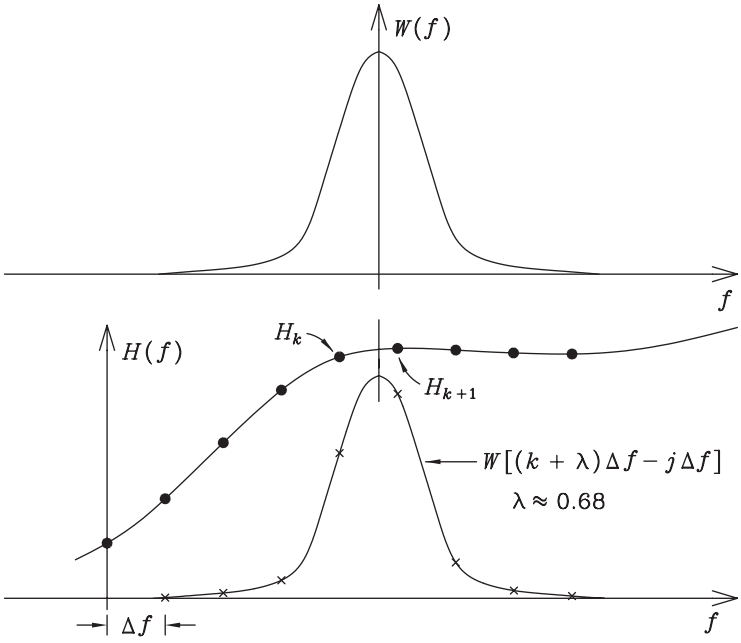


Figure 11.5: Interpolation of spectrum $H(f)$ by convolution with $W(f)$. The known data H_k are separated by the distance Δf. Values of interpolated data are given by the sum $\sum_{j=-\infty}^{\infty} H_j W\,[(k+\lambda)\Delta f - j\Delta f]\,\Delta f$.

The interpolation method where zero data are added to the Fourier transform of the spectrum, in t-domain, can also be expressed as a convolution in f-domain. We can consider that in this method the Fourier transform in t-domain is multiplied by the boxcar function

$$\Pi_{2N\Delta t}(t) = \begin{cases} 0, & |t| > N\Delta t, \\ 1, & |t| \le N\Delta t. \end{cases} \tag{11.3}$$

The interval in the t-domain is

$$\Delta t = \frac{1}{2N\Delta f},$$

where N is the number of original data (on the positive half of the spectrum), and Δf the sampling interval of the original spectrum. Multiplication in t-domain corresponds to convolution in f-domain with

$$W(f) = \mathcal{F}^{-1}\{\Pi_{2N\Delta t}(t)\}. \tag{11.4}$$

Knowing the Fourier transform of a boxcar function, we can write

$$W(f) = 2N\Delta t \,\mathrm{sinc}(2\pi f N \Delta t), \tag{11.5}$$

or,

$$\boxed{W(f) = \frac{1}{\Delta f}\,\mathrm{sinc}\left(\frac{\pi f}{\Delta f}\right).} \tag{11.6}$$

The interpolated spectrum is the convolution

$$\begin{aligned} H'(f) &= \int_{-\infty}^{\infty} H(u)W(f-u)\,\mathrm{d}u \approx \Delta f \sum_{j=-\infty}^{\infty} H_j W(f - j\Delta f) \\ &= \sum_{j=-\infty}^{\infty} H_j \,\mathrm{sinc}\left[\frac{\pi}{\Delta f}(f - j\Delta f)\right]. \end{aligned}$$

At the point $f = (k+\lambda)\Delta f$ the interpolated spectral value is

$$\boxed{H'_{k+\lambda} \approx \sum_{j=-\infty}^{\infty} H_j \,\mathrm{sinc}[\pi(k+\lambda-j)].} \tag{11.7}$$

The interpolated spectral value which is obtained by adding zero data to the Fourier transform in t-domain is equal to the value which is obtained by convolution of the original spectral data with the sinc function of Equation 11.5. We can also notice that if $\lambda = 0$, then Equation 11.7 gives $H'_k = H_k$. In practice, we cannot use Equation 11.7 as such, because the sum has an infinite number of terms.

Interpolation by convolution with the sinc function of Equation 11.5 is illustrated in Figure 11.6. If $\lambda = 0$, then $H'_k = H_k$, because $\mathrm{sinc}(\pi l) = 0$, $l = \pm 1, \pm 2, \pm 3, \ldots$ If $\lambda \neq 0$, then, generally, all the measured data H_j contribute to the interpolated spectral value $H'_{k+\lambda}$, because $\mathrm{sinc}\,[\pi(k+\lambda-j)] \neq 0$ at any values j and k.

Finally note that zero-padding the signal is only necessary when using FFT, since it gives the spectrum only at the predetermined points $k\Delta f = k/(2N\Delta t)$. If we compute discrete Fourier transform by other means, this restriction no longer exists, and once we have obtained the signal data h_j we may calculate the spectrum at any point f we wish from Equation 3.3.

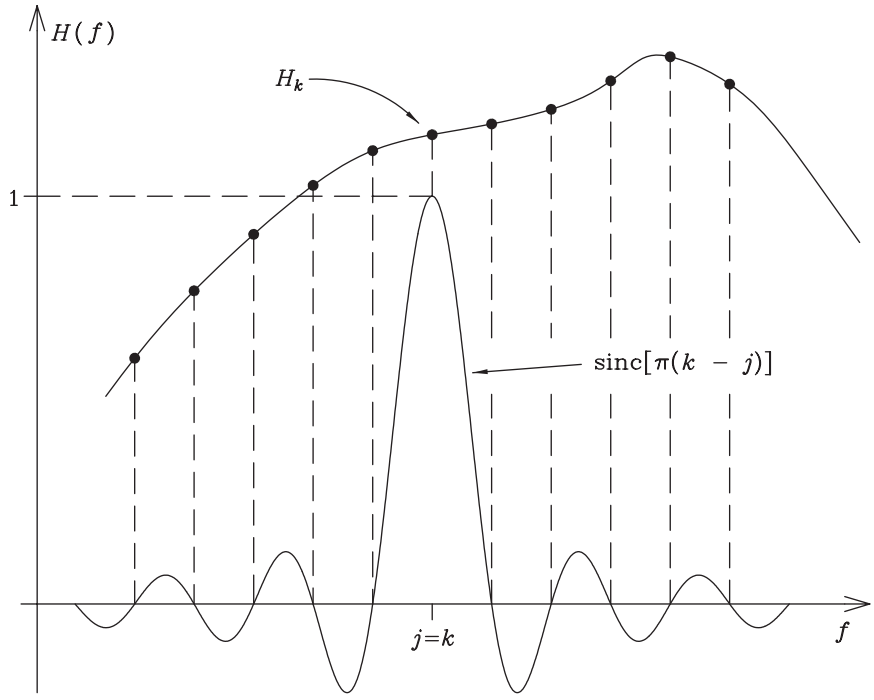


Figure 11.6: Interpolation $H'_{k+\lambda} \approx \sum_{j=-\infty}^{\infty} H_j \operatorname{sinc}[\pi(k+\lambda-j)]$ of the spectrum $H(f)$. If $\lambda = 0$, then $H'_k = H_k$, because $\operatorname{sinc}(\pi l) = 0$ ($l = \pm 1, \pm 2, \pm 3, \ldots$).

11.2 Mathematical filtering

A measured spectrum generally contains undesired components, noise, in addition to the actual spectrum. The measured spectrum is $H_0(f) + N(f)$, where $H_0(f)$ is the actual spectrum and $N(f)$ is noise.

Band limited white noise consists of noise at all frequencies in a limited band, below a maximum frequency f_{noise}. If the maximum frequency of the actual signal is $f_{\max}$, we have

$$\begin{cases} H_0(f) = 0, & |f| \geq f_{\max}, \\ N(f) = 0, & |f| \geq f_{\text{noise}}. \end{cases} \tag{11.8}$$

Figure 11.7 shows an example of a spectrum which contains band limited white noise. In this example, the maximum frequency of the noise $f_{\text{noise}} = 3 f_{\max}$.

The registered signal which gives the noisy spectrum is $h_0(t) + n(t)$, where $h_0(t)$ is the actual signal and $n(t)$ the noise. Figure 11.8 shows such a signal.

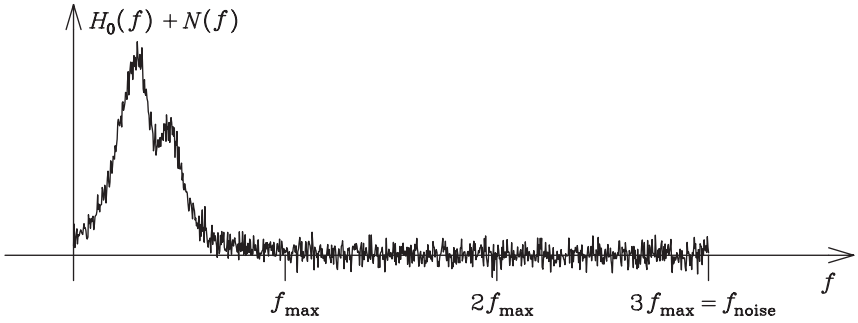


Figure 11.7: A spectrum $H_0(f) + N(f)$, consisting of the actual spectrum $H_0(f)$ and noise $N(f)$. $H_0(f) = 0$ at $|f| \geq f_{\max}$, and the band limited white noise $N(f) = 0$ at $|f| \geq f_{\text{noise}}$.

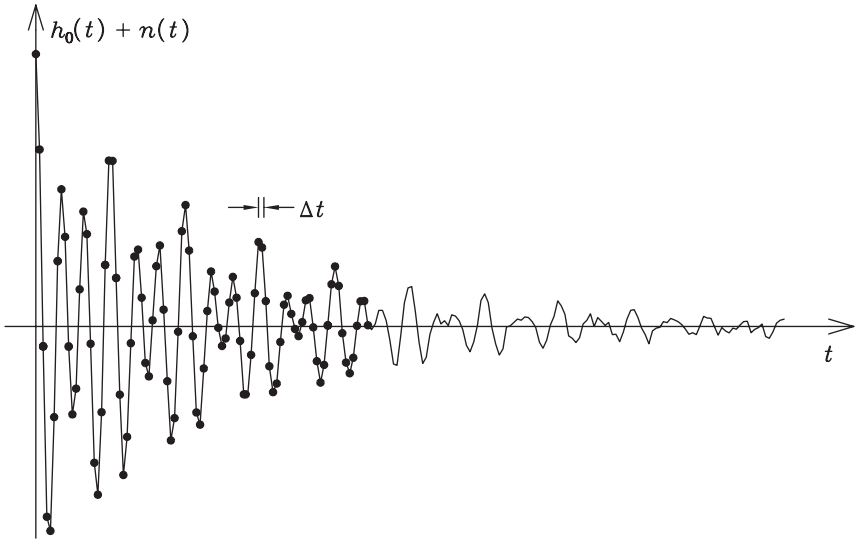


Figure 11.8: A registered signal $h_0(t)+n(t)$, where $h_0(t)$ is the actual signal and $n(t)$ is the contribution of noise in the signal. Δt is the sampling interval.

If we know that the maximum frequency of the spectrum is $f_{\max}$, then a natural choice for the sampling interval would be $\Delta t = 1/(2f_{\max})$, because then the spectrum would not be aliased. However, if $f_{\text{noise}} > f_{\max}$, then noise will be aliased, and the amplitude of noise in the spectrum will grow. In the case of Figure 11.7, where $f_{\text{noise}} = 3f_{\max}$, the amplitude of noise in the spectrum would grow $\sqrt{3}$-fold due to aliasing. Generally, if $f_{\text{noise}} = nf_{\max}$, then the sampling interval $\Delta t = 1/(2f_{\max})$ increases the amplitude of the noise $\sqrt{n}$-fold.

The aliasing of noise can be avoided by using the shorter sampling interval

$$\Delta t = \frac{1}{2f_{\text{noise}}}, \tag{11.9}$$

but then the number of data increases. In the example of Figure 11.7, we would have $\Delta t = 1/(6f_{\max})$, and the number of data would be increased 3-fold.

An alternative solution for the elimination of high-frequency noise is low-pass filtering. Low-pass filtering is a special case of *band-pass filtering,* which filters away all other frequencies of the spectrum than the desired bands containing information.

Let us examine a spectrum $H(f) = H_0(f) + N(f)$, which consists of an actual spectrum $H_0(f)$ limited to a band $f_a < f < f_b$, and of band limited white noise $N(f)$ at frequencies $f < f_{\text{noise}}$. Figure 11.9 shows an example of such a spectrum. The filtered spectrum of $H(f)$ is

$$H'(f) = G(f)H(f), \tag{11.10}$$

where the *filter function*, or the *transfer function*, $G(f)$ is

$$G(f) = \begin{cases} 1, & f_a \le |f| \le f_b, \\ 0, & |f| < f_a, \text{ or } |f| > f_b, \end{cases} \tag{11.11}$$

which is shown in Figure 11.10.

Taking the Fourier transforms of both sides of Equation 11.10 and applying the convolution theorem we obtain

$$\mathcal{F}\{H'(f)\} = \mathcal{F}\{G(f)\} * \mathcal{F}\{H(f)\}, \tag{11.12}$$

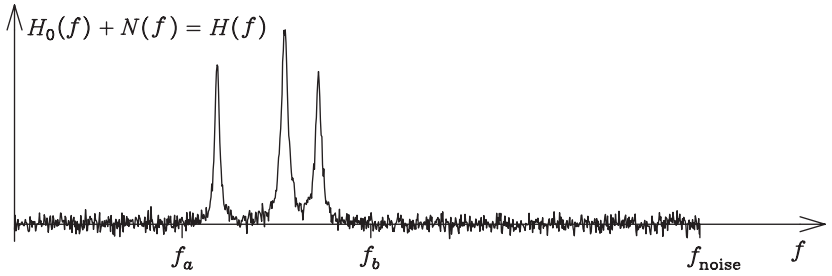


Figure 11.9: A spectrum $H(f) = H_0(f) + N(f)$, which consists of an actual spectrum $H_0(f)$ limited to a band $f_a < f < f_b$, and of band limited white noise $N(f)$ at frequencies $f < f_{\text{noise}}$. The mirror image $H(-f) = H(f)$ is not shown in the figure.

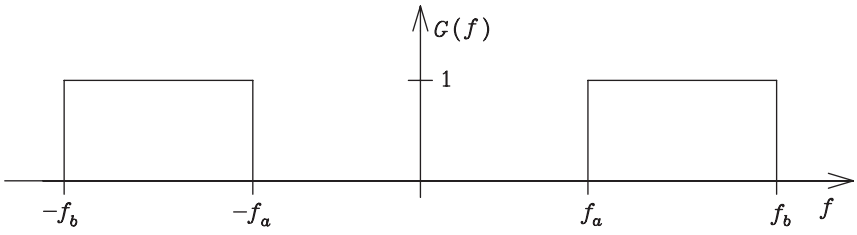


Figure 11.10: The transfer function $G(f)$ for band-pass filtering of the spectrum shown in Figure 11.9.

or

$$h'(t) = g(t) * h(t), \tag{11.13}$$

where $h(t) = h_0(t) + n(t)$ is the signal which corresponds to the unfiltered spectrum, $h'(t)$ is the signal which corresponds to the filtered spectrum, and $g(t) = \mathcal{F}\{G(f)\}$ is the *smoothing function,* or the *impulse response*. $h'(t)$ is called a *smoothed* signal. We can compute $g(t)$ using the shift theorem or the modulation theorem, and obtain

$$\boxed{g(t) = 2(f_b - f_a)\,\mathrm{sinc}\,[\pi(f_b - f_a)t]\cos[\pi(f_b + f_a)t]\ .} \tag{11.14}$$

There are two alternative procedures how a noisy signal, $h(t)$, can be smoothed after it has been registered, in order to obtain a signal which corresponds to a filtered spectrum.

Procedure 1: Take the inverse Fourier transform of the signal $h(t)$ which contains noise. Multiply the inverse Fourier transform by the transfer function $G(f)$. Take the Fourier transform of the product. The Fourier transform of the product is the smoothed signal $h'(t)$:

$$h'(t) = \mathcal{F}\{\underbrace{G(f)\underbrace{\mathcal{F}^{-1}\{h(t)\}}_{H(f)}}_{H'(f)}\}. \tag{11.15}$$

If this procedure is used, the filter is as sharp as possible. In this procedure, however, aliasing of noise $n(t)$ is problematic. In order to avoid aliasing of noise, we would have to use a long signal portion and a large number of data.

Procedure 2: Use the Equation 11.13 and calculate the convolution

$$h'(t) = g(t) * h(t) = \int_{-\infty}^{\infty} g(u)h(t-u)\,\mathrm{d}u. \tag{11.16}$$

In this procedure aliasing of noise $n(t)$ is avoided, because it is filtered away from the signal.

If Procedure 2 is applied, it is, in practice, necessary to truncate the convolution integral, and the smoothed signal becomes

$$h''(t) = \int_{-T_0}^{T_0} g(u)h(t-u)\,\mathrm{d}u = \int_{-\infty}^{\infty} \Pi_{2T_0}(u)g(u)h(t-u)\,\mathrm{d}u = \left[\underbrace{\Pi_{2T_0}(t)g(t)}_{g'(t)}\right] * h(t), \tag{11.17}$$

where

$$\Pi_{2T_0}(t) = \begin{cases} 1, & |t| \le T_0, \\ 0, & |t| > T_0. \end{cases} \tag{11.18}$$

The number of data points in this region must be odd, because otherwise the transfer function will not be real. The smoothed signal becomes

$$\begin{aligned} H''(f) &= \mathcal{F}^{-1}\{h''(t)\} \overset{(11.17)}{=} \mathcal{F}^{-1}\{\Pi_{2T_0}(t)g(t)\}H(f) \\ &= \left[\mathcal{F}^{-1}\{\Pi_{2T_0}(t)\} * G(f)\right] H(f) = G'(f)H(f), \end{aligned} \tag{11.19}$$

and the true transfer function

$$\boxed{G'(f) = \mathcal{F}^{-1}\{\Pi_{2T_0}(t)\} * G(f) = 2T_0 \operatorname{sinc}(2\pi f T_0) * G(f)\,.} \tag{11.20}$$

We notice that the truncation of the convolution integral causes that the true transfer function $G'(f)$ is the convolution of the optimal filter function $G(f)$ and a sinc function. This is illustrated in Figure 11.11. The true filter function has Gibbs oscillations around $f = f_a$ and $f = f_b$.

The Gibbs oscillations of $G'(f)$ may be reduced by using a weighted truncation function instead of the boxcar function $\Pi_{2T_0}(t)$. The truncation function can be, for example, the triangular function

$$\Lambda_{T_0}(t) = \begin{cases} 0, & |t| > T_0, \\ 1 - \dfrac{|t|}{T_0}, & |t| \le T_0. \end{cases} \tag{11.21}$$

We know that

$$\mathcal{F}^{-1}\{\Lambda_{T_0}(t)\} = T_0 \operatorname{sinc}^2(\pi f T_0), \tag{11.22}$$

and, consequently, the true transfer function of the spectrum, if triangular function is used in truncation, is

$$G''(f) = T_0 \operatorname{sinc}^2(\pi f T_0) * G(f). \tag{11.23}$$

This transfer function is illustrated in Figure 11.12.

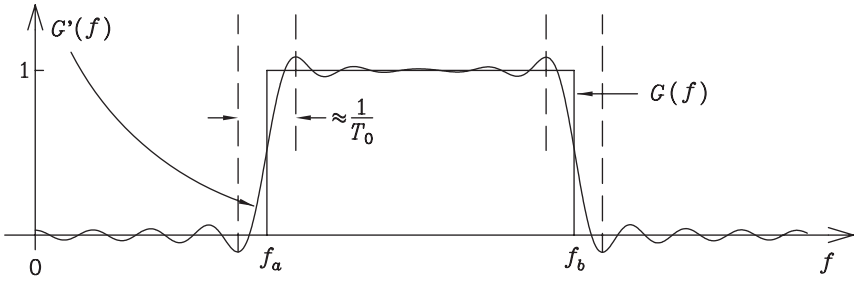


Figure 11.11: If the convolution integral in Equation 11.16 is truncated by a boxcar function, then the true transfer function $G'(f)$ is the convolution of the optimal filter function $G(f)$ and a sinc function: $G'(f) = 2T_0 \operatorname{sinc}(2\pi f T_0) * G(f)$.

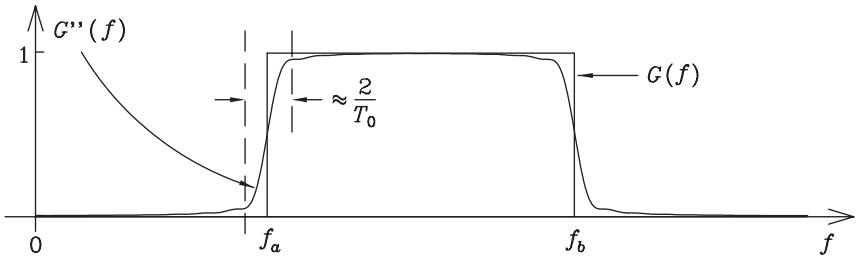


Figure 11.12: If the convolution integral in Equation 11.16 is truncated by a triangular function, then the true transfer function $G''(f)$ is the convolution of the optimal filter function $G(f)$ and a sinc^2 function: $G''(f) = T_0 \,\mathrm{sinc}^2(\pi f T_0) * G(f)$.

Table 11.1 summarizes the terminology of mathematical filtering.

A fast computer is able to calculate the smoothed signal from the convolution integral in Equation 11.17 in almost real time, with only a short delay T_0. This mathematical band-pass filter has a strong sharpness of approximately $1/T_0$. The minimum frequency f_a and the maximum frequency f_b of the band can be chosen almost arbitrarily. The filter may be programmed to a low-pass filter or a high-pass filter. The phase of the transfer function $G'(f)$ or $G''(f)$ is zero.

The convolution integral in Equation 11.16, which gives the smoothed signal, can be

Table 11.1: Summary of terminology.

$H(f)$		$h(t)$
spectral domain	$\mathcal{F}^{-1}$	signal domain
frequency domain	$\Leftarrow$	time domain
f-domain		t-domain
filtering =	$\mathcal{F}$	*smoothing* =
multiplying by filter function or transfer function $G(f)$	$\Rightarrow$	*convolution* with smoothing function or impulse response $g(t)$
$G(f)H(f)$		$\int_{-\infty}^{\infty} g(u)h(t-u)\,\mathrm{d}u$

written in the discrete form as

$$h'_k = \sum_{j=-\infty}^{\infty} g_j h_{k-j} \Delta t, \tag{11.24}$$

where $h_k = h(k\Delta t)$ and $\Delta t = 1/(2f_{\max})$. The transfer function $G(f)$ can be calculated from the discrete impulse response g_j in the continuous form

$$G(f) = \sum_{j=-\infty}^{\infty} g_j e^{-i2\pi f j \Delta t} \Delta t = \sum_{j=-\infty}^{\infty} g_j e^{-i\pi f j / f_{\max}} \Delta t, \tag{11.25}$$

or in the discrete form, where $f = k\Delta f = k/(2T)$, and $T = N\Delta t$,

$$G_k = \sum_{j=-N}^{N} g_j e^{-i\pi j k/N} \Delta t. \tag{11.26}$$

If calculations are performed by FFT, then $N = 2^m$, where m is a positive integer.

Example 11.1: Averaging three successive signal data is a method which is sometimes applied in order to smooth a signal. What is the transfer function (filter function) of this method?

Solution. We can apply the discrete convolution integral in Equation 11.24. The smoothed signal is

$$h'_k = \Delta t \sum_{j=-\infty}^{\infty} g_j h_{k-j} = \frac{h_{k-1} + h_k + h_{k+1}}{3}.$$

We can see that the discrete impulse response is

$$g_{-1} = g_0 = g_1 = \frac{1}{3\Delta t}.$$

Applying Equation 11.25, we can calculate the transfer function

$$\begin{aligned} G(f) &= \sum_{j=-\infty}^{\infty} g_j e^{-i2\pi f j \Delta t} \Delta t = \frac{1}{3} e^{i2\pi f \Delta t} + \frac{1}{3} + \frac{1}{3} e^{-i2\pi f \Delta t} \\ &= \frac{1}{3} [1 + 2\cos(2\pi f \Delta t)] = \frac{1}{3} [1 + 2\cos(\pi f / f_{\max})]. \end{aligned} \tag{11.27}$$

This transfer function is shown in Figure 11.13. The transfer function is periodic with the period $2f_{\max} = 1/\Delta t$. This transfer function seems very bad, and even reaches zero at some frequency value. This smoothing method cannot be recommended.

It is possible to find this same transfer function by an alternative method: starting from a continuous impulse response, and performing a *discontinuity correction.* We must, however, be careful in finding the exactly right function. If the continuous impulse response which

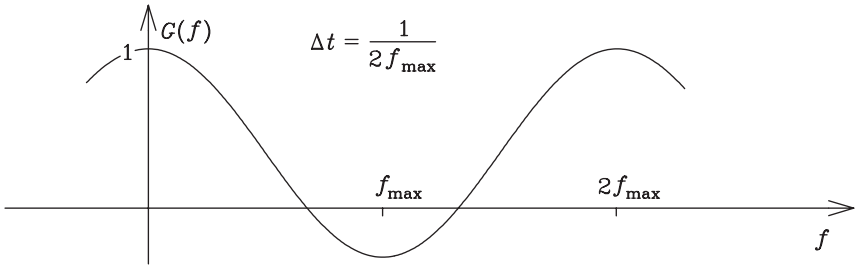


Figure 11.13: The transfer function $G(f) = [1 + 2\cos(\pi f/f_{\max})]/3$, which is obtained, if the signal is smoothed by averaging three successive data.

corresponds to averaging three successive data is thought to be the boxcar function of length $2\Delta t$ and height $1/(3\Delta t)$, shown in Figure 11.14, we obtain the transfer function

$$G(f) = \mathcal{F}^{-1}\{g(t)\} = \frac{2\Delta t}{3\Delta t}\ \mathrm{sinc}(2\pi f \Delta t).$$

The transfer function corresponding to the discrete impulse response with sampling interval Δt is obtained from the transfer function corresponding to the continuous $g(t)$ by using Equation 3.14:

$$\begin{aligned} G^{\Delta t}(f) &\overset{(3.14)}{=} \sum_{j=-\infty}^{\infty} G\left(f - \frac{j}{\Delta t}\right) = \frac{2}{3}\sum_{j=-\infty}^{\infty} \mathrm{sinc}\left[2\pi\left(f - \frac{j}{\Delta t}\right)\Delta t\right] \\ &\neq \frac{1}{3}[1 + 2\cos(2\pi f \Delta t)]. \end{aligned} \tag{11.28}$$

We did not obtain the right transfer function, since, for example, $G^{\Delta t}(f_{\max}) = 0$.

In order to obtain the correct transfer function, Equation 11.27, we must make a discontinuity correction and use as the continuous impulse response the boxcar function of length $3\Delta t$

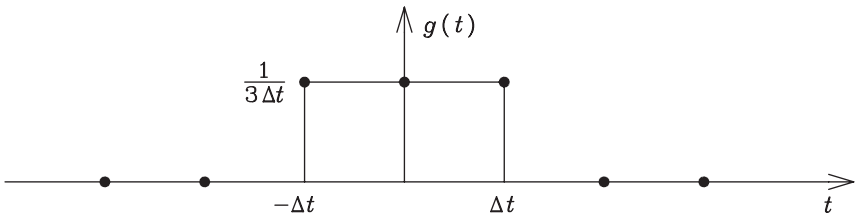


Figure 11.14: The wrong impulse response $g(t)$, which does not correspond to the situation of smoothing a signal by averaging three successive data with sampling interval Δt.

and height $1/(3\Delta t)$, shown in Figure 11.15. With the help of Equation 9.22, we can calculate the transfer function

$$\begin{aligned} G^{\Delta t}(f) &= \frac{3\Delta t}{3\Delta t}\sum_{j=-\infty}^{\infty}\operatorname{sinc}\left[2\pi\left(f-\frac{j}{\Delta t}\right)\frac{3}{2}\Delta t\right] \\ &= \frac{\sin(3\pi f\Delta t)}{3\sin(\pi f\Delta t)} = \frac{1}{3}\left[1+2\cos(2\pi f\Delta t)\right]. \end{aligned} \tag{11.29}$$

We can see that now we have obtained the right $G(f)$ (Equation 11.27).

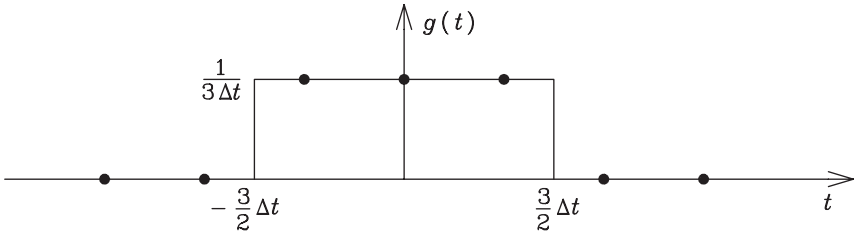


Figure 11.15: The correct impulse response $g(t)$, which corresponds to the situation of smoothing a signal by averaging three successive data with sampling interval Δt.

Example 11.2: Smooth a signal by averaging N successive data, in such a way that each datum is replaced by the average of the datum itself and $N-1$ previous data. N is an odd integer.

Solution. The smoothed signal is

$$h'_k = \frac{1}{N}\sum_{j=0}^{N-1} h_{k-j}.$$

The discrete impulse response is

$$g_j = \frac{1}{N\Delta t}, \qquad j = 0, 1, 2, \ldots, N-1,$$

and the transfer function is

$$G(f) = \sum_{j=0}^{N-1} g_j e^{-i2\pi f j\Delta t}\Delta t = \frac{1}{N}\sum_{j=0}^{N-1}(e^{-i2\pi f\Delta t})^j = \frac{1}{N}\,\frac{1-e^{-i2\pi f\Delta t N}}{1-e^{-i2\pi f\Delta t}}, \tag{11.30}$$

because the sum of the geometrical series is

$$\sum_{n=0}^{N-1} x^n = \frac{1-x^N}{1-x}.$$

The transfer function can also be written in the forms

$$\begin{aligned} G(f) &= \frac{1}{N}\frac{(1-e^{-i2\pi f\Delta tN})}{(1-e^{-i2\pi f\Delta t})} = \frac{1}{N}\frac{e^{-i\pi f\Delta tN}(e^{i\pi f\Delta tN}-e^{-i\pi f\Delta tN})}{e^{-i\pi f\Delta t}(e^{i\pi f\Delta t}-e^{-i\pi f\Delta t})} \\ &= \frac{1}{N}\frac{2i\sin(\pi f\Delta tN)}{2i\sin(\pi f\Delta t)}e^{-i(N-1)\pi f\Delta t}. \end{aligned} \tag{11.31}$$

The same transfer function will be found, if we use the continuous impulse response shown in Figure 11.16. The impulse response is a boxcar function of length $N\Delta t$ and height $1/(N\Delta t)$, which has been shifted $\frac{N-1}{2}\Delta t$ from the origin. The inverse Fourier transform of the boxcar function around the origin is

$$\mathcal{F}^{-1}\{g(t+\frac{N-1}{2}\Delta t)\} = N\Delta t\,\mathrm{sinc}(\pi f N\Delta t)/(N\Delta t)$$

and, applying the shift theorem,

$$\mathcal{F}^{-1}\{g(t)\} = \mathrm{sinc}(\pi f N\Delta t)e^{-i2\pi f(N-1)\Delta t/2} = G(f).$$

Consequently, the transfer function corresponding to the discrete impulse response with interval Δt is

$$G^{\Delta t}(f) = \sum_{j=-\infty}^{\infty}\mathrm{sinc}\left[\pi N\left(f-\frac{j}{\Delta t}\right)\Delta t\right]e^{-i(N-1)\pi f\Delta t}.$$

Remembering Equation 9.22, we obtain

$$G^{\Delta t}(f) = \frac{1}{N}\frac{\sin(N\pi f\Delta t)}{\sin(\pi f\Delta t)}e^{-i(N-1)\pi f\Delta t}. \tag{11.32}$$

We can see that we obtained the same result as in Equation 11.31.

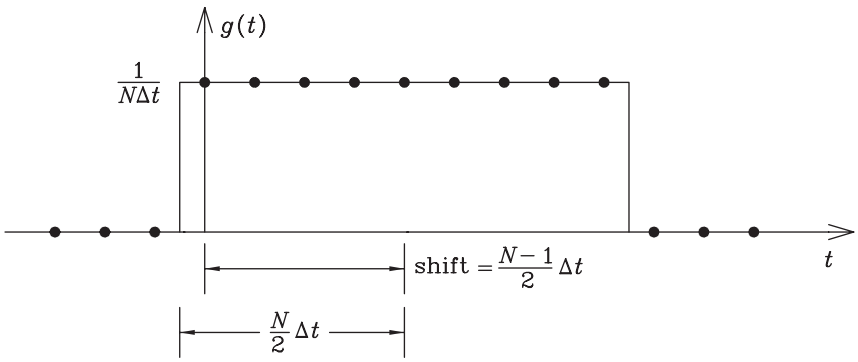


Figure 11.16: The continuous impulse response $g(t)$, which corresponds to the situation of smoothing a signal by averaging N successive data, chosen as in Example 11.2.

11.3 Mathematical smoothing

A measured spectrum often consists of spectral lines, which have a certain line shape, and of white noise. Such a spectrum is illustrated in Figure 11.17. *Smoothing* of a spectrum $H(f)$ is an operation where the spectrum is convolved with a smoothing function $W(f)$ in order to reduce the rapidly oscillating random noise of the data.

A spectrum with noise can be written as $H(f) = H_0(f) + N(f)$, where $H_0(f)$ is the true spectrum and $N(f)$ is the white noise spectrum. The smoothing operation in the frequency domain can be expressed as

$$H'(f) = W(f) * H(f), \tag{11.33}$$

where $H'(f)$ is the smoothed spectrum.

According to the convolution theorem, the Fourier transform of a convolution is a simple multiplication. Consequently, convolution of a spectrum in the frequency domain is the same as multiplication of a signal in the time domain:

$$\mathcal{F}\{H'(f)\} = \mathcal{F}\{W(f)\}\mathcal{F}\{H(f)\} = A(t)h(t). \tag{11.34}$$

$A(t) = \mathcal{F}\{W(f)\}$ is the weight function, by which the original signal $h(t) = \mathcal{F}\{H(f)\}$ is multiplied. This operation can be regarded as filtering of the signal. Smoothing of a spectrum corresponds to filtering of the signal, in the same way as filtering of a spectrum corresponds to smoothing of the signal.

Multiplication of the signal $h(t) = \mathcal{F}\{H(f)\}$ by the weight function $A(t) = \mathcal{F}\{W(f)\}$ is the same operation as is performed in the apodization method.

The aim of smoothing is to enhance the signal-to-noise ratio (S/N) by reducing the noise as much as possible, but distorting the true spectral line shape as little as possible.

The simplest way of smoothing a spectrum is the simple truncation of the signal. This makes sense, because the ratio of true information and noise is generally the best near $t = 0$, and the worst at large values of t. In truncation, the signal $h(t)$ is multiplied by a boxcar

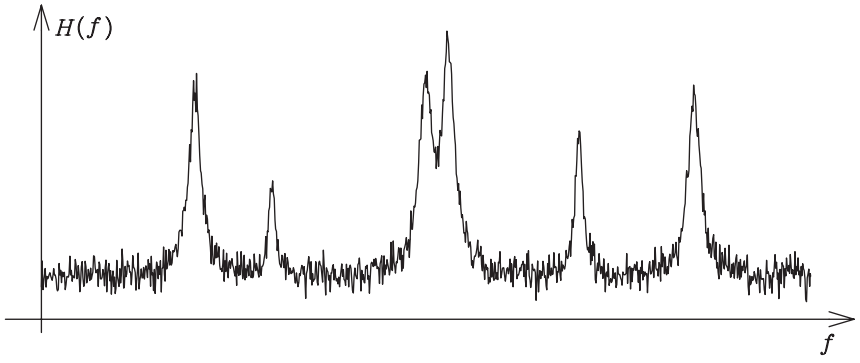


Figure 11.17: A spectrum which consists of spectral lines and white noise.

function

$$\Pi_{2T_i}(t) = \begin{cases} 1, & |t| \le T_i, \\ 0, & |t| > T_i. \end{cases} \tag{11.35}$$

Multiplication by a boxcar function is the weakest form of apodization, which does not modify the signal or the noise in the retained portion of the signal.

An *optimal smoothing* is achieved, if the signal is truncated at such a point T_i beyond which all information $h(t)$ disappears under the noise $n(t)$. This is illustrated in Figure 11.18. At the point T_i the rms, or the root mean square, of the noise

$$\sqrt{\overline{n^2(t)}} = \sqrt{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} n^2(t)\,\mathrm{d}t} \tag{11.36}$$

is equal to the amplitude of the signal.

The signal of an optimally smoothed spectrum is

$$h'(t) = \Pi_{2T_i}(t)h(t), \tag{11.37}$$

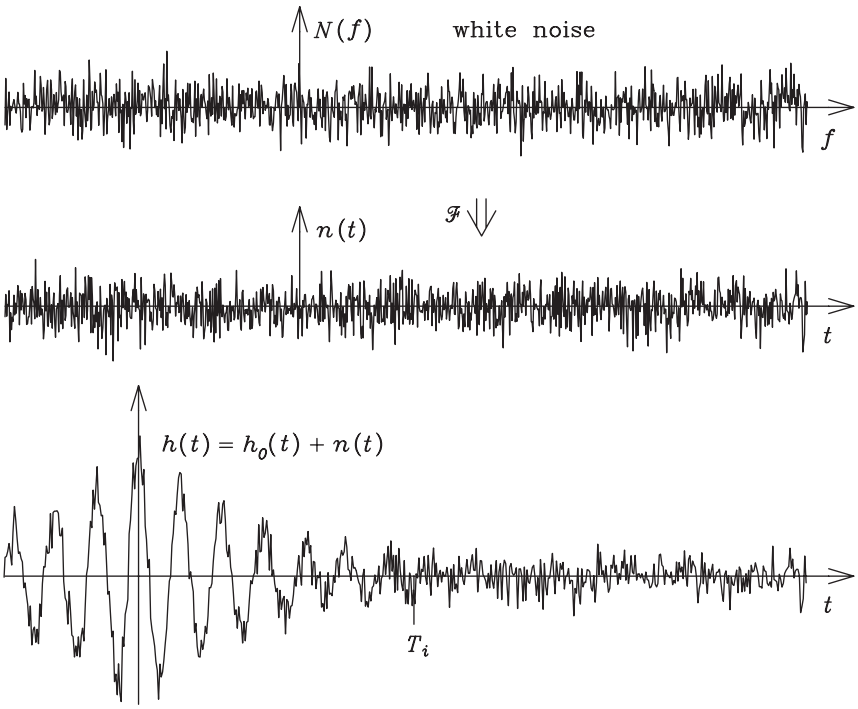


Figure 11.18: White noise spectrum $N(f)$ in f-domain, white signal noise $n(t) = \mathcal{F}\{N(f)\}$ in t-domain, and the noise-corrupted signal $h(t) = h_0(t) + n(t)$. Beyond the point T_i all information $h_0(t)$ disappears under the noise $n(t)$.

and, consequently, the optimally smoothed spectrum is

$$H'(f) = \mathcal{F}^{-1}\{\Pi_{2T_i}(t)h(t)\} = \mathcal{F}^{-1}\{\Pi_{2T_i}(t)\} * H(f) = 2T_i \,\mathrm{sinc}(2\pi f T_i) * H(f). \quad (11.38)$$

We can see that the smoothing function $W(f)$ of the spectrum is a sinc function.

A conventional way to approximate the spectral line shape is to assume that it is *Lorentzian*, shown in Figure 11.19. The spectrum is written as a sum of Lorentzian lines, that is,

$$H_0(f) = \sum_j A_j \left[\frac{\sigma_j/\pi}{\sigma_j^2 + (f - f_j)^2} + \frac{\sigma_j/\pi}{\sigma_j^2 + (f + f_j)^2} \right] = H_0(-f), \quad (11.39)$$

where A_j is the area of a line, f_j is the position of a line, and the full width at half maximum,

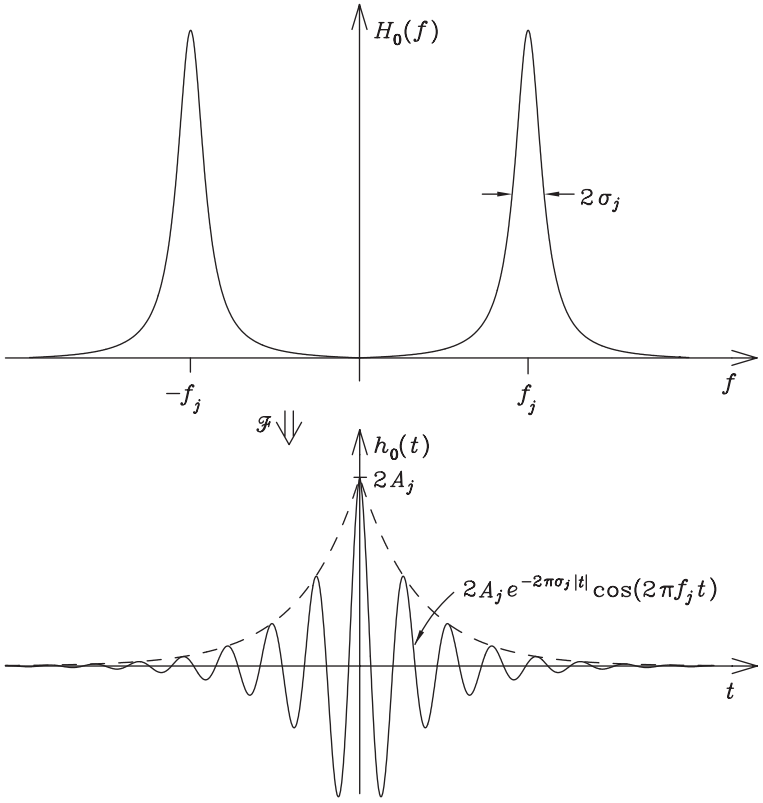


Figure 11.19: A symmetric pair of Lorentzian spectral lines. The position of the lines is $\mp f_j$, the area A_j, and the FWHM $2\sigma_j$. The Fourier transform of the lines is $2A_j e^{-2\pi\sigma_j|t|}\cos(2\pi f_j t)$.

FWHM, of a line is $2\sigma_j$. The Fourier transform of a Lorentzian line is

$$\begin{aligned} & \mathcal{F}\left\{A_j \frac{\sigma_j/\pi}{\sigma_j^2+(f-f_j)^2} + A_j \frac{\sigma_j/\pi}{\sigma_j^2+(f+f_j)^2}\right\} \\ = \quad & \mathcal{F}\left\{\frac{\sigma_j/\pi}{\sigma_j^2+f^2}\right\} A_j \left[e^{i2\pi f_j t} + e^{-i2\pi f_j t}\right] \\ = \quad & e^{-2\pi\sigma_j|t|}\, 2A_j \cos(2\pi f_j t) = h_0(t). \end{aligned} \tag{11.40}$$

We can see that the signal which corresponds to a Lorentzian spectral line is an exponentially decaying cosine function. This is also illustrated in Figure 11.19.

If the spectral line shape is Lorentzian, then the optimal smoothing is obtained, if

$$2A_j e^{-2\pi\sigma_j T_i} = \sqrt{\overline{n^2(t)}}. \tag{11.41}$$

Figure 11.20 shows a simulation of a smoothing operation. In practice, it is impossible to

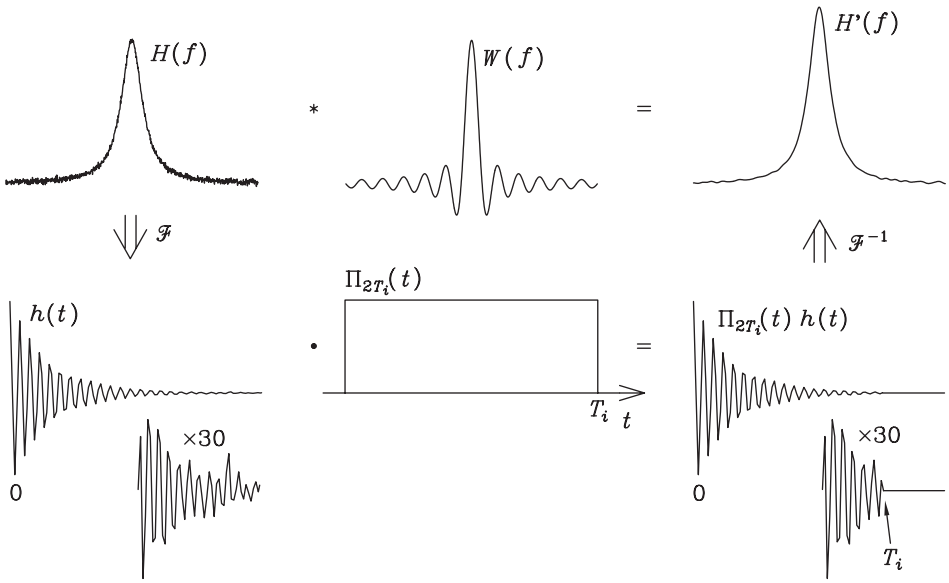


Figure 11.20: The spectrum $H(f)$, which has a signal-to-noise ratio of 20, can be smoothed to the spectrum $H'(f)$ by two alternative procedures. In the first procedure, the smoothed spectrum is obtained by convolving the original spectrum by a smoothing function $W(f)$, which in this simulation is a sinc function. This is a laborious procedure. In the second procedure the spectrum is Fourier transformed into the signal $h(t) = \mathcal{F}\{H(f)\}$, and multiplied by a boxcar function $\Pi_{2T_i}(t)$. The smoothed spectrum is obtained as the inverse Fourier transform of the product $\Pi_{2T_i}(t)h(t)$. The weak portions of the signals are shown 30-fold magnified around T_i.

carry out sinc smoothing in the frequency domain exactly, because the convolution has to be truncated. The optimal smoothing is much easier to perform in the time domain. This is both fast and accurate.

11.4 Distortion and (S/N) enhancement in smoothing

Let us next consider, how much the spectral line shape is distorted by the smoothing operation. The height of a general noise-free spectral line shape $W_0(f)$ is given by

$$W_0(0) = \int_{-\infty}^{\infty} I_0(t)\,\mathrm{d}t, \tag{11.42}$$

where $I_0(t) = \mathcal{F}\{W_0(f)\}$ is the noise-free signal corresponding to the noise-free spectral line shape. In a real measurement, the signal cannot be infinite, but it is limited by a window function $A_0(t)$. This is instrumental distortion. In other words, the spectral line shape is smoothed by an intrinsic smoothing function $\mathcal{F}^{-1}\{A_0(t)\}$ of the recording system. Very often the signal is merely truncated, and $A_0(t)$ is a boxcar function. The recorded line shape of a noise-free line is

$$W_\mathrm{R}(f) = \int_{-\infty}^{\infty} A_0(t) I_0(t) e^{-i2\pi f t}\,\mathrm{d}t. \tag{11.43}$$

Its peak height is

$$W_\mathrm{R}(0) = \int_{-\infty}^{\infty} A_0(t) I_0(t)\,\mathrm{d}t. \tag{11.44}$$

After smoothing by a smoothing function $W(f) = \mathcal{F}^{-1}\{A(t)\}$ the noise-free line shape is

$$W'(f) = \mathcal{F}^{-1}\{A(t)A_0(t)I_0(t)\} = \int_{-\infty}^{\infty} A(t)A_0(t)I_0(t)e^{-i2\pi f t}\,\mathrm{d}t, \tag{11.45}$$

and its peak height is

$$W'(0) = \int_{-\infty}^{\infty} A(t)A_0(t)I_0(t)\,\mathrm{d}t. \tag{11.46}$$

The relative distortion of the recorded line shape $W_\mathrm{R}(f)$ by smoothing can be defined as

$$\Delta(f) = \frac{W'(f) - W_\mathrm{R}(f)}{W_\mathrm{R}(0)}, \tag{11.47}$$

and the maximum relative distortion is

$$\Delta(0) = \frac{W'(0) - W_{\mathrm{R}}(0)}{W_{\mathrm{R}}(0)} = \frac{\int_{-\infty}^{\infty} [A(t) - 1]\, A_0(t) I_0(t)\, \mathrm{d}t}{\int_{-\infty}^{\infty} A_0(t) I_0(t)\, \mathrm{d}t}. \tag{11.48}$$

It is also possible to determine the relative distortion $\delta_{1/2}$ of the FWHM of the line shape in smoothing. It can be shown to be [5]

$$\delta_{1/2} = \frac{\mathrm{FWHM}_{\mathrm{smoothed}} - \mathrm{FWHM}_{\mathrm{recorded}}}{\mathrm{FWHM}_{\mathrm{recorded}}} \approx \frac{-W_{\mathrm{R}}(0)\, \Delta(0)}{W'(0)} = \frac{-\Delta(0)}{\Delta(0) + 1}. \tag{11.49}$$

Another estimate of the efficiency of the smoothing operation is the enhancement of the signal-to-noise ratio. Let us assume that the random white noise $N(f)$ and its Fourier transform, the noise signal $n(t)$, both are even and real. According to the Parseval's relation, the noise power spectrum $N^2(f)$ and the noise power interferogram $n^2(t)$ are related by

$$\int_{-\infty}^{\infty} N^2(f)\, \mathrm{d}f = \int_{-\infty}^{\infty} n^2(t)\, \mathrm{d}t. \tag{11.50}$$

An experimental spectrum is, in practice, bandlimited between $-f_m$ and f_m. Applying the Parseval's relation to the noise power spectrum $N_{\mathrm{R}}^2(f)$ after smoothing by the intrinsic smoothing function of the system, and taking into account that random white noise varies much more rapidly than the window function $A_0(t)$, we obtain that

$$\begin{aligned} \int_{-f_m}^{f_m} N_{\mathrm{R}}^2(f)\, \mathrm{d}f &= \int_{-\infty}^{\infty} A_0^2(t) n^2(t)\, \mathrm{d}t \\ &\approx \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} n^2(t)\, \mathrm{d}t \int_{-\infty}^{\infty} A_0^2(t)\, \mathrm{d}t \\ &= \overline{n^2} \int_{-\infty}^{\infty} A_0^2(t)\, \mathrm{d}t. \end{aligned} \tag{11.51}$$

The rms value of the noise in the recorded spectrum is

$$\sqrt{\overline{N_{\mathrm{R}}^2}} = \sqrt{\frac{\overline{n^2}}{2f_m} \int_{-\infty}^{\infty} A_0^2(t)\, \mathrm{d}t}. \tag{11.52}$$

Using the maximum value (Equation 11.44) of a line as the signal intensity, we obtain the

signal-to-noise ratio of the recorded line

$$(S/N)_{\mathrm{R}} = \frac{W_{\mathrm{R}}(0)}{\sqrt{\overline{N_{\mathrm{R}}^2}}} = \frac{\sqrt{2f_m}\int_{-\infty}^{\infty} A_0(t)I_0(t)\,\mathrm{d}t}{\sqrt{\overline{n^2}}\sqrt{\int_{-\infty}^{\infty} A_0^2(t)\,\mathrm{d}t}}. \tag{11.53}$$

After smoothing by $W(f) = \mathcal{F}^{-1}\{A(t)\}$ the signal $I_0(t)$ has been multiplied by both $A_0(t)$ and $A(t)$, and analogously to Equation 11.53, the signal-to-noise ratio of the smoothed spectrum is

$$(S/N)' = \frac{W'(0)}{\sqrt{\overline{N_A^2}}} = \frac{\sqrt{2f_m}\int_{-\infty}^{\infty} A(t)A_0(t)I_0(t)\,\mathrm{d}t}{\sqrt{\overline{n^2}}\sqrt{\int_{-\infty}^{\infty} A(t)^2A_0^2(t)\,\mathrm{d}t}}. \tag{11.54}$$

Thus, the general formula of the signal-to-noise ratio enhancement Q resulting from the smoothing operation is

$$\begin{aligned} Q = \frac{(S/N)'}{(S/N)_{\mathrm{R}}} &= \frac{\int_{-\infty}^{\infty} A(t)A_0(t)I_0(t)\,\mathrm{d}t\sqrt{\int_{-\infty}^{\infty} A_0^2(t)\,\mathrm{d}t}}{\int_{-\infty}^{\infty} A_0(t)I_0(t)\,\mathrm{d}t\sqrt{\int_{-\infty}^{\infty} A^2(t)A_0^2(t)\,\mathrm{d}t}} \\ &= \frac{W'(0)\sqrt{\int_{-\infty}^{\infty} A_0^2(t)\,\mathrm{d}t}}{W_{\mathrm{R}}(0)\sqrt{\int_{-\infty}^{\infty} A^2(t)A_0^2(t)\,\mathrm{d}t}}. \end{aligned} \tag{11.55}$$

We can see that the efficiency of the smoothing operation depends on the smoothing functions $\mathcal{F}^{-1}\{A(t)\}$ and $\mathcal{F}^{-1}\{A_0(t)\}$, as well as on the original line shape $W_0(f) = \mathcal{F}^{-1}\{I_0(t)\}$.

In Fourier transform spectroscopy, $A_0(t)$ is the apodization function which is applied when computing the spectrum $H(f)$. Usually, it is the boxcar function

$$A_0(t) = \begin{cases} 0, & |t| > T_0, \\ 1, & |t| \le T_0. \end{cases} \tag{11.56}$$

In this case, the signal-to-noise ratio enhancement of the smoothing operation is

$$Q_0 = \frac{\int_{-T_0}^{T_0} A(t)I_0(t)\,\mathrm{d}t\sqrt{2T_0}}{\int_{-T_0}^{T_0} I_0(t)\,\mathrm{d}t\sqrt{\int_{-T_0}^{T_0} A^2(t)\,\mathrm{d}t}}. \tag{11.57}$$

Generally, if we wish to calculate $\Delta(f)$, $\Delta(0)$, or $\delta_{1/2}$, we can approximate $A_0(t)$ by one (and T_0 by infinite), because $I_0(t)$ is damped much faster than $A_0(t)$. In the calculation of Q, however, the shape of $A_0(t)$ is essential, because the noise $n(t)$ is not damped, and $A_0(t)$ determines how much the spectrum contains noise before smoothing.

Smoothing is often described by the *smoothing parameter* K, which is the ratio

$$K = \frac{\text{FWHM of the original line}}{\text{FWHM of the smoothing function}}. \tag{11.58}$$

Smoothing by truncation of the registered signal is characterized by the parameter K_0, which is the ratio

$$K_0 = \frac{\text{FWHM of the original line}}{\text{FWHM of sinc function caused by truncation}}. \tag{11.59}$$

Example 11.3: A Lorentzian line is smoothed by a sinc function. What is the maximum relative distortion, if we know the smoothing parameter K?

Solution. The Lorentzian line is of the form $W_0(f) = \dfrac{\sigma/\pi}{\sigma^2 + f^2}$. Its FWHM is 2σ. The signal which corresponds to this line shape is $I_0(t) = \mathcal{F}\{W_0(f)\} = e^{-2\pi\sigma|t|}$.

The FWHM of a sinc-shaped smoothing function $W(f) = 2T\,\mathrm{sinc}(\pi 2Tf)$ is approximately $1.2067/(2T)$. We obtain

$$K \approx \frac{2\sigma}{1.2067/(2T)}.$$

Consequently,

$$T \approx \frac{1.2067K}{4\sigma}.$$

The Fourier transform of the sinc-shaped smoothing function is the boxcar function

$$A(t) = \begin{cases} 1, & |t| \le T, \\ 0, & |t| > T. \end{cases}$$

This is the apodization function.

Assuming that the recorded line shape is the same as the true line shape (there is no truncation), the maximum relative distortion of the line shape is

$$\begin{aligned} \Delta(0) &= \frac{W_0(0) * W(0) - W_0(0)}{W_0(0)} = \frac{\int_{-\infty}^{\infty} [A(t) - 1]\, e^{-2\pi\sigma|t|}\, \mathrm{d}t}{1/(\pi\sigma)} \\ &= -\pi\sigma 2 \int_T^{\infty} e^{-2\pi\sigma t}\, \mathrm{d}t = \frac{-2\pi\sigma}{-2\pi\sigma}(0 - e^{-2\pi\sigma T}) = -e^{-2\pi\sigma T} \\ &\approx -e^{-1.2067\pi K/2}. \end{aligned}$$

Example 11.4: A Lorentzian line is smoothed by a sinc function. The original signal is registered in the interval $[-T_0, T_0]$ (boxcar truncation). Show that the signal-to-noise ratio enhancement of the smoothing is

$$Q_0 \approx \sqrt{K_0/K}\ [1 + \Delta(0)]\,,$$

where

$$K = \frac{\text{FWHM of the original line}}{\text{FWHM of the smoothing function}},$$

and

$$K_0 = \frac{\text{FWHM of the original line}}{\text{FWHM of sinc function caused by truncation}}.$$

$\Delta(0)$ is the maximum relative distortion of the line shape.

Solution. The line is, as in the previous example, $W_0(f) = \dfrac{\sigma/\pi}{\sigma^2 + f^2}$. Its FWHM is 2σ. The signal which corresponds to this line shape is $I_0 = \mathcal{F}\{W_0(f)\} = e^{-2\pi\sigma|t|}$.

The signal is truncated by the boxcar function

$$A_0(t) = \begin{cases} 1, & |t| \le T_0, \\ 0, & |t| > T_0. \end{cases}$$

The FWHM of the intrinsic smoothing function $\mathcal{F}^{-1}\{A_0(t)\} = 2T_0\,\mathrm{sinc}(\pi 2T_0 f)$ is approximately $1.2067/(2T_0)$. We obtain that

$$K_0 \approx \frac{2\sigma}{1.2067/(2T_0)},$$

where 2σ is the FWHM of a Lorentzian line. Consequently,

$$T_0 \approx \frac{1.2067 K_0}{4\sigma}.$$

As in the previous example, the smoothing function is $W(f) = 2T\,\mathrm{sinc}(\pi 2Tf)$. Its FWHM is approximately $1.2067/(2T)$, and

$$T \approx \frac{1.2067 K}{4\sigma}.$$

The corresponding apodization function is, again,

$$A(t) = \begin{cases} 1, & |t| \le T, \\ 0, & |t| > T. \end{cases}$$

The truncation boxcar must be much longer than the apodization boxcar: $T_0 \gg T$.

In order to be able to apply Equation 11.57, which gives the signal-to-noise ratio enhancement of smoothing in the case of boxcar truncation, we need to find two quantities. Since

$A_0(t)$ is a boxcar function, the maximum relative distortion of the line shape, Equation 11.48, can be written as

$$\Delta(0) = \frac{\int_{-T_0}^{T_0} [A(t) - 1]\, I_0(t)\, \mathrm{d}t}{\int_{-T_0}^{T_0} I_0(t)\, \mathrm{d}t}.$$

From this equation we get that

$$\int_{-T_0}^{T_0} A(t) I_0(t)\, \mathrm{d}t = \int_{-T_0}^{T_0} I_0(t)\, \mathrm{d}t\, [\Delta(0) + 1]\,.$$

For the boxcar apodization function, since $T \ll T_0$,

$$\int_{-T_0}^{T_0} A^2(t) dt = \int_{-T}^{T} A^2(t) dt = 2T \approx \frac{1.2067K}{2\sigma}.$$

Inserting these in Equation 11.57, we find that

$$\begin{aligned} Q_0 &= \frac{\int_{-T_0}^{T_0} A(t) I_0(t)\, \mathrm{d}t \sqrt{2T_0}}{\int_{-T_0}^{T_0} I_0(t)\, \mathrm{d}t \sqrt{\int_{-T_0}^{T_0} A^2(t)\, \mathrm{d}t}} = \frac{\int_{-T_0}^{T_0} I_0(t)\, \mathrm{d}t\ [\Delta(0) + 1] \sqrt{2T_0}}{\int_{-T_0}^{T_0} I_0(t)\, \mathrm{d}t \sqrt{2T}} \\ &\approx \frac{\sqrt{1.2067K_0/(2\sigma)}}{\sqrt{1.2067K/(2\sigma)}}\, [\Delta(0) + 1] = \sqrt{K_0/K}\, [\Delta(0) + 1]\,. \end{aligned}$$

Example 11.5: The height of a Lorentzian spectral line is $H_0(0) = S$. The FWHM of the line is 2σ. The signal was in the recording multiplied by a boxcar in the interval $[-T_0, T_0]$. The spectrum also contains white noise of standard deviation N, and the signal-to-noise ratio is thus S/N. The line is smoothed by a sinc function $W(f) = 2T\ \mathrm{sinc}(\pi 2Tf)$. As shown in the previous example, the signal-to-noise ratio is in smoothing changed by the factor

$$Q_0 \approx \sqrt{K_0/K}\ [1 + \Delta(0)]\,.$$

The change of the height of the line is $\Delta(0) \approx -e^{-1.2067\pi K/2}$. (a) What is the standard deviation N' of the noise after smoothing? (b) How should the apodization truncation point T be chosen, if we want the standard deviation N' of the smoothed noise to be equal to the distortion of the height of the line?

Solution. (a) The new signal-to-noise ratio after smoothing is

$$\frac{S'}{N'} = Q_0 \frac{S}{N}\,.$$

On the other hand, the height of the line after smoothing is

$$S' = [1 + \Delta(0)]\ S.$$

Consequently,

$$N' = \frac{N}{S}\,\frac{S'}{Q_0} \approx \sqrt{K/K_0}\,N.$$

Remembering that $K \approx \dfrac{2\sigma}{1.2067/(2T)}$ and $K_0 \approx \dfrac{2\sigma}{1.2067/(2T_0)}$, we can write the standard deviation of noise after smoothing as

$$N' \approx \sqrt{T/T_0}\,N.$$

(b) We want that

$$N' \approx \sqrt{T/T_0}\,N = |\Delta(0)S| = e^{-1.2067\pi K/2}\,S = e^{-2\pi\sigma T}\,S.$$

Consequently,

$$e^{2\pi\sigma T} = \sqrt{T_0/T}\,\frac{S}{N}.$$

This yields

$$T = \frac{1}{2\pi\sigma}\,\ln\left(\sqrt{T_0/T}\,\frac{S}{N}\right).$$

The truncation point T should fulfill this condition.

11.5 Comparison of smoothing functions

In the previous section, in Examples 11.3. and 11.4, we noticed that smoothing of a spectrum with a sinc function corresponds to apodization of the signal by a boxcar function. If a Lorentzian line $W_0(f) = \dfrac{\sigma/\pi}{\sigma^2 + f^2}$ is smoothed with a sinc function, the maximum relative distortion of the line shape is, as we found out,

$$\Delta(0) \approx -e^{-1.207\,\pi K/2}, \tag{11.60}$$

where K is the smoothing parameter. The signal-to-noise ratio enhancement of sinc smoothing was shown to be

$$Q_0 \approx \sqrt{K_0/K}\left[1 - e^{-1.207\,\pi K/2}\right], \tag{11.61}$$

where K_0 is the smoothing parameter of the truncation of the registered signal.

If a spectrum is smoothed with the sinc^2 function $W(f) = T\,\text{sinc}^2(\pi T f)$, then the corresponding apodization function of the signal is the triangular function

$$A(t) = \Lambda_T(t) = \begin{cases} 1 - \dfrac{|t|}{T}, & |t| \le T, \\ 0, & |t| > T. \end{cases} \tag{11.62}$$

It can be shown (Problem 12) that in this case the maximum relative distortion of a Lorentzian line is

$$\Delta(0) \approx \frac{2}{1.7718\,\pi K}\left[\exp\left(-\frac{1.7718\,\pi K}{2}\right) - 1\right]. \tag{11.63}$$

Expressed using this maximum relative distortion, the signal-to-noise ratio enhancement of the sinc^2 smoothing of a Lorentzian line can be shown to be (Problem 15)

$$Q_0 \approx \sqrt{\frac{3 \times 1.2067\,K_0}{1.7718\,K}}\,[1 + \Delta(0)]\,. \tag{11.64}$$

If a spectrum is smoothed by a Lorentzian function $W(f) = \dfrac{\sigma/(K\pi)}{(\sigma/K)^2 + f^2}$, then the corresponding signal is multiplied by the exponential apodization function

$$A(f) = e^{-2\pi\sigma|t|/K}. \tag{11.65}$$

The maximum relative distortion of a Lorentzian line in smoothing with the Lorentzian function can be shown to be

$$\Delta(0) = \frac{-1}{K+1}, \tag{11.66}$$

and the signal-to-noise ratio enhancement can be shown to be

$$Q_0 = \sqrt{\frac{1.207\,\pi K_0}{K}}\,\frac{K}{K+1}. \tag{11.67}$$

Figure 11.21 shows the maximum relative distortion $-\Delta(0)$ of a Lorentzian line as a function of the smoothing parameter K in the cases where the line is smoothed with the three smoothing functions discussed above. We can see that the maximum distortion depends strongly on the choice of the smoothing function.

Figure 11.22 shows the behavior of the signal-to-noise ratio enhancement Q_0 as a function of K in the same three cases. Also Q_0 depends on the smoothing function. Apodization by an exponential function (smoothing by Lorentzian function) is a very strong form of apodization, and it produces efficient suppression of noise, but, simultaneously, strongly distorts the line shape. Boxcar apodization (smoothing by sinc function) is the weakest type of apodization, where the retained portions of the signal and noise are not modified. Triangular apodization (smoothing by sinc^2 function) represents intermediate apodization between these two extremes.

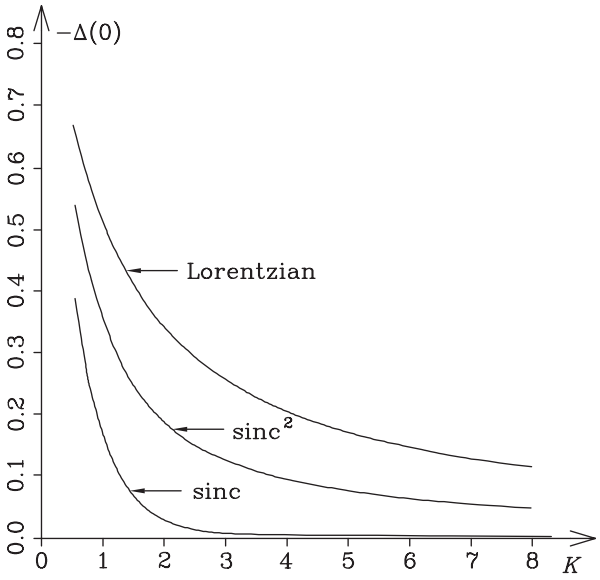


Figure 11.21: Maximum relative distortion $-\Delta(0)$ as a function of the smoothing parameter K in the cases where a Lorentzian spectrum is smoothed with a Lorentzian, sinc^2, or sinc smoothing function.

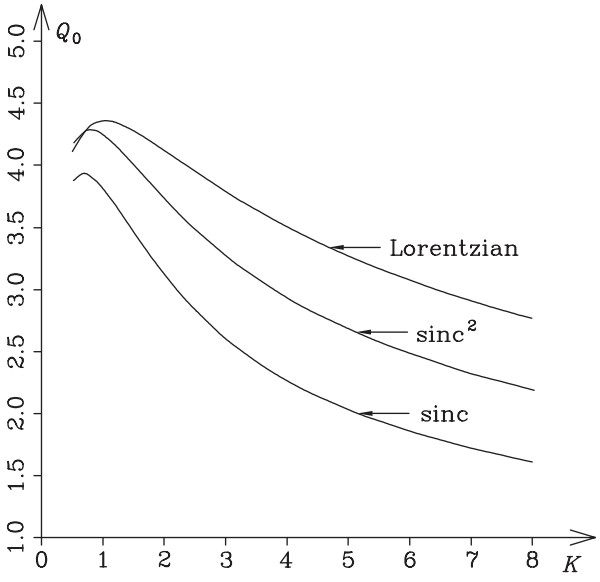


Figure 11.22: Signal-to-noise ratio enhancement Q_0 as a function of the smoothing parameter K in the cases where a Lorentzian spectrum is smoothed with a Lorentzian, sinc^2, or sinc smoothing function.

Figure 11.23 illustrates the relative efficiency of the three smoothing functions in the case where the maximum relative distortions $\Delta(0)$ are, in each case, -2 %. Theoretical values of the signal-to-noise ratio enhancement, given by Equations 11.67, 11.64, and 11.61 are 2.73, 3.31, and 6.93 for the Lorentzian, sinc^2, and sinc smoothing, respectively. In each case, the maximum relative distortions are the same, by there are large differences in the signal-to-noise ratio enhancement. If we wish to smooth with minimal distortions, the sinc smoothing gives the best result. On the other hand, a more profound examination would reveal that the shape of the relative distortion curve $\Delta(f)$ would be the most attractive in Lorentzian smoothing [5]. Figure 11.23 also demonstrates how smoothing takes place simultaneously for all spectral lines, in spite of their different positions and intensities.

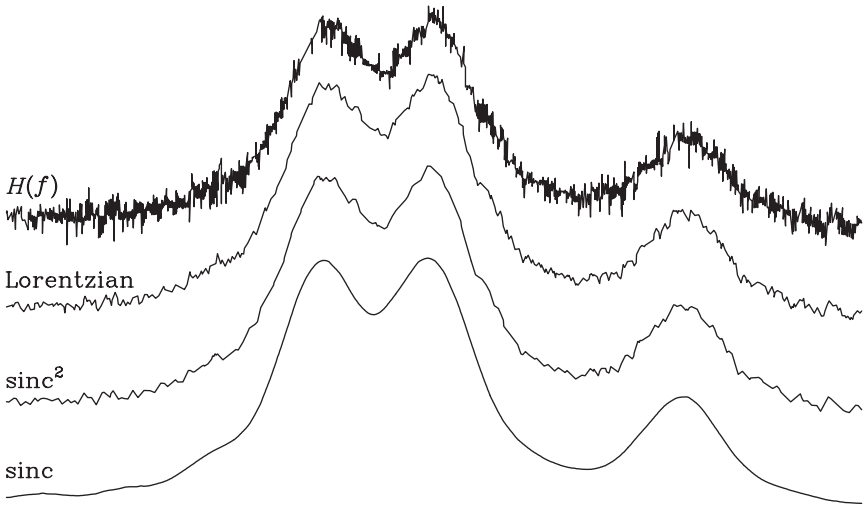


Figure 11.23: Illustration of the efficiency of Lorentzian, sinc^2, and sinc smoothing, applied to a spectrum $H(f)$ consisting of three Lorentzian lines with $K_0 = 100$. The K-values of the smoothing functions were chosen in such a way that the maximum relative distortion $\Delta(0)$ is -2 % in each case. The theoretical enhancement of the signal-to-noise ratio is 2.73, 3.31, and 6.93 for Lorentzian, sinc^2 and sinc smoothing, respectively. [5]

11.6 Elimination of a background

A spectrum $H(f)$ often contains a slowly varying background. In the signal $h(t)$ the background information is near $t = 0$. Consequently, the background can be eliminated by suppressing the beginning of the signal. This process is shown in Figure 11.24. If the FWHM of the background is Δf_0, then the background disappears, if the values of the signal are replaced by zero at $|t| < 1.21/(2\Delta f_0)$.

The removal of part of the signal causes a distortion of spectral lines. The broadest spectral lines experience the largest distortion. The maximum relative distortion of a spectral line with FWHM Δf, due to the shortening of the significant part of the signal, is approximately $\Delta f/(\Delta f_0)$. The relative distortion of narrow spectral lines is smaller than this.

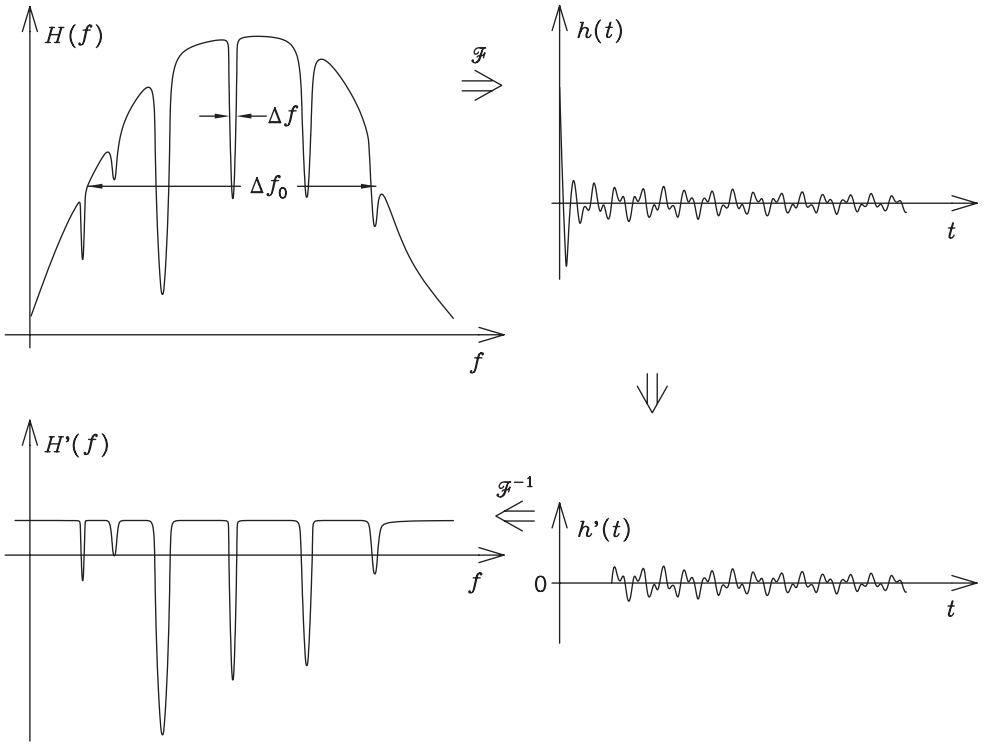


Figure 11.24: The background of the spectrum $H(f)$ may be eliminated by first taking the Fourier transform of the spectrum $\mathcal{F}\{H(f)\} = h(t)$, and then replacing the values of the beginning of the signal $h(t)$ by zero. The inverse Fourier transform of the manipulated signal $h'(t)$ gives a spectrum $H'(f)$ without the background: $\mathcal{F}^{-1}\{h'(t)\} = H'(f)$.

In practice, the distortion of the spectral lines due to the elimination of the background is often smaller than the noise of the spectrum. However, in some cases the background of the measured spectrum cannot be considered slowly varying. If the derivative of the background is large at some points, then the background cannot be eliminated from the whole spectrum. We may eliminate it part by part from those areas where it is slowly varying, and process the areas with high background derivative in some other way.

11.7 Elimination of an interference pattern

A spectrum sometimes contains an interference pattern. It may be caused by, for example, interference in filters or windows, which are frequently used in optical and IR instruments.

If a registered spectrum contains an interference pattern with a period F in f-domain, then the corresponding signal has a peak at $t = 1/F$ in t-domain. A practical way to eliminate the interference pattern is to take the Fourier transform of the spectrum, remove the peak at

$t = 1/F$ by replacing the values of the signal around $t = 1/F$ by the value zero, and return to f-domain by inverse Fourier transform. This procedure is demonstrated in Figure 11.25. The result of this mathematical procedure is a spectrum without the interference pattern.

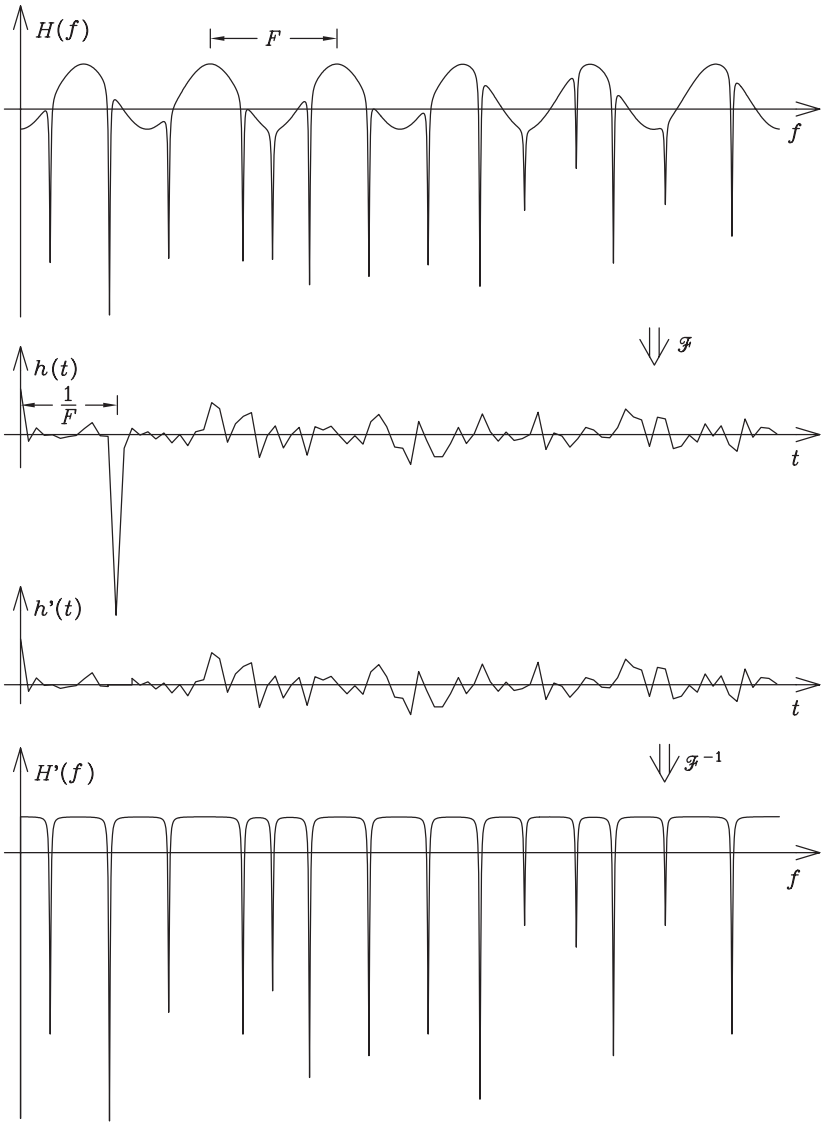


Figure 11.25: Elimination of an interference pattern. The spectrum $H(f)$, which contains an interference pattern of the period F, is Fourier transformed to the signal $h(t)$. The interference peak is removed by replacing the values of the signal around $t = 1/F$ by zero. The new signal $h'(t)$ is inverse Fourier transformed. The resulting spectrum $H'(f)$ contains no interference pattern.

If the width of the interfered area in the spectrum is Δf_0, as illustrated in Figure 11.26, then the width of the interference peak in the signal is approximately $1/\Delta f_0$. This is approximately the width of the signal area which should be removed. If the FWHM of the broadest true spectral line is Δf (see Figure 11.26), then its relative distortion due to the interference pattern removal is $\Delta f/\Delta f_0$. Narrower lines have smaller relative distortions.

To be precise, zeroing the bad samples does not eliminate background or interference completely, since the true, undistorted samples are not exactly zeroes. An alternative approach would be to regenerate the corrupted signal (in FTS, interferogram) samples by using linear prediction, which is treated in detail in Chapter 13.

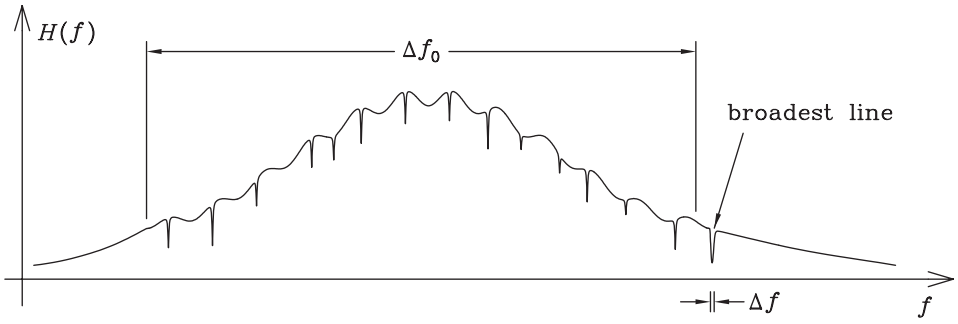


Figure 11.26: Determination of the relative distortion of a line in the removal of an interference pattern. If the width of the interference area in the spectrum is Δf_0, and the width of the broadest spectral line is Δf, then the relative distortion of the broadest line is $\Delta f/(\Delta f_0)$.

11.8 Deconvolution

Deconvolution is the inverse operation of convolution. Let us assume that an observed spectrum is

$$H_W(f) = W(f) * H(f) = \int_{-\infty}^{\infty} W(f')H(f - f')\,\mathrm{d}f', \tag{11.68}$$

where $H(f)$ is the real spectrum, and $W(f)$ an instrumental function, due to the measurement system. The true spectrum may be computed from the measured spectrum by deconvolution, if the instrumental function is known.

Deconvolution in the frequency domain is a very difficult operation. An easier way is to perform the operation in the time domain. If we take Fourier transform of both sides of Equation 11.68, and apply the convolution theorem, we obtain

$$\mathcal{F}\{H_W(f)\} = \mathcal{F}\{W(f)\}\mathcal{F}\{H(f)\}. \tag{11.69}$$

Consequently, the original spectrum can be expressed as

$$\boxed{H(f) = \mathcal{F}^{-1}\{\mathcal{F}\{H(f)\}\} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{H_W(f)\}}{\mathcal{F}\{W(f)\}}\right\},} \tag{11.70}$$

if $\mathcal{F}\{W(f)\} \neq 0$ at every value of t.

We can see that deconvolution is possible only if $\mathcal{F}\{W(f)\} \neq 0$ at every value of t. The validity of this condition is, however, difficult to observe in f-domain! In practice, the noise of the measured spectrum $H_W(f)$ will limit the use of deconvolution, since the noise will "explode" at the points where $\mathcal{F}\{W(t)\}$ is close to zero. This can be avoided by smoothing the spectrum. We can use the optimal smoothing, and multiply the signal in t-domain by a boxcar function $\Pi_{2T_i}(t)$ (Equation 11.35). As demonstrated in Figure 11.27, we can choose

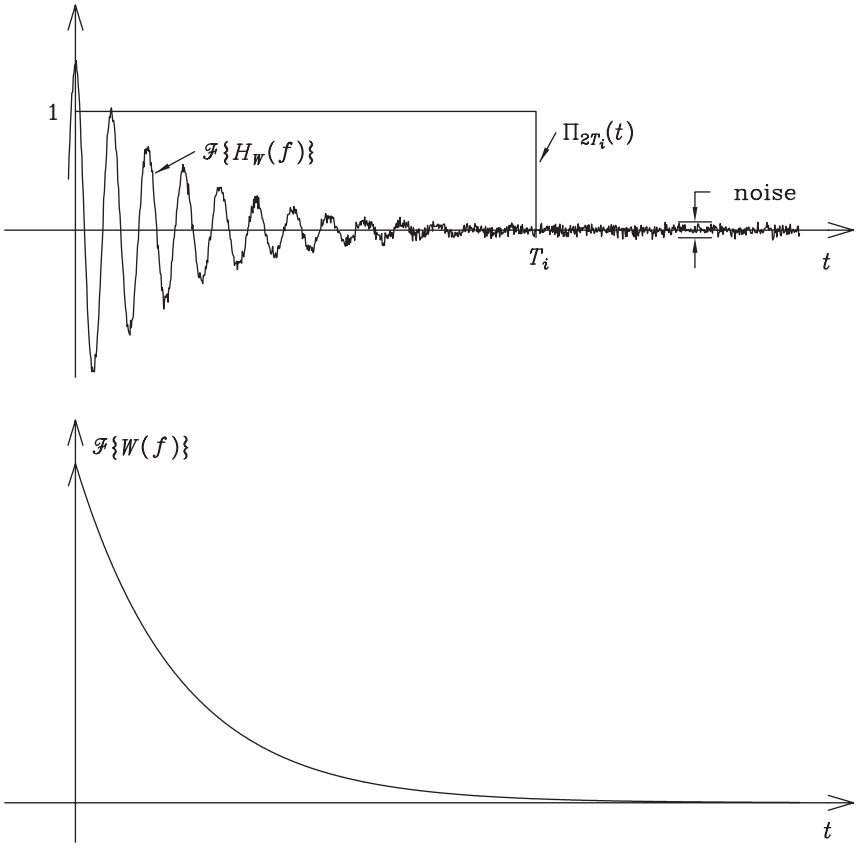


Figure 11.27: The Fourier transform of the measured spectrum $\mathcal{F}\{H_W(f)\}$, and the truncation boxcar function $\Pi_{2T_i}(t)$. Behind the point T_i the signal $\mathcal{F}\{H_W(f)\}$ is mostly noise. The lower curve is the Fourier transform of the instrumental function $\mathcal{F}\{W(f)\}$.

the point T_i in such a way that at this point the value of the signal $\approx \sqrt{\overline{n^2(t)}}$, that is, behind T_i the signal $\mathcal{F}\{W(f)\}$ is smaller than noise.

The deconvolution of the smoothed measured spectrum $H_W^{\mathrm{s}}(f)$ gives

$$H^{\mathrm{s}}(f) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{H_W^{\mathrm{s}}(f)\}}{\mathcal{F}\{W(f)\}}\right\} = \mathcal{F}^{-1}\left\{\frac{\Pi_{2T_i}(t)\mathcal{F}\{H_W(f)\}}{\mathcal{F}\{W(f)\}}\right\}, \tag{11.71}$$

because

$$H_W^{\mathrm{s}}(f) = \mathcal{F}^{-1}\{\Pi_{2T_i}(t)\} * H_W(f). \tag{11.72}$$

From Equations 11.71 and 11.70 we obtain

$$H^{\mathrm{s}}(f) = \mathcal{F}^{-1}\{\Pi_{2T_i}(t)\} * H(f) = 2T_i\,\mathrm{sinc}(\pi f T_i) * H(f). \tag{11.73}$$

If the Fourier transform of the instrumental function $\mathcal{F}\{W(f)\}$ equals zero behind a certain value of t (or at certain values of t), then the undistorted spectrum cannot be computed exactly by deconvolution, or by any other means, even if there would be no noise. This is natural, because in the area where $\mathcal{F}\{W(f)\} = 0$ the undistorted signal $\mathcal{F}\{H(f)\}$ is multiplied by zero and is permanently lost.

Let us examine, as an example, a grating. The instrumental function or the convolving function $W(\nu)$ of a grating, in the wavenumber domain, is (remember Equation 9.21)

$$W(\nu) \propto L\,\mathrm{sinc}^2(\pi\nu L), \tag{11.74}$$

where

$$L = Nd\,\sin\theta. \tag{11.75}$$

Here N is the number of slits in the grating, d is the distance of the slits, and θ is the angle of the diffracted wave. If the true spectrum is $H(\nu)$, then the measured spectrum is the convolution $L\,\mathrm{sinc}^2(\pi\nu L) * H(\nu)$. The Fourier transform of this is the measured signal $\mathcal{F}\{L\,\mathrm{sinc}^2(\pi\nu L)\}\mathcal{F}\{H(\nu)\}$. On the other hand, we know that $\mathcal{F}\{L\,\mathrm{sinc}^2(\pi\nu L)\}$ is the triangular function shown in Figure 11.28. Consequently, a grating loses information $h(x)$ in the area where $|x| > Nd\,\sin\theta$.

Generally, the aim of deconvolution is to compute the true spectrum $H(f)$ from the measured spectrum $H_W(f)$, that is, to correct the distortion caused by the instrumental function $W(f)$, which generally is known. This is usually successful, especially, if the instrumental function is much narrower than the spectral lines. This situation is illustrated in Figure 11.29.

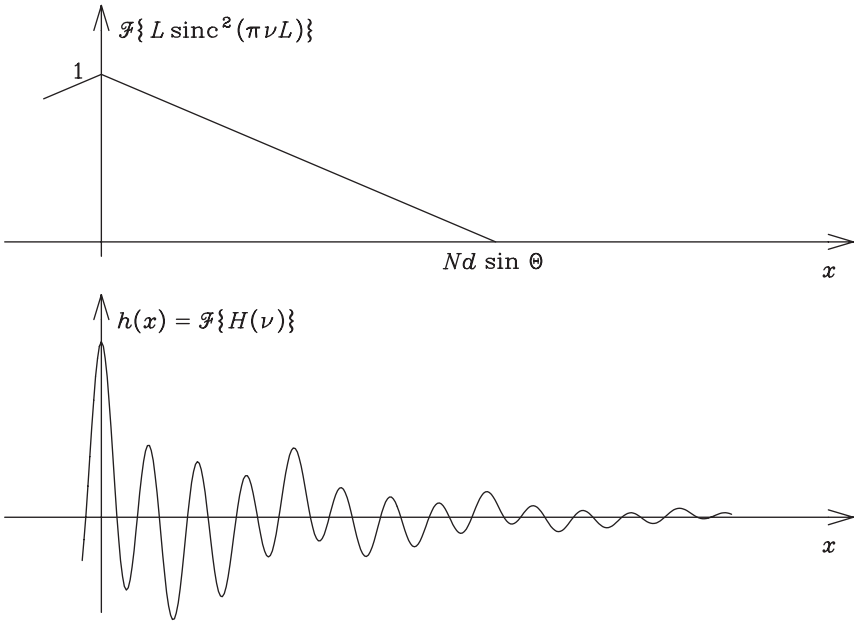


Figure 11.28: The Fourier transform of the instrumental function $\mathcal{F}\{L\,\mathrm{sinc}^2(\pi\nu L)\}$, and the true signal $h(x) = \mathcal{F}\{H(\nu)\}$. The measured signal is the product of these two functions.

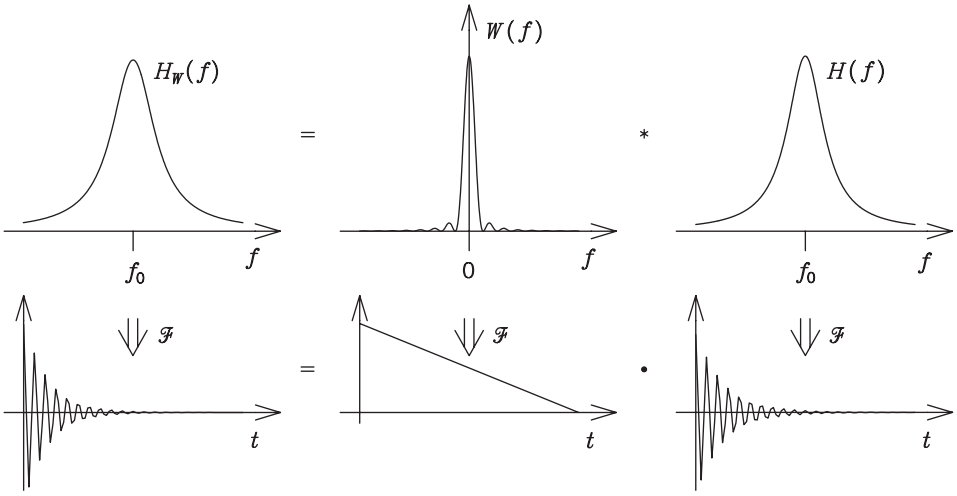


Figure 11.29: If the instrumental function $W(f)$ has a small FWHM, then it distorts the spectral lines of the true spectrum $H(f)$ only slightly, and the measured spectrum $H_W(f) \approx H(f)$.

Problems

1. The discrete Fourier transformation of the vector $\boldsymbol{f} = (f_0, \ldots, f_{2N-1})$ gives the transform vector $\boldsymbol{F} = (F_0, \ldots, F_{2N-1})$. Let us assume that the number of data of $\boldsymbol{f}$ is increased k-fold in the following way: $2kN - 2N - 1$ zero components are added in the middle of the original vector, behind the datum N, and the value of the datum N is divided into two equal parts and given to datum N and datum $2kN - N$ of the new vector. The new vector is

$$\boldsymbol{f}' = \left(f_0, \ldots, f_{N-1}, \frac{1}{2} f_N, 0, \ldots, 0, \frac{1}{2} f_N, f_{N+1}, \ldots, f_{2N-1} \right).$$

   Let us denote its Fourier transform vector $\boldsymbol{F}'$.

   Show that $F'_{kj} = F_j$, $j = 0, \ldots, 2N - 1$. (This means that the data of the vector $\boldsymbol{F}$ are not changed, but $k - 1$ new data are interpolated in each data interval.)

2. The spectrum of a real signal under examination differs from zero only in the frequency band [1000 Hz, 1600 Hz]. The spectrum does, however, contain noise up to the frequency of 4000 Hz. This noise will be filtered by discrete convolution in the signal domain.

   (a) How large can we let the convolution sampling interval Δt be before high-frequency noise will be aliased on the signal?

   (b) The spectrum is calculated by Fourier transforming the convolved signal. What is the largest value of Δt which gives an undistorted spectrum? (The spectrum is allowed to be mirrored.)

3. A noisy signal is smoothed by discrete convolution in such a way that a datum h_k is replaced by the average of the data $h_{k-n}, \ldots, h_{k+n}$. Compute the corresponding transfer function (filter function) by directly calculating the discrete Fourier transform of the impulse response. The number of data of the transformation is M.

4. Derive the transfer function requested in the previous problem in an alternative way: first calculate the continuous Fourier transform of a continuous impulse response and then make a discontinuity correction, *i.e.*, take into account aliasing due to sampling.

   Hint: Use Eq. 9.22.

5. A signal is smoothed by discrete convolution in such a way that a smoothed datum is

$$h'_j = \frac{1}{2} h_{j-2} - \frac{1}{2} h_{j-1} + h_j - \frac{1}{2} h_{j+1} + \frac{1}{2} h_{j+2}.$$

   Find the discrete transfer function (vector) $\boldsymbol{G}$, by which the discrete Fourier transform of $\boldsymbol{h}$ is multiplied. Outline the function $\boldsymbol{G}$ in the interval $[-M/2, M/2]$, where M is the number of data of $\boldsymbol{G}$.

6. A noisy signal vector $\boldsymbol{s}$ is smoothed by discrete convolution in such a way that a datum s_k is replaced by the datum

$$s'_k = \Delta t \sum_{j=-\infty}^{\infty} g_j s_{k-j},$$

where $g_j = F\,\mathrm{sinc}^2(\pi F j \Delta t)$ and $\dfrac{1}{2\Delta t} < F < \dfrac{1}{\Delta t}$. Draw the corresponding transfer function.

Hint: First determine the continuous Fourier transform of a continuous impulse response and then make the discontinuity correction.

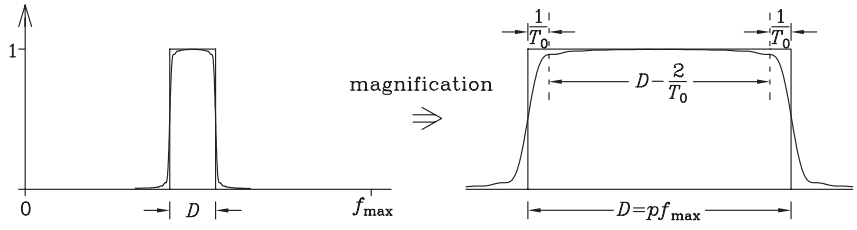
7. The discrete smoothing of a signal is made according to

$$h'_k = \frac{1}{3} h_k + \frac{1}{3} h_{k-1} + \frac{1}{3} h_{k-2}.$$

Determine the relative damping of the amplitude of a signal in this filter at the frequency $f = f_{\max}/2 = 1/(4\Delta t)$.

8. Samples of a signal are taken in the intervals Δt in such a way that the highest frequency which is not aliased is $1/(2\Delta t) = f_{\max}$. We wish to filter by discrete convolution a spectral band of width D which satisfies $D/f_{\max} = p$, $0 < p < 1$ (band-pass filtering). The convolution is truncated by a triangular weight function which goes to zero at the points $\pm T_0 = \pm k\Delta t$ and encloses $m = 2k - 1$ non-zero impulse response data.

Determine, as a function of p and the length of the convolution m, how large part of the band, *i.e.*, $\dfrac{D - 2/T_0}{D}$, is not seriously distorted (see the following picture).



9. A female radio journalist whose voice contains frequencies 4 kHz–10 kHz is making a radio program about summer in Finland. The sounds of her surroundings contain hum of wind in trees (0 Hz–1 kHz) and whine of mosquitoes (9 kHz–16 kHz).

   (a) What is the maximum sampling interval Δt which gives an undistorted signal?

   (b) The annoying lisp of drunks (1 kHz–4 kHz) is filtered away by convolving the signal in (almost) real time. Find the convolution function.

   (c) The convolution integral is truncated at the points $\pm T_0$ (boxcar truncation). What is the minimum number of data in the convolution in order to keep the width $1/T_0$ of the slope of the filter function under the maximum of 100 Hz? Use the value of Δt obtained in (a).

10. Show that if two Lorentz functions are convolved with each other, and the FWHMs of these functions are $2\sigma_1$ and $2\sigma_2$, then the resulting function is a Lorentz function, and its FWHM is $2\sigma_1 + 2\sigma_2$.

11. A Lorentzian-shaped spectral line is smoothed by a sinc function. The FWHM of this sinc function is half of the FWHM of the Lorentzian line. What is the maximum relative distortion of the line?

12. Show that if a Lorentzian line is smoothed by a sinc^2 function, then the maximum relative distortion is (Equation 11.63)

$$\Delta(0) \approx \frac{2}{1.7718\pi K}\left[\exp\left(-\frac{1.7718\pi K}{2}\right) - 1\right],$$

where K is the smoothing parameter.

Hint: The FWHM of the function $T\,\text{sinc}^2(\pi T f)$ is approximately $1.7718/(2T)$.

13. A Lorentzian-shaped spectral line is smoothed by a sinc^2 function. The FWHM of this sinc^2 function is half of the FWHM of the Lorentzian line. What is the maximum relative distortion of the line?

14. A Lorentzian-shaped spectral line is smoothed by a sinc function. The FWHM of this sinc function is half of the FWHM of the Lorentzian line. What is the signal-to-noise ratio enhancement in the smoothing, if the smoothing parameter of truncation is $K_0 = 20$.

15. A Lorentzian line is smoothed by a sinc^2 function. The original signal is registered in the interval $[-T_0, T_0]$ (boxcar truncation). Show that the signal-to-noise ratio enhancement in the smoothing is (Equation 11.64)

$$Q_0 \approx \sqrt{\frac{3 \times 1.2067 K_0}{1.7718 K}}\,[1 + \Delta(0)]\,,$$

where K and K_0 are the smoothing parameters, and $\Delta(0)$ is the maximum relative distortion of the line shape.

Hint: The FWHM of the function $T\,\text{sinc}^2(\pi T f)$ is approximately $1.7718/(2T)$. The FWHM of the function $2T_0\,\text{sinc}(\pi 2T_0 f)$ is approximately $1.2067/(2T_0)$. You can make the approximation $e^{-\pi K_0/2} \approx 0$.

16. The height of a Lorentzian spectral line is $H_0(0) = S$. The FWHM of the line is 2σ. The signal was in the recording multiplied by a boxcar in the interval $[-T_0, T_0]$. The spectrum also contains white noise of standard deviation N, so that the original signal-to-noise ratio is S/N. The line is smoothed by the function $W(f) = T\,\text{sinc}^2(\pi T f)$, so that the signal-to-noise ratio is changed by the factor

$$Q_0 \approx \sqrt{\frac{3 \times 1.2067 K_0}{1.7718 K}}\,[1 + \Delta(0)]\,,$$

where

$$\Delta(0) \approx \frac{2}{1.7718\pi K}\left[\exp\left(-\frac{1.7718\pi K}{2}\right) - 1\right]$$

is the maximum relative distortion of the line shape.

(a) Find the standard deviation N' of noise after smoothing as a function of N, T, and T_0.

(b) Let us choose the point T in such a way that the standard deviation N' of the smoothed noise is equal to the distortion of the height of the line. Show that in this case T is determined by equation

$$\frac{1}{1 - e^{-2\pi\sigma T}} = \frac{(3T_0)^{1/2}}{2\pi\sigma T^{3/2}} \frac{S}{N}.$$

17. A Lorentzian spectral line of FWHM 10 cm^{-1} was recorded by a Fourier transform spectrometer. The resolution was 0.1 cm^{-1} and the signal-to-noise ratio was 500. What would the signal-to-noise ratio have been if the resolution had been 1 cm^{-1}? We assume that the spectrometer has a point-like light source.

18. A spectrum which consists of Lorentzian spectral lines is smoothed. In recording, the signal has been truncated by a boxcar in $[-T_0, T_0]$. The smoothing is made by truncating the signal by a shorter boxcar $[-T, T]$. The optimal truncation point T_e where the distortion of the height of a line is equal to the standard deviation of the smoothed noise is, approximately, given by the equation (derived in Example 11.5)

$$T_e \approx \frac{1}{2\pi\sigma} \ln\left[\sqrt{\frac{T_0}{T_e}}\,(S/N)\right],$$

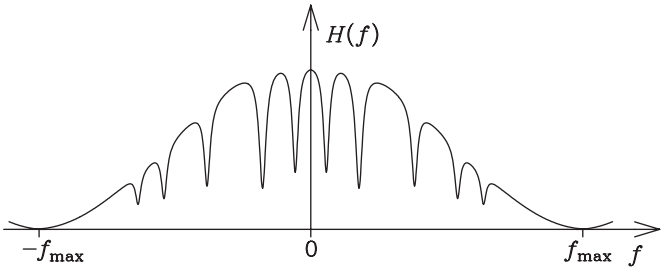
where (S/N) is the signal-to-noise ratio of the unsmoothed spectrum.

Show that if the signal-to-noise ratio fulfills the condition $(S/N) \gg \sqrt{T_0/T_e}$, then the optimal smoothing gives the rule $K_{\text{opt}} \gtrapprox \log_{10}(S/N)$ (*i.e.*, the value of the optimal smoothing parameter is slightly larger than $\log_{10}(S/N)$).

19. In the measurement of a weak cosine wave, the standard deviation of white noise, or the noise rms-value σ_t, is ten-fold compared to the amplitude of the cosine wave. How many signal samples should we register in order to obtain a signal-to-noise ratio 10, *i.e.*, in order that the spectral line of the signal would be ten times as high as the standard deviation σ_f of the noise in the spectral domain?

Hint: The noises in the signal and the spectrum are connected by the discrete Parseval's theorem $\Delta t \sum_j |n_j|^2 = \Delta f \sum_k |N_k|^2$, or $\Delta t \sigma_t^2 = \Delta f \sigma_f^2$.

20. The spectrum of the following picture has a slowly changing background, which is of the form $1+\cos(\pi f/f_{\max})$. What is the minimum number of points which must be given the value zero in the t-domain in order to remove the background?



21. A Fourier transform spectrometer has an optical filter of thickness d and refractive index in the interesting wavenumber region n. Reflections at the surfaces of the filter produce in the spectrum a cosine wave which has a maximum at those wavenumbers at which the beam which has been reflected back and forth has constructive interference. At which value of the optical path difference x does the signal contain a disturbance?

22. Show that the area of the convolution of two functions f and g is the product of the areas of the functions f and g. Also show that the area of the deconvolution of two functions is the ratio of the areas of the two functions.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

12 Fourier self-deconvolution (FSD)

12.1 Principle of FSD

Narrowing a spectral line in the spectral domain is equivalent to stretching the corresponding signal, the decaying wave, in the signal domain (see also Chapter 13). Making the decaying wave decay more slowly in t-domain results in a narrower line in f-domain. This operation can be done by dividing the wave by some smoothly decaying curve. In the spectral domain this operation appears as deconvolution. In the same way as convolution always broadens spectral lines, its inverse operation, deconvolution, always narrows the lines.

In *Fourier self-deconvolution*, FSD, the spectrum is deconvolved by the line shape of the spectrum itself. The line shape is a line situated in the origin, and it has an area equal to one. In FSD, the signal in t-domain is divided by the Fourier transform of the line shape function. A return to f-domain by inverse Fourier transform gives a spectrum with narrower spectral lines. The goal of FSD is to enhance the spectral resolution, that is, make the spectral lines narrower, by mathematical means, in such a way that the frequency and the integrated intensity of each line are preserved.

Let us assume that we recorded the line shape $W_{\mathrm{R}}(f)$. We remember that we can write $W_{\mathrm{R}}(f) = W_{\mathrm{R}}(f) * \delta(f)$. If we carry out deconvolution of $W_{\mathrm{R}}(f)$ by the line shape $W_{\mathrm{R}}(f)$ itself, Equation 11.70 gives

$$\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{W_{\mathrm{R}}(f)\}}{\mathcal{F}\{W_{\mathrm{R}}(f)\}}\right\} = \mathcal{F}^{-1}\{1\} = \delta(f),$$

and the result is Dirac's delta function. Total removal of the original line shape results in infinitely narrow lines. In practice, however, we obtain a sinc function, which originates from the inevitable truncation of the now non-decaying wave. In addition, the noise of the original spectrum as well as the highly oscillating sinc line require smoothing. The spectrum should be apodized by some window function in order to make the spectral lines smoother.

Let us assume that a registered spectrum $H_{\mathrm{R}}(f)$ consists of lines $A_i W_{\mathrm{R}}(f - f_i)$:

$$H_{\mathrm{R}}(f) = \sum_i A_i W_{\mathrm{R}}(f - f_i), \tag{12.1}$$

where A_i is the area (intensity), and f_i the position of the ith line. $W_{\mathrm{R}}(f)$ is the line shape function of the registered spectrum. The area of $W_{\mathrm{R}}(f)$ is equal to one. We are assuming that all the lines have this same line shape. The Fourier self-deconvolved spectrum $H_{\mathrm{FSD}}(f)$ can

then be expressed as

$$\boxed{H_{\mathrm{FSD}}(f) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{W(f)\}\mathcal{F}\{H_{\mathrm{R}}(f)\}}{\mathcal{F}\{W_{\mathrm{R}}(f)\}}\right\},} \tag{12.2}$$

where $W(f)$ is the smoothing function, which is also the desired new line shape. The area of $W(f)$ is equal to one. In t-domain, $\mathcal{F}\{W(f)\}$ is the window function by which the signal is apodized in order to smooth the spectrum.

Since deconvolution is equivalent to division in t-domain, it is clearly a linear operation. Deconvolving a linear combination of spectral lines of various heights simultaneously gives the same result as deconvolving each line separately, providing that they have a common line shape. The relative heights of the lines remain unchanged. Since the shift theorem states that shifting a line in f-domain corresponds to multiplication by an exponential wave in t-domain, the positions of the lines may also be arbitrary. FSD is able to narrow an overlapping set of an unknown number of spectral lines without the need to know their heights and positions.

Since the operation is linear, we can examine only one spectral line. A registered spectrum with one spectral line can be written as

$$H_{\mathrm{R}}(f) = A_0 W_{\mathrm{R}}(f - f_0). \tag{12.3}$$

By applying the shift theorem, we obtain that the new spectrum after FSD is

$$\begin{aligned} H_{\mathrm{FSD}}(f) &= \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{W(f)\}\mathcal{F}\{A_0 W_{\mathrm{R}}(f - f_0)\}}{\mathcal{F}\{W_{\mathrm{R}}(f)\}}\right\} \\ &\overset{\dagger}{=} \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{W(f)\}A_0\mathcal{F}\{W_{\mathrm{R}}(f)\}e^{i2\pi f_0 t}}{\mathcal{F}\{W_{\mathrm{R}}(f)\}}\right\} \\ &\overset{\dagger}{=} \mathcal{F}^{-1}\{\mathcal{F}\{A_0 W(f - f_0)\}\} = A_0 W(f - f_0). \end{aligned} \tag{12.4}$$

($\dagger \Leftrightarrow$ shift theorem.) We can see that FSD is, indeed, the operation

$$A_0 W_{\mathrm{R}}(f - f_0) \rightarrow A_0 W(f - f_0). \tag{12.5}$$

A registered spectral line of the shape $W_{\mathrm{R}}(f)$ is turned into a narrower spectral line of the shape $W(f)$, the area and the position remaining unchanged.

The principle of the FSD procedure is illustrated in Figure 12.1. The registered line is deconvolved by its own line shape by division by $\mathcal{F}\{W_{\mathrm{R}}(f)\}$ in t-domain, resulting, in the absence of truncation, in Dirac's delta function $A_0\delta(f - f_0)$. This Dirac's delta function is smoothed by multiplication by $\mathcal{F}\{W(f)\}$ in t-domain, resulting in the final line shape $W(f)$.

We must, however, remember, that in practice the line shape $W_{\mathrm{R}}(f)$ of a registered spectrum is not the true original line shape, but the convolution of the instrumental function of the measurement system $W_{\mathrm{inst}}(f)$ and the true line shape $W_{\mathrm{true}}(f)$, as illustrated in Figure 12.2.

The main goal of FSD is to reduce the half-widths of the spectral lines. The efficiency of this is given by the *resolution enhancement factor*

$$K = \frac{\sigma_{\mathrm{R}}}{\sigma} = \frac{\text{FWHM of the registered line}}{\text{FWHM of the smoothing function}}. \tag{12.6}$$

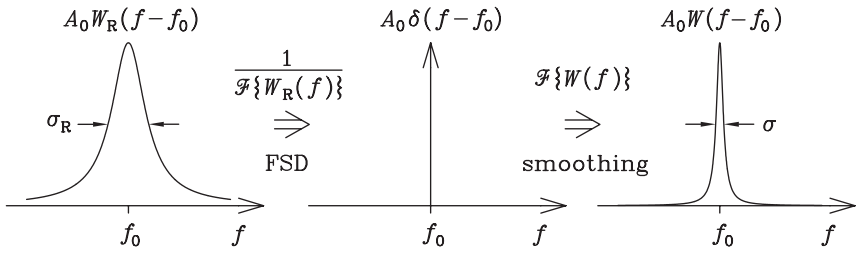


Figure 12.1: The procedure of FSD. The registered spectral line $A_0W_R(f - f_0)$ is Fourier self-deconvolved, resulting in Dirac's delta peak $A_0\delta(f - f_0)$, and then smoothed, resulting in the final line $A_0W(f - f_0)$.

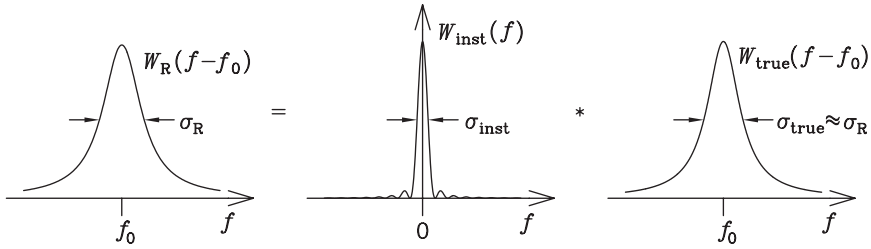


Figure 12.2: The line shape $W_R(f)$ of a registered high-resolution spectrum $H_R(f)$ is the convolution of the instrumental function of the measurement system $W_{inst}(f)$ and the true line shape $W_{true}(f)$.

This may be simply called the K-value.

The theoretical resolution (see Equation 6.31) cannot be exceeded by FSD. This sets an upper limit to the K-value. If the recorded signal is

$$H_R(f) = W_{inst}(f) * H_{true}(f), \tag{12.7}$$

where $W_{inst}(f)$ is an instrumental function, then

$$K \leq K_0 = \frac{\sigma_R}{\sigma_{inst}} = \frac{\text{FWHM of the registered line}}{\text{FWHM of the instrumental function}}. \tag{12.8}$$

Let us now examine examples of FSD. In the next two examples, as well as in a few other examples later, FSD is applied to Fourier transform infrared spectroscopy. Conventionally, the spectrum is then expressed as $E(\nu)$ as a function of the wavenumber ν $\left[\text{cm}^{-1}\right]$, and the signal is $I(x)$ as a function of the optical path difference x [cm]. However, we shall, for consistency, denote the spectrum and the signal with the more general symbols $H(f)$ and $h(t)$, respectively. The phenomena demonstrated in the examples are, after all, generally valid.

Figure 12.3 shows two simulations of FSD, where the spectra contain no noise, and the instrumental resolution is assumed to be infinite. Two simulated band contours $H_R(f)$ are each composed of two Lorentzian lines of half-widths 8 cm^{-1}. In figure (a) the lines are 6 cm^{-1} apart and in figure (b) 2 cm^{-1} apart. The proportion of the intensities of the lines is 0.75. FSD is applied to both spectra. The window function which is used in apodization (smoothing) is

$$\mathcal{F}\{W(f)\} = [1 - (|t|/T)]^2 , \tag{12.9}$$

where $T = 0.8$ cm. The achieved spectral resolution enhancement is

$$K = \frac{\sigma_R}{\sigma} = \frac{8\,\mathrm{cm}^{-1}}{1.5\,\mathrm{cm}^{-1}} = 5.3. \tag{12.10}$$

We can see that in both cases the frequencies of the component lines are recovered by FSD. Even in figure (b) the relative intensities of the lines are recovered.

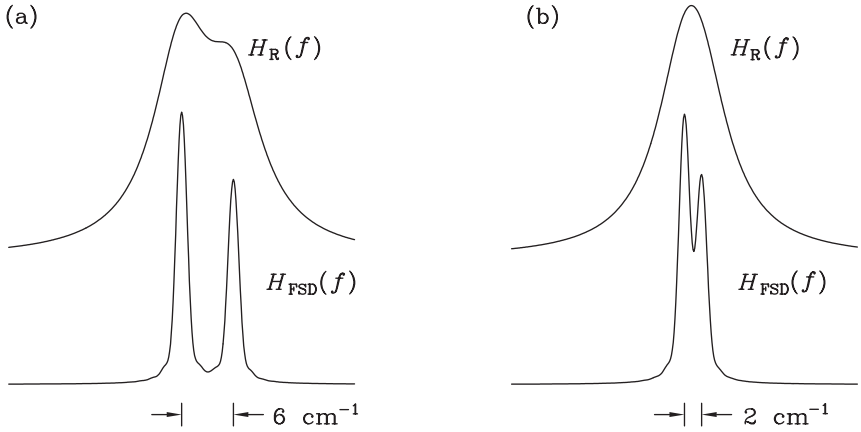


Figure 12.3: A simulated example of FSD. The spectra contain no noise, and the instrumental resolution is assumed to be infinite. Two simulated band contours $H_R(f)$ are each composed of two Lorentzian lines of half-widths 8 cm^{-1}. In figure (a) the lines are 6 cm^{-1} apart and in figure (b) 2 cm^{-1} apart. The proportion of the intensities of the lines is 0.75. FSD enhances the spectral resolution in both cases by a factor of 5.3, and the new FWHM is 1.5 cm^{-1}.

Figure 12.4(a) shows a real, experimental spectrum of liquid chlorobenzene. It is measured with an instrumental resolution of $\sigma_{inst} = 1.2$ cm^{-1}, and the signal-to-noise-ratio is 3000. The FWHM of the spectral lines due to the Brownian motion of liquid is approximately 9.8 cm^{-1}, and the line shape is assumed to be Lorentzian. The spectrum is Fourier self-deconvolved, using the smoothing function

$$W(f) = 16\sqrt{\frac{\pi}{2}}\,\frac{T}{(2\pi f T)^{5/2}}\,J_{5/2}(2\pi f T), \tag{12.11}$$

where $J_{5/2}$ is the Bessel function of the order $5/2$. Figures 12.4(b) and (c) show the spectrum after FSD, where we have chosen $K = 3.6$ and $K = 4$, respectively.

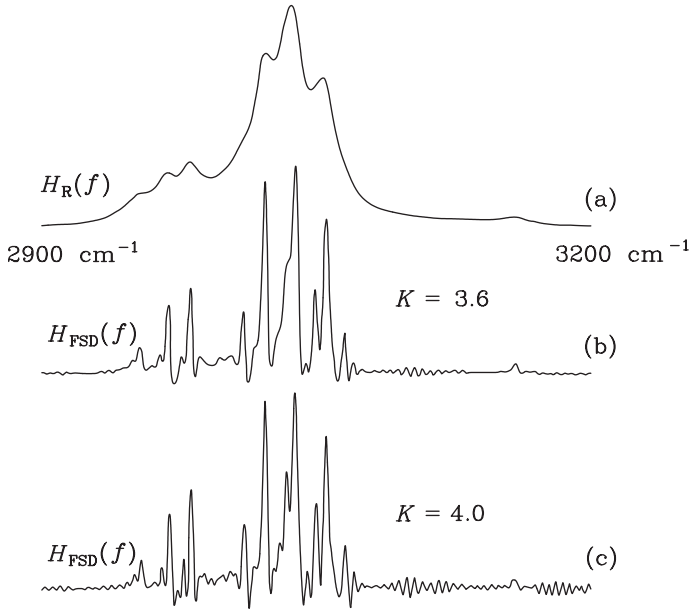


Figure 12.4: Fourier self-deconvolution of the C–H stretching region of the spectrum of liquid chlorobenzene, consisting of 9.8 cm^{-1}-wide Lorentzian lines. (a) Experimental spectrum with an instrumental resolution of 1.2 cm^{-1} and a signal-to-noise ratio (S/N) of about 3000. (b) Fourier self-deconvolved spectrum with a Bessel-type smoothing function and $K = 3.6$. (c) Fourier self-deconvolved spectrum with a Bessel-type smoothing function and $K = 4$.

Figure 12.5 demonstrates stepwise the procedure of FSD in f- and t-domains. The behaviour of random white noise is shown separately. If we know the line shape $W_R(f)$ of the recorded spectrum $H_R(f)$, the only option which may be freely chosen in the FSD procedure is the desired line shape $W(f)$. The signal-to-noise ratio of the Fourier self-deconvolved spectrum depends strongly on the shape and the half-width of $W(f)$. This is easy to understand by examining the behavior of noise in Figure 12.5. Consequently, the correct choice of $W(f)$ is very important in FSD.

Figure 12.6 shows eight different window functions $A(t) = \mathcal{F}\{W(f)\}$ and the corresponding smoothing functions (new line shapes) $W(f)$, which may be used in FSD. Figure 12.6 also lists the FWHM of the new line shapes, as well as the relative magnitudes of the strongest sidelobes.

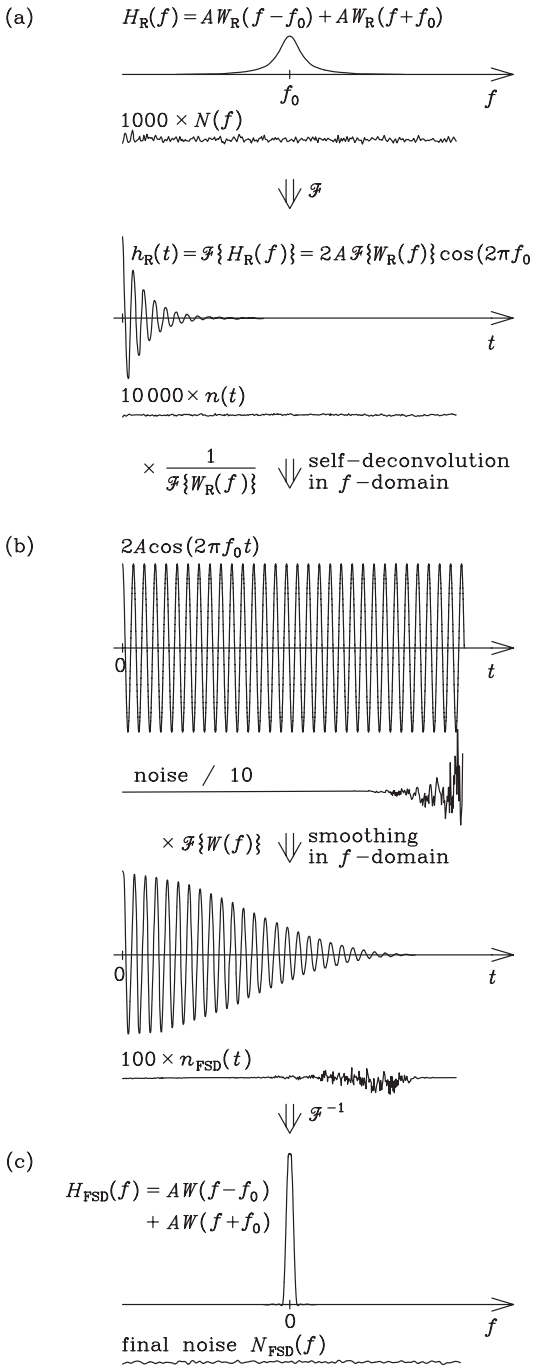


Figure 12.5: Illustration of the various steps (a) – (c) of the Fourier self-deconvolution (FSD) procedure.

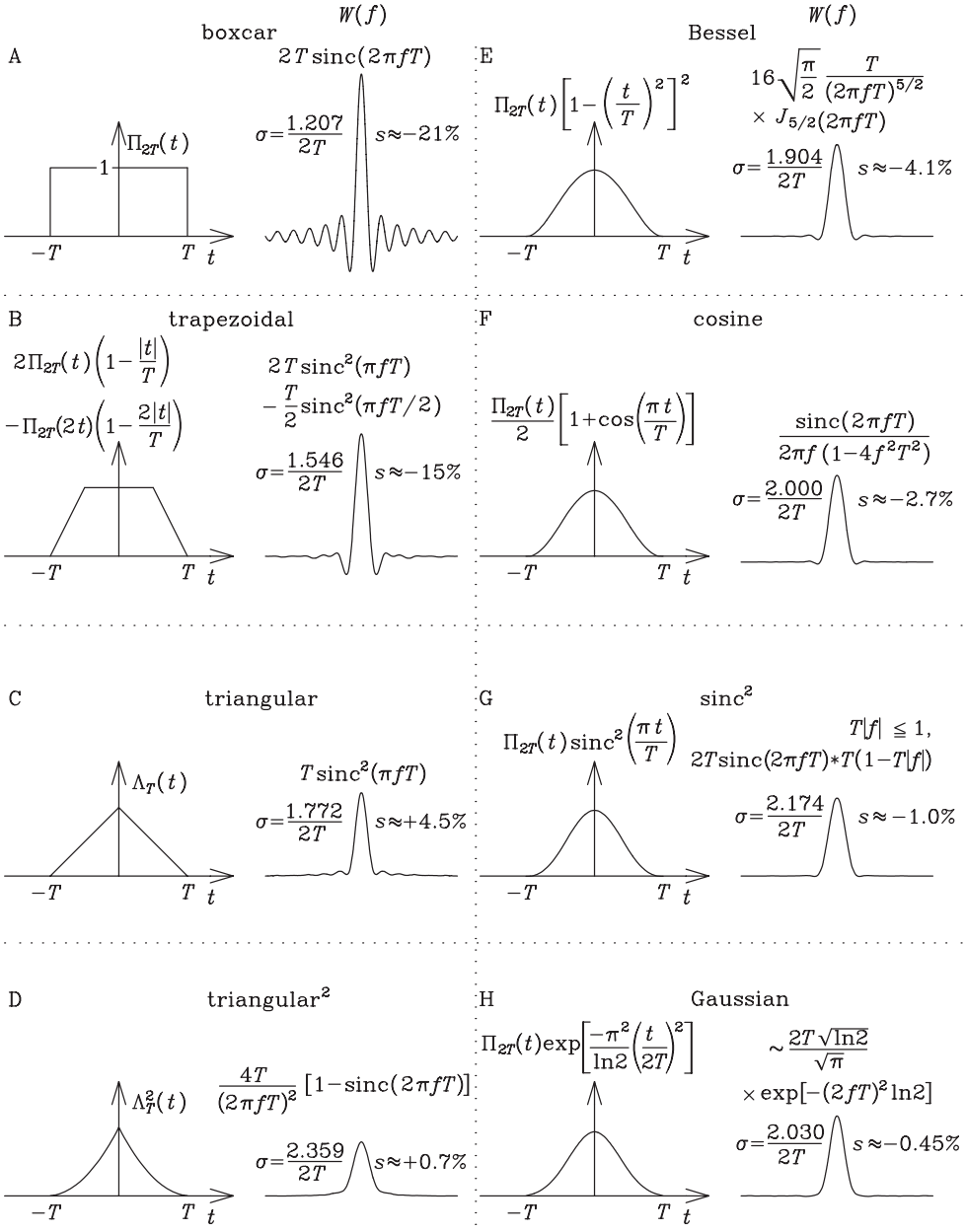


Figure 12.6: Various $A(t) = \mathcal{F}\{W(f)\}$ functions A – H (left-hand column) and the corresponding line shape functions $W(f)$ (right-hand column) used in Fourier self-deconvolution. The half-width σ of the line shape function and the relative magnitude s [%] of the strongest sidelobe of the line shape function are listed in the right-hand column.

12.2 Signal-to-noise ratio in FSD

The signal-to-noise ratio $(S/N)_{\mathrm{FSD}}$ of the Fourier self-deconvolved spectrum $H_{\mathrm{FSD}}(f)$ largely determines the usefulness of an FSD operation. The general formula of the enhancement, or alternatively, decrease, of the signal-to-noise ratio by FSD is obtained from Equation 11.55 by replacing $A(t)$ by $A(t)/[A_0(t)I_0(t)]$, where $A(t) = \mathcal{F}\{W(f)\}$, as in smoothing, until the point $T = T_0$, which defines the region where the signal is registered. This yields

$$Q = \frac{(S/N)_{\mathrm{FSD}}}{(S/N)_{\mathrm{R}}} = \frac{\int_{-T_0}^{T_0} A(t)\,\mathrm{d}t \sqrt{\int_{-\infty}^{\infty} A_0^2(t)\,\mathrm{d}t}}{\int_{-\infty}^{\infty} A_0(t)I_0(t)\,\mathrm{d}t \sqrt{\int_{-T_0}^{T_0} A^2(t)/I_0^2(t)\,\mathrm{d}t}}. \tag{12.12}$$

If $A_0(t)$ is the boxcar function

$$A_0(t) = \begin{cases} 0, & |t| > T_0, \\ 1, & |t| \le T_0, \end{cases} \tag{12.13}$$

then the signal-to-noise ratio enhancement, or decrease, by FSD is

$$Q_0 = \frac{\int_{-T_0}^{T_0} A(t)\,\mathrm{d}t \sqrt{T_0}}{\sqrt{\int_0^{T_0} A^2(t) I_0^{-2}(t)\,\mathrm{d}t} \int_{-T_0}^{T_0} I_0(t)\,\mathrm{d}t}. \tag{12.14}$$

The value of Q_0 depends on the choice of the shape of smoothing function and on the choice of the resolution enhancement factor. Q_0 is often written as a function of the resolution enhancement factor K and the maximum K-value K_0: $Q_0(K, K_0)$.

Let us consider the Lorentzian line shape

$$W_0(f) = \frac{\sigma_0/\pi}{\sigma_0^2 + f^2}. \tag{12.15}$$

Its Fourier transform is exponential:

$$I_0(t) = \mathcal{F}\{W_0(f)\} = e^{-2\pi\sigma_0|t|}. \tag{12.16}$$

Since the signal is truncated, the registered line shape is

$$W_{\mathrm{R}}(f) = [2T_0\,\mathrm{sinc}(2\pi T_0 f)] * \frac{\sigma_0/\pi}{\sigma_0^2 + f^2}. \tag{12.17}$$

If we further assume that the instrumental resolution is much higher than the spectral resolution, that is, $\sigma_{\mathrm{inst}} \ll \sigma_0$, then the registered spectral resolution $\sigma_{\mathrm{R}} \approx \sigma_0$. In this case, the maximum resolution enhancement factor is

$$K_0 = \frac{\sigma_{\mathrm{R}}}{\sigma_{\mathrm{inst}}} \approx \frac{2\sigma_0}{\sigma_{\mathrm{inst}}} = \frac{2\sigma_0}{1.207/(2T_0)}. \tag{12.18}$$

Then, Equation 12.14 gives, if T_0 is large,

$$Q_0(K, K_0) \approx \frac{\pi\sqrt{1.207K_0\sigma_0}\int_0^\infty A(t)\,\mathrm{d}t}{\sqrt{\int_0^\infty A^2(t)e^{4\pi\sigma_0 t}\,\mathrm{d}t}}. \tag{12.19}$$

Figure 12.7 shows plots of computed $Q_0(K, K_0)$ values as a function of K. In this example, the line shape is Lorentzian and $K_0 = 5$. The K-curves are shown for the eight different

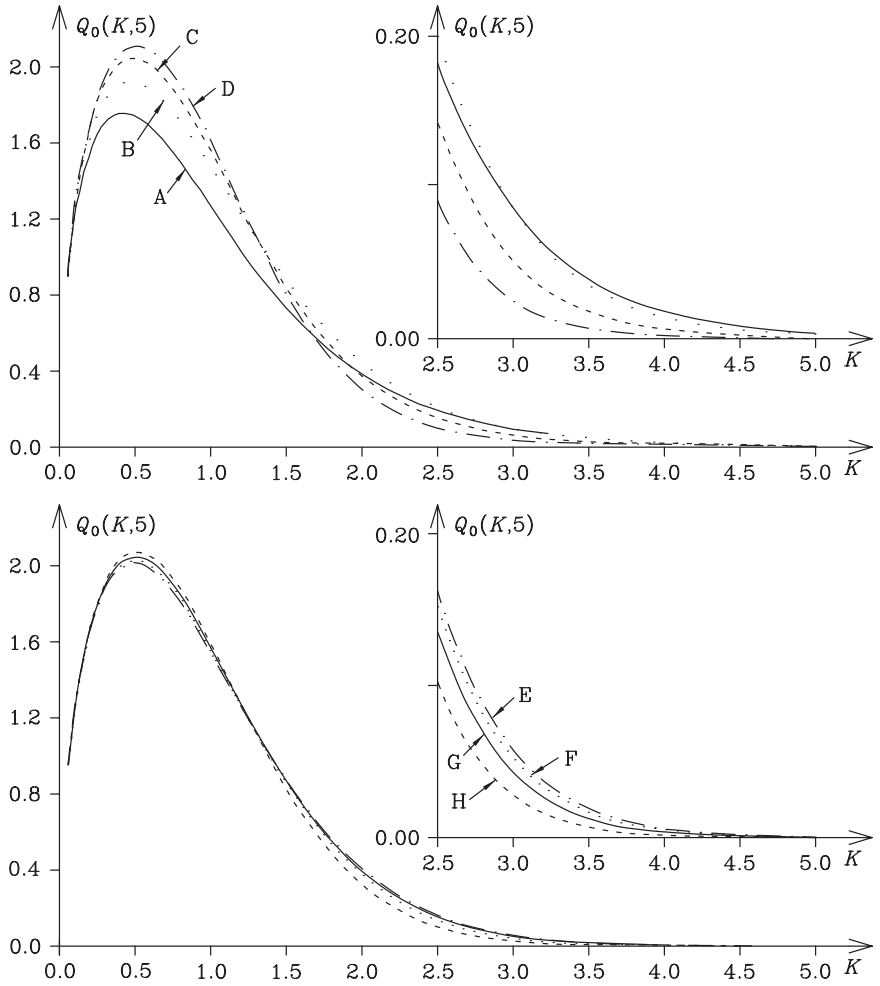


Figure 12.7: $Q_0(K, 5)$ curves computed using Equation 12.19 and the eight functions $A(x)$ shown in Fig. 12.6. Upper part: A: boxcar, B: trapezoidal, C: triangular, D: triangular^2. Lower part: E: Bessel, F: cosine, G: sinc^2, H: Gaussian. [6]

smoothing functions of Figure 12.6. We can see that when $K > 1$, then Q_0 decreases rapidly and nonlinearly as a function of K. At high K-values, Q_0 decreases nearly exponentially. The higher the resolution enhancement is, the worse the signal-to-noise ratio becomes.

It is evident from Equation 12.19, that $Q_0(K, K_0)$ at other values of K_0 behaves in a similar way to the curves at $K_0 = 5$, because

$$Q_0(K, K_0) = \sqrt{\frac{K_0}{5}}\, Q_0(K, 5). \tag{12.20}$$

Figure 12.8 (a) illustrates the behavior of noise in Fourier self-deconvolved spectra at various K-values. It is a simulation, where a Lorentzian line with $K_0 = 20$ and $(S/N) \approx 2980$ is Fourier self-deconvolved at seven different values of K. The shape of the window function $A(t)$ used in smoothing is sinc^2. The increase of noise in the Fourier self-deconvolved spectrum at higher K-values is evident. We can also see that at high K-values the noise is no longer random, but components with periodicity of about $1/T$, where T is the t-domain point after which $A(t)$ stays as zero, dominate. This is approximately the width of the new line shape. This phenomenon can also be seen in Figure 12.8 (b), which shows the noise interferograms obtained from the original noise interferogram $n(t)$ by Fourier self-deconvolution with the same parameters as in (a). The amplitude of the noise interferogram has a maximum just below the point T, and this results in the periodicity of $1/T$ in the noise spectrum.

The Q_0 curves in Figure 12.7 show that the value of Q_0 depends not only on the K-value, but also on the shape of the chosen smoothing function $W(f)$. If K is small, less than 2.5, then the maximum factor by which the Q_0 values by different smoothing functions differ is about 2.2. At higher values of K, however, the differences are much larger. The higher K is, the more critical is the choice of the smoothing function.

The effect of the shape of the smoothing function on the signal-to-noise ratio of the Fourier self-deconvolved spectrum is illustrated also in Figure 12.9. The figure shows an experimental IR spectrum $H_R(f)$ of liquid chlorobenzene. The spectral lines are assumed to be Lorentzian with $\sigma_R \approx 2\sigma_0 = 9.8\ \text{cm}^{-1}$. The spectrum is Fourier self-deconvolved using eight different smoothing functions, at the same value $K = 4$. We can see that there is more than an order-of-magnitude difference in $(S/N)_{FSD}$ between, for example, the triangular[2] and the boxcar weighting.

It is clear that, in practice, the maximum value of K is limited by the noise in the recorded spectrum. If the signal-to-noise ratio of H_R is (S/N), then the maximal recommended narrowing is given by

$$\boxed{K_{max} \approx \log_{10}(S/N)\,.} \tag{12.21}$$

By examining the Fourier self-deconvolved spectra, we can find that some of the spectra contain strong sidelobes. The most dominant sidelobes are obtained, if the weight function is the boxcar function. At high K-values, the enhancement, or decrease, of the signal-to-noise ratio is roughly proportional to the magnitude $|s|$ of the strongest sidelobe of the smoothing function $W(f)$:

$$\frac{Q_0}{|s|} \approx \text{constant}. \tag{12.22}$$

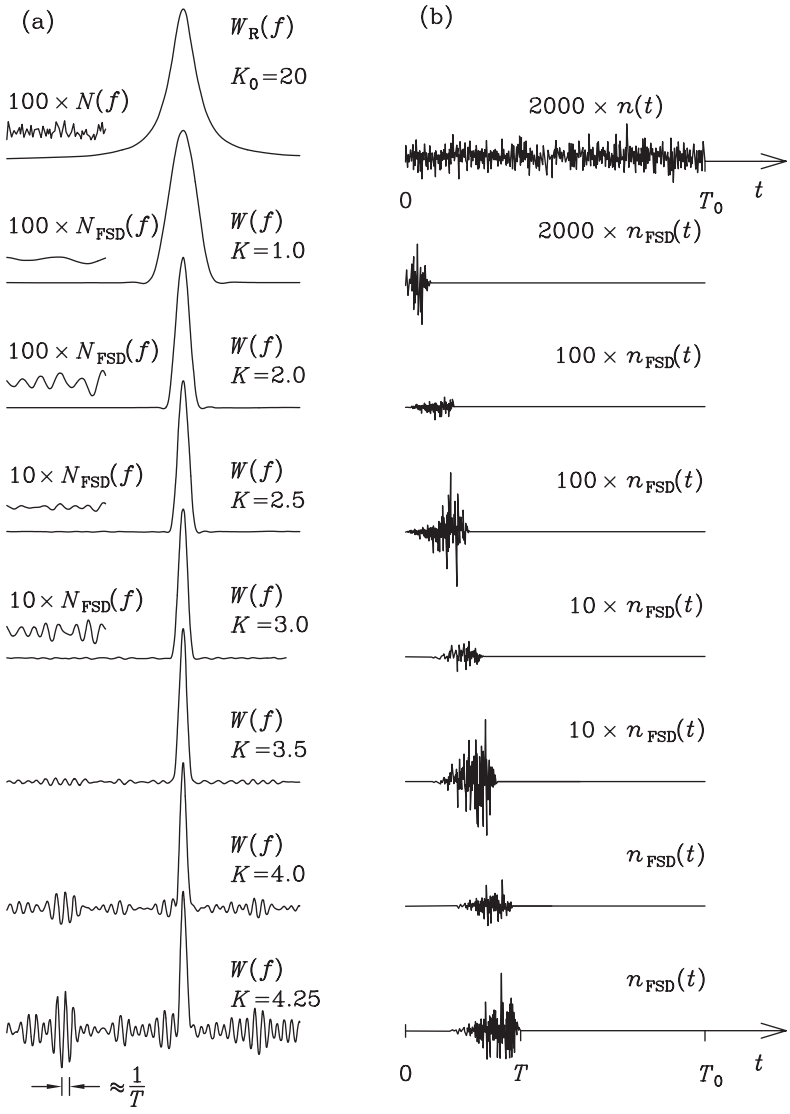


Figure 12.8: (a) A Lorentzian line $W_R(f)$ (top) with random noise, with $(S/N) \approx 2980$ and $K_0 = 20$, and Fourier self-deconvolved spectra at $K = 1.0$, 2.0, 2.5, 3.0, 3.5, 4.0, and 4.25, obtained using the sinc^2 function as $A(t)$. (b) The noise interferogram $n(t)$ (top) and the noise interferograms $n(t) = \Pi(t)\,\mathrm{sinc}^2(\pi t/T)\exp(\pi\sigma_R|t|)n(t)$ after Fourier self-deconvolution at same parameters as in (a). The source of the periodicity of about $1/T$ of the noise in (a) is clear from the corresponding noise interferograms (b). [6]

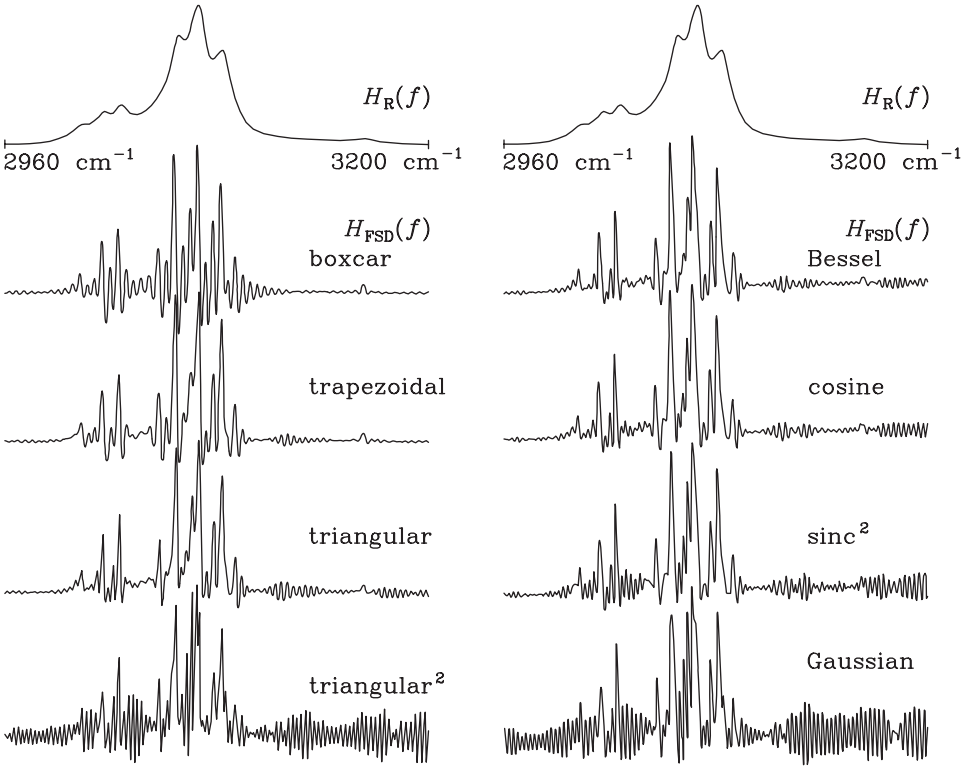


Figure 12.9: Comparison of the effect of the shape of the weighting function $A(t)$ on the $(S/N)_{\mathrm{FSD}}$ of a Fourier self-deconvolved spectrum. $H_{\mathrm{R}}(f)$ is an experimentally registered IR spectrum of liquid chlorobenzene in the C–H stretching region. The spectral lines are assumed to be Lorentzian, with $\sigma_{\mathrm{R}} \approx 2\sigma_0 = 9.8\ \mathrm{cm}^{-1}$. $K = 4$ is chosen. The Fourier self-deconvolved spectra are computed using eight different weight functions $A(t)$ of Fig. 12.6. [6]

An optimal situation is achieved, when the sidelobe and the noise are of similar magnitude:

$$(S/N)_{\mathrm{FSD}} \approx 1/|s|. \tag{12.23}$$

Example 12.1: A spectrum consists of spectral lines which are all of the same Lorentzian shape

$$W_{\mathrm{R}}(f) = \frac{\sigma/\pi}{\sigma^2 + f^2}.$$

By which function should we multiply the signal, if we wish the new line shape to be a sinc function, and the spectral resolution enhancement factor to be $K = 2$? The areas of the lines should remain unchanged.

Solution. The signal amplitude which gives the Lorentzian line shape is

$$I_{\mathrm{R}}(t) = \mathcal{F}\{W_{\mathrm{R}}(f)\} = e^{-2\pi\sigma|t|}.$$

The desired new line shape is

$$W(f) = 2T\,\mathrm{sinc}(\pi 2Tf).$$

The corresponding signal amplitude is the boxcar

$$A(t) = \Pi_{2T}(t) = \begin{cases} 0, & |t| > T, \\ 1, & |t| \leq T. \end{cases}$$

In order to obtain sinc-shaped lines, we must multiply the signal by the function

$$A(t)/I_{\mathrm{R}}(t) = \Pi_{2T}(t)\, e^{2\pi\sigma|t|}.$$

Since the value of this function in the origin is one, it retains the line areas, as required. The truncation point T is determined by the condition

$$K = \frac{\text{FWHM of the registered line}}{\text{FWHM of the smoothing function}} \approx \frac{2\sigma}{1.2067/2T} = 2.$$

This yields

$$T \approx \frac{1.2067}{2\sigma}.$$

12.3 Underdeconvolution and overdeconvolution

In Fourier self-deconvolution, the line shape of the spectrum is assumed to be known. If the assumed line shape is not correct, then the final line shape of the Fourier self-deconvolved spectrum differs from the desired line shape $W(f)$. However, even then the resolution is enhanced, and the frequency and the area of a line are not distorted.

It may sometimes be difficult to estimate the half-width of the spectral lines from the observed spectrum. The half-width of a single line is easy to measure, but the half-width of several overlapping lines may be impossible to measure. In addition, the half-widths of different lines often differ from each other. If the half-width of a line is underestimated, then the Fourier self-deconvolved spectrum has a line shape close to the original line shape, and the spectral resolution is lower than predicted. This situation is called *underdeconvolution*. If the half-width of a line is overestimated, then the spectral resolution after Fourier self-deconvolvolution may be slightly higher than expected, but the Fourier self-deconvolved line has strong negative sidelobes. This situation is called *overdeconvolution*.

Figure 12.10 shows simulated examples of underdeconvolution and overdeconvolution. Figure 12.10 (a) shows two FTS spectra with Lorentzian lines. In Figure 12.10 (b), (c), and (d) these spectra have been Fourier self-deconvolved, using a triangular squared weight function. In 12.10 (b) the half-width of the lines is 25 % underestimated, and overlapping lines are not adequately resolved. In 12.10 (c) the correct half-width is used. In 12.10 (d) the half-width is 25 % overestimated, and the lines have strong sidelobes.

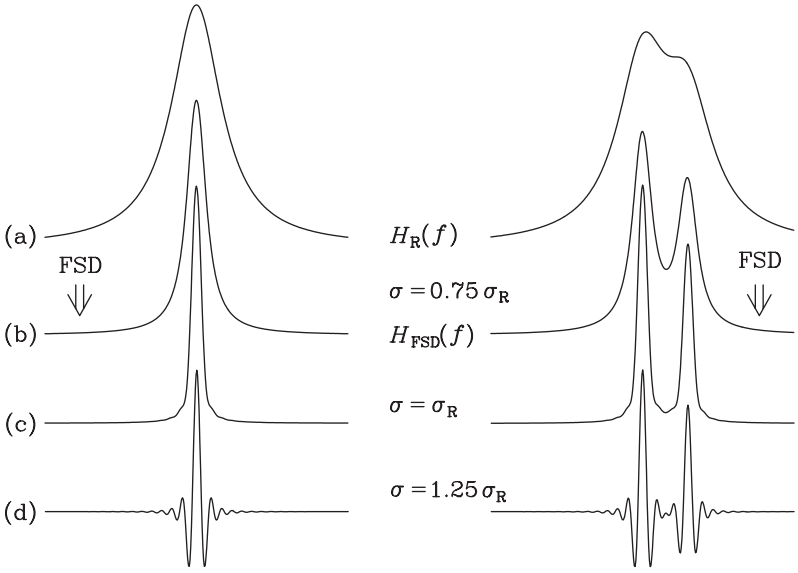


Figure 12.10: The effects of incorrect half-width in Fourier self-deconvolution (FSD). (a) Simulated spectra, consisting of a single Lorentzian line (left), and of two overlapping Lorentzian lines (right), all of the half-width $\sigma_R = 8\ \mathrm{cm}^{-1}$. These spectra are Fourier self-deconvolved, using a triangular^2 weight function with $T = 0.8$ cm. In (b) the assumed half-width of the lines σ is 25 % too small, in (c) it is correct, and in (d) it is 25 % too large.

12.4 Band separation

Band separation is an application where the usefulness of Fourier self-deconvolution is very obvious. From the Fourier self-deconvolved spectrum we can easily separate a single line, without touching other lines. The desired line may be removed, even if the lines in the original spectrum would overlap. The frequency and the area of the line are preserved. After separation of the desired spectral band, we can restore the line shape information by making an inverse Fourier self-deconvolution. The result is the original line shape, even if the assumed line shape in FSD would not be correct.

Figure 12.11 shows an FTS simulation which illustrates the band separation procedure. The original spectrum consists of two overlapping Lorentzian lines at the wavenumbers f_1 and f_2, with $\sigma_R = 8\ \mathrm{cm}^{-1}$. The weight function used in FSD is triangular^2, with $T = 0.8$ cm. The original spectrum is first Fourier self-deconvolved by multiplying the corresponding interferogram in the signal domain by $(1-|t|/T)^2 e^{\sigma_R \pi |t|}$ and then transforming back to the spectral domain. After that, the line at f_2 is removed. The remaining spectrum, consisting of the line at f_1, is convolved by multiplying the corresponding interferogram by $1/\left[(1-|t|/T)^2 e^{\sigma_R \pi |t|}\right]$ and then transforming back to the spectral domain. The result is the original spectrum with only one Lorentzian line.

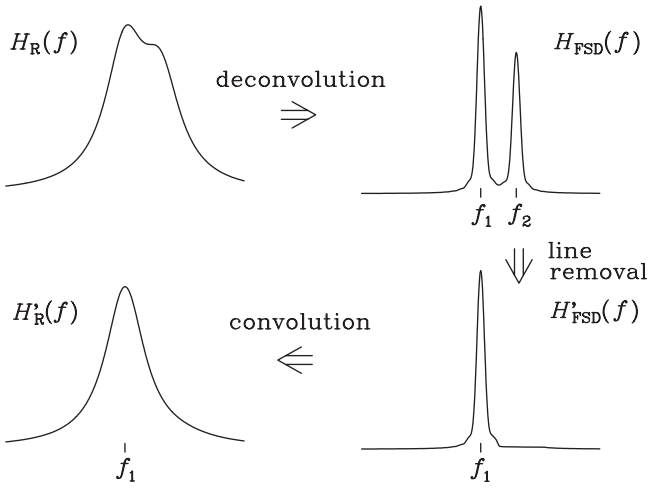


Figure 12.11: Illustration of the band separation procedure. The original spectrum $H_R(f)$ consists of two Lorentzian lines of the half-width 8 cm^{-1} at the distance 6 cm^{-1} from each other. One of the lines is removed from the Fourier self-deconvolved spectrum $H_{FSD}(f)$. The remaining spectrum is Fourier self-deconvolved back to the spectrum $H'_R(f)$, which differs from the original, because it now consists of only one Lorentzian line.

12.5 Fourier complex self-deconvolution

Fourier complex self-deconvolution, FCSD, is Fourier self-deconvolution where the registered line shape function $W_R(f)$ is asymmetric. In this case, the Fourier transform of the line shape function is complex:

$$\mathcal{F}\{W_R(f)\} = \mathrm{Re}(t) + i\,\mathrm{Im}(t), \tag{12.24}$$

where $\mathrm{Re}(t)$ and $\mathrm{Im}(t)$ are the real and imaginary parts of $\mathcal{F}\{W_R(f)\}$, respectively. The desired line shape $W(f)$ after FCSD is, however, symmetric. The change of the line shape by FCSD is illustrated in Figure 12.12.

Using Equation 12.24, Equation 12.2 of FSD can be written as

$$H_{FCSD}(f) = \mathcal{F}^{-1}\left\{\frac{[\mathrm{Re}(t) - i\,\mathrm{Im}(t)]\,\mathcal{F}\{W(f)\}\mathcal{F}\{H_R(f)\}}{[\mathrm{Re}(t)]^2 + [\mathrm{Im}(t)]^2}\right\}, \tag{12.25}$$

where H_{FCSD} is the spectrum after FCSD. This spectrum is still real, even though $\mathcal{F}\{W_R\}$ is complex.

FCSD is used instead of conventional FSD if the line shape is asymmetric and complicated. The goal is to make the line shape symmetric, and also to enhance spectral resolution. The decrease of the signal-to-noise ratio $(S/N)_{FCSD}$ as a function of the resolution enhancement factor K is in FCSD quite similar to conventional FSD, only somewhat faster.

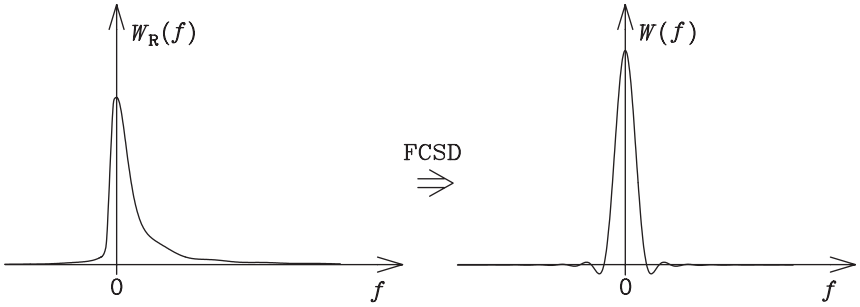


Figure 12.12: The principle of FCSD: an asymmetric line shape $W_R(f)$ is changed to a new, symmetric line shape $W(f)$.

Figure 12.13 shows a simulated example of FCSD. The original spectrum $H_R(f)$ consists of two overlapping lines and one singlet line, all of the same asymmetric line shape $W_R(f)$. The spectrum also contains some noise. After FCSD, all three lines are clearly resolved, and have a symmetric line shape $W(f)$.

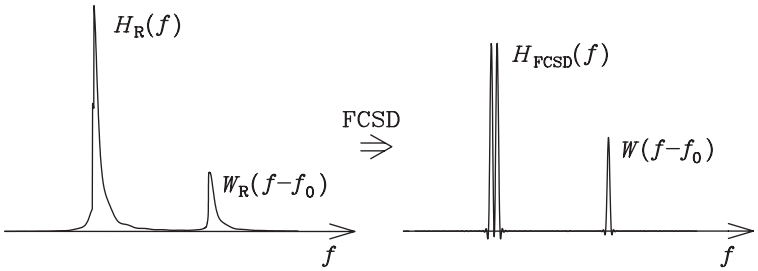


Figure 12.13: Simulated example of FCSD. The original spectrum $H_R(f)$ consists of two overlapping lines and one singlet line, all of the same asymmetric line shape $W_R(f)$. The spectrum also contains some noise. After FCSD, the spectrum $H_{FCSD}(f)$ consists of three separate lines of a symmetric line shape $W(f)$.

Example 12.2: Show that the Fourier transform $\mathcal{F}\{W_R(f)\}$ of an asymmetric line shape $W_R(f)$ has a non-zero imaginary part.

Solution. Any function can be expressed as the following sum:

$$W_R(f) = \frac{1}{2}[W_R(f) + W_R(-f)] + \frac{1}{2}[W_R(f) - W_R(-f)] = G_1(f) + G_2(f),$$

where

$$G_1(f) = \frac{1}{2}[W_R(f) + W_R(-f)]$$

is a symmetric function, and

$$G_2(f) = \frac{1}{2}\left[W_{\mathrm{R}}(f) - W_{\mathrm{R}}(-f)\right]$$

is an antisymmetric function.

If the line shape function is asymmetric, then $G_2(f)$ is a non-zero function. The Fourier transform of the asymmetric line shape function is then

$$\begin{aligned}
\mathcal{F}\{W_{\mathrm{R}}(f)\} &= \mathcal{F}\{G_1(f)\} + \mathcal{F}\{G_2(f)\} \\
&= \int_{-\infty}^{\infty} \cos(2\pi tf) G_1(f)\mathrm{d}f + i\int_{-\infty}^{\infty} \sin(2\pi tf) G_1(f)\mathrm{d}f \\
&\quad + \int_{-\infty}^{\infty} \cos(2\pi tf) G_2(f)\mathrm{d}f + i\int_{-\infty}^{\infty} \sin(2\pi tf) G_2(f)\mathrm{d}f \\
&= \int_{-\infty}^{\infty} \cos(2\pi tf) G_1(f)\mathrm{d}f + i\int_{-\infty}^{\infty} \sin(2\pi tf) G_2(f)\mathrm{d}f.
\end{aligned}$$

The term $\int_{-\infty}^{\infty} \cos(2\pi tf) G_1(f)\mathrm{d}f$ is real, and the term $i\int_{-\infty}^{\infty} A\sin(2\pi tf) G_2(f)\mathrm{d}f$ is imaginary. The following two integrals vanish. Since $G_2(f)$ is a non-zero function, then also the integral $i\int_{-\infty}^{\infty} \sin(2\pi tf) G_2(f)\mathrm{d}f$ is not identically zero.

12.6 Even-order derivatives and FSD

FSD is not the only method which is used in resolution enhancement of a recorded spectrum; another often employed is the computation of even-order *derivatives*.

The most practical way to compute the derivative of a spectrum is to use Fourier transforms. According to the derivative theorem, differentiation in the spectral domain corresponds to multiplying by $(-i2\pi t)$ in the signal domain. If a spectrum $H(f)$ in f-domain is differentiated k times, then the corresponding signal $h(t) = \mathcal{F}\{H(f)\}$ in t-domain is multiplied by the function

$$\boxed{A(t) = (-i2\pi t)^k.} \tag{12.26}$$

We can write the kth derivative as

$$\frac{\mathrm{d}^k}{\mathrm{d}f^k} H(f) = \int_{-\infty}^{\infty} (-i2\pi t)^k h(t) e^{-i2\pi ft}\,\mathrm{d}t. \tag{12.27}$$

In practice, differentiation alone does not lead to a good result, if the spectrum contains noise. In this case, the derivative of the spectrum looks like pure noise. In addition to differentiation, it is necessary to smooth the spectrum; this is demonstrated in Figure 12.14.

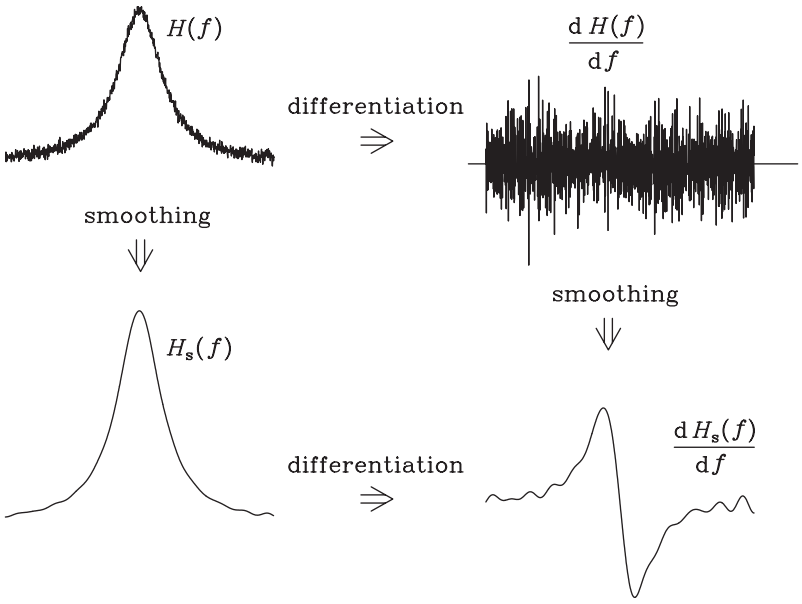


Figure 12.14: The derivative $\mathrm{d}H(f)/\mathrm{d}f$ of an unsmoothed spectrum $H(f)$ looks like pure noise, whereas the derivative $\mathrm{d}H_\mathrm{S}(f)/\mathrm{d}f$ of a smoothed spectrum $H_\mathrm{S}(f)$ may be useful.

It is practical to make also smoothing of the spectrum in t-domain, by multiplying the corresponding signal by a window function. The best window function is a boxcar function $\Pi(t)$, because it distorts the shape of the derivative the least (see Section 12.5). Hence, combined derivation and smoothing of a spectrum $H(f)$ in f-domain is done by multiplying the corresponding signal $h(t)$ in t-domain by the function

$$A_\Pi(t) = (-i2\pi t)^k \Pi(t). \tag{12.28}$$

The line shape of the derivative of a spectrum is complicated compared to the line shape after FSD, where it can be chosen by using the desired smoothing function. The kth derivative of a smooth spectral line has $k+1$ positive or negative peaks, and k zero-crossings. If k is odd, the derivative is antisymmetric, and one of the zero-crossings lies exactly at the line position. If k is even, the derivative is symmetric, and the highest peak (by absolute value) lies at the line position.

In FSD, the resolution enhancement factor

$$K = \frac{\text{FWHM of the original line}}{\text{FWHM of the new line}}$$

is the same as the smoothing parameter

$$K = \frac{\text{FWHM of the original line}}{\text{FWHM of the smoothing function}},$$

since the smoothing function is the same as the new line shape, and it may be chosen freely. In derivation, the resolution enhancement factor is not the same as the smoothing parameter, and it can obtain only some fixed values. If the line shape is Lorentzian, the second, fourth, and sixth derivatives are, respectively, 2.7, 3.9, and 5.3 times narrower than the original line.

The signal-to-noise ratio of even-order derivative spectra are approximately the same as those of Fourier self-deconvolved spectra with the same value of K. In both methods, the maximum value of K is limited by the signal-to-noise ratio of the original spectrum. The optimum number of derivations of a Lorentzian line with $(S/N) \approx 200$, 2000, and 20 000 are, respectively, two, four, and six.

Figure 12.15 is a comparison of even-order derivative spectra and Fourier self-deconvolved spectra of noisy Lorentzian lines with three different values of (S/N). We can see that the signal-to-noise ratios are comparable in both methods, but FSD yields a far better line shape.

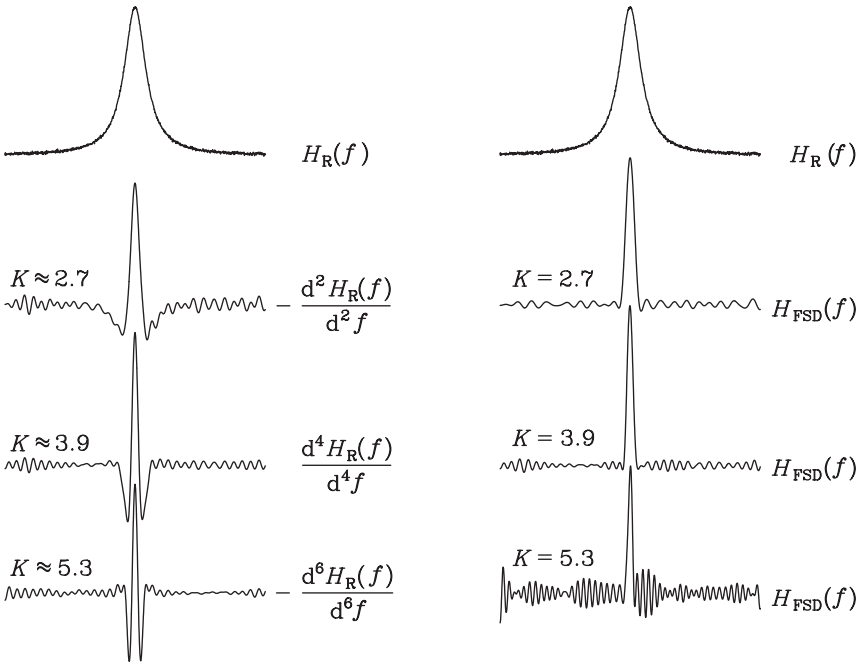


Figure 12.15: Comparison of derivation and FSD. Left: a Lorentzian line with $(S/N) = 200$, the second derivative of a Lorentzian line with $(S/N) = 200$, the fourth derivative of a Lorentzian line with $(S/N) = 2000$, and the sixth derivative of a Lorentzian line with $(S/N) = 20\,000$. The orders of differentiation are optimal to the signal-to-noise ratios. Right: a Lorentzian line with $(S/N) = 200$, and Fourier self-deconvolutions of Lorentzian lines with $(S/N) = 200$, $(S/N) = 2000$, and $(S/N) = 20\,000$. The line shape after FSD was Bessel function (see Figure 12.6), and the value K was in each case chosen to be the same as in the left column.

Example 12.3: A spectrum consists of spectral lines which are all of the same Lorentzian shape

$$\text{constant} \times L(f) = \frac{\text{constant}}{\sigma^2 + f^2}.$$

In order to enhance spectral resolution, the spectrum is differentiated twice and multiplied by (-1). (a) Determine the new line shape $-L^{(2)}(f)$ (without signal truncation). (b) Find an estimate for the FWHM of $-L^{(2)}(f)$, and also for the resolution enhancement factor K.

Solution. (a) The first derivative of $L(f)$ is

$$L^{(1)}(f) = -\frac{2f}{(\sigma^2 + f^2)^2}.$$

The second derivative is

$$L^{(2)}(f) = -\frac{2}{(\sigma^2 + f^2)^2} + \frac{8f^2}{(\sigma^2 + f^2)^3}.$$

Consequently, the new line shape is

$$-L^{(2)}(f) = \frac{2\sigma^2 - 6f^2}{(\sigma^2 + f^2)^3}.$$

(b) Let us first find the point $f_{1/2}$ where $-L^{(2)}(f_{1/2}) = -\frac{1}{2}L^{(2)}(0)$. The FWHM of the new line is, then, $2f_{1/2}$.

At $f = 0$:

$$-L^{(2)}(0) = \frac{2\sigma^2}{\sigma^6} = \frac{2}{\sigma^4}.$$

Consequently, the condition $-L^{(2)}(f_{1/2}) = -\frac{1}{2}L^{(2)}(0)$ gives

$$\frac{2\sigma^2 - 6f_{1/2}^2}{(\sigma^2 + f_{1/2}^2)^3} = \frac{1}{\sigma^4},$$

or

$$\sigma^6 - 9\sigma^4 f_{1/2}^2 - 3\sigma^2 f_{1/2}^4 - f_{1/2}^6 = 0.$$

This equation has the root, approximately,

$$f_{1/2} \approx \frac{\sigma}{3.0550}.$$

The resolution enhancement factor is, thus,

$$K = \frac{2\sigma}{2f_{1/2}} \approx \frac{2\sigma}{2\sigma/3.0550} = 3.0550.$$

This resolution enhancement factor is, in practice, only a theoretical upper limit, because the spectrum must be smoothed in addition to differentiation, that is, the signal must be truncated, and this makes the lines wider and the true value of K smaller.

Problems

1. A spectrum consists of spectral lines which are all of the same shape

$$W_{\mathrm{R}}(f) = B\,\frac{\sigma^2}{\sigma^2 + f^2},$$

where the constant B is the height of a line. By which function should we multiply the signal, if we wish to obtain a new line shape $W(f) = BT\,\mathrm{sinc}^2(\pi T f)$?

2. A spectrum consists of spectral lines which all are of Lorentzian shape and which all have the same FWHM of 2σ. In order to change the shape of the lines from Lorentzian to triangular, Fourier self-deconvolution is applied. By which function should the signal be multiplied, if the spectral resolution enhancement factor should be K, and

   (a) the areas of the lines should not be changed, or

   (b) the heights of the lines should not be changed?

3. A spectrum consists of spectral lines which all are of Lorentzian shape and have the same FWHM of 2σ. Fourier self-deconvolution is applied in such a way that also the new lines are of Lorentzian shape, but their FWHM is $2\sigma' < 2\sigma$. Compute the signal-to-noise ratio enhancement factor Q_0 as a function of the smoothing parameters K and K_0, if the signal was in registration truncated in the time interval $[-T_0, T_0]$. You can approximate $e^{-2\pi\sigma' T_0} \approx 0$.

4. Show that the signal-to-noise-ratio of Lorentzian lines is enhanced in FSD by the factor (Eq. 12.19)

$$Q_0 = \frac{(S/N)'}{(S/N)} \approx \frac{\pi\sqrt{1.2067 K_0 \sigma}\int_0^\infty A(t)\,\mathrm{d}t}{\sqrt{\int_0^\infty A^2(t) e^{4\pi\sigma t}\,\mathrm{d}t}},$$

if the signal has been multiplied by a boxcar function in registration.

Hint: If a noisy signal is apodized (multiplied) by a function $B(t)$, then the rms-value of the noise in the spectrum is multiplied by $\sqrt{\int_{-\infty}^{\infty} B^2(t)\,\mathrm{d}t}$. You can assume $e^{-2\pi\sigma T_0} \approx 0$. The function $A(t)$ is symmetric.

5. As suggested by A. Losev [7], asymmetric spectral lines can be described by the line profile

$$W_{\mathrm{R}}(f) = \frac{B}{e^{-af} + e^{bf}},$$

where a, b, and B are positive real constants. The constant B is determined by the condition that the area of the line shape function must be equal to one, *i.e.*, $\mathcal{F}\{W_{\mathrm{R}}(f)\} = 1$ at $t = 0$.

If a spectrum consists of lines of this shape, by which function should the signal be divided in FSD?

Hint: From mathematical tables,

$$\int_0^\infty \frac{\cosh(Cx)}{\cosh(Dx)}\,\mathrm{d}x = \frac{\pi}{2D}\,\frac{1}{\cos\left[C\pi/(2D)\right]}, \qquad \mathrm{Re}(D) > \mathrm{Re}(C).$$

6. The part of a spectrum which contains spectral lines of height A_1, A_2, and A_3 at wavenumbers ν_1, ν_2, and ν_3, respectively, has been separated from the rest of the spectrum. All these three lines are of the line shape $W_\mathrm{R}(\nu)$, and the piece of spectrum can be expressed as

$$S(\nu) = A_1 W_\mathrm{R}(\nu - \nu_1) + A_2 W_\mathrm{R}(\nu - \nu_2) + A_3 W_\mathrm{R}(\nu - \nu_3).$$

   (a) With which function should the line shape function $W_\mathrm{R}(\nu)$ be convolved in order to obtain the spectrum $S(\nu)$ with three lines?

   (b) By which function should the signal corresponding to the spectrum $S(\nu)$ be divided in order to leave the pure line shape function $W_\mathrm{R}(\nu)$ in the spectral domain?

7. Applying the derivative theorem, find the derivative function of the function

$$\Pi_L(x - s) = \begin{cases} 1, & |x - s| \le L/2, \\ 0, & |x - s| > L/2, \end{cases}$$

   which is the boxcar function of width L and height one, shifted by s from the origin.

8. A Lorentzian spectral line is differentiated twice and multiplied by (-1). Determine the relative height of the negative sidelobes of the new line shape, *i.e.*, $-O/S$, where O is the depth of the sidelobes, and S is the height of the line.

9. All the lines of a line spectrum are of the Gaussian shape, *i.e.*, if shifted to origin they are of the form $AG(\nu) = Ae^{-C\nu^2}$, where A and C are real constants, and $C > 0$. In order to enhance spectral resolution, the spectrum is differentiated twice and multiplied by (-1). The new spectral lines are of the form $-AG^{(2)}(\nu)$. The effect of smoothing can be neglected.

   (a) What is the relative height of the negative sidelobes of the differentiated line shape?

   (b) Find an estimate for the FWHM of $-G^{(2)}(\nu)$, as well as an estimate for the resolution enhancement factor

$$K = \frac{\text{FWHM of original line shape}}{\text{FWHM of new line shape}}.$$

10. A spectrum $S(\nu)$ is differentiated and, in addition, smoothed by multiplication by the boxcar function $\Pi_{2L}(x)$ in the signal domain. Show that this whole operation is a convolution in the spectral domain. Which function is the spectrum $S(\nu)$ convolved with?

    Hint: Use the derivative and the convolution theorems.

11. Applying the derivative theorem, determine the first and the second derivative functions of the function

$$\Lambda_L(x-s) = \begin{cases} 1-|x-s|/L, & |x-s| \le L, \\ 0, & |x-s| > L, \end{cases}$$

which is the one-unit high triangular function of FWHM L, shifted by s from the origin.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

13 Linear prediction

13.1 Linear prediction and extrapolation

Linear prediction[1] is a means of estimating a missing sample of a sampled function. As the name implies, the missing sample is calculated as a linear combination of the known samples. Usually the known samples either all precede or all follow the missing sample. Correspondingly, we speak about forward linear prediction or backward linear prediction. In Fourier transform interferometry, the sequence of samples will always be a series or equidistant interferogram samples I_j, and in NMR it is the FID signal, but we will call it more generally signal. The forward predicted estimate I'_j of a sample I_j can be written as

$$\boxed{I'_j = \Delta x \sum_{l=1}^{M} h_l I_{j-l}\,,} \tag{13.1}$$

and the backward predicted estimate as

$$\boxed{I'_j = \Delta x \sum_{l=1}^{M} h_l^* I_{j+l}\,.} \tag{13.2}$$

The reason for complex conjugating the coefficients in the backward equation will be explained in the following section. This detail is not important for us, however, since we will restrict ourselves to real coefficients h_l.

The usefulness of linear prediction in spectroscopy comes from its property that it lends itself readily to extrapolation. By applying linear prediction repeatedly, using predicted sample estimates to predict more estimates, we are able to extrapolate a measured signal. We call the signal *predictable* if h_l-coefficients exist such that the predicted estimate is free of errors, or $I'_j = I_j$ independent of j. Thus, predictability means that the signal can be extrapolated arbitrarily far without errors. The prediction equations, Equations 13.1 and 13.2, then define the *one-point extrapolation*.

[1] Chapter 13 is contributed by Dr. Pekka Saarinen, Dept. of Applied Physics, University of Turku, Finland

13.2 Extrapolation of linear combinations of waves

According to what was mentioned above, if in Equation 13.1 the estimate is errorless, then we can extrapolate the signal sample by sample as far as we wish. This kind of situation is very advantageous, and it is worth examining under what conditions the equidistant samples $I_j = I(j\Delta x)$ can be extrapolated in this way. We begin with the exponential wave $I(x) = A\exp(i2\pi\nu x + i\phi)$, where A is a real constant. Clearly, the ratio of two successive samples is in this case constant (which we denote by $\Delta x C_\nu$), since

$$\frac{I_{j+1}}{I_j} = \frac{e^{i2\pi\nu(j+1)\Delta x + i\phi}}{e^{i2\pi\nu j\Delta x + i\phi}} = e^{i2\pi\nu\Delta x} = \Delta x C_\nu, \tag{13.3}$$

which is independent of j. Thus, signal $A\exp(i2\pi\nu x + i\phi)$ can be extrapolated without errors by using only one previous sample in each prediction. Then, $M = 1$ and $h_1 = C_\nu = \exp(i2\pi\nu\Delta x)/\Delta x$.

Next, we shall discuss the predictability of a much more realistic signal that is a linear combination of exponential waves with different wavenumbers. We saw above that one exponential wave (with arbitrary amplitude and phase shift) can always be extrapolated. If we manage to show that adding one new exponential wave to a predictable signal $I(x)$ does not destroy its predictability, then we can add the waves one by one, and it follows that any linear combination of such waves can be extrapolated by linear prediction. To this end we assume that $I(x)$ can be extrapolated using M h_l-coefficients $h_1, \ldots, h_M$. We add to it the exponential wave $e(x) = A\exp(i2\pi\nu x + i\phi)$ and try to extrapolate the summed signal $I(x) + e(x)$. According to Equation 13.3, equidistant samples taken from $e(x)$ can be extrapolated by multiplying the previous sample by $\Delta x C_\nu = \exp(i2\pi\nu\Delta x)$. We thus have

$$\begin{cases} I_j &= \Delta x h_1 I_{j-1} + \Delta x h_2 I_{j-2} + \cdots + \Delta x h_M I_{j-M}, \\ e_j &= \Delta x C_\nu e_{j-1}\,, \end{cases} \tag{13.4}$$

where j is any integer. We now claim that equidistant samples taken from $I(x) + e(x)$ can be extrapolated by using the $M + 1$ coefficients $h'_1, \ldots, h'_{M+1}$, where

$$\begin{cases} h'_1 &= h_1 + C_\nu, \\ h'_2 &= h_2 - \Delta x h_1 C_\nu, \\ h'_3 &= h_3 - \Delta x h_2 C_\nu, \\ &\vdots \\ h'_M &= h_M - \Delta x h_{M-1} C_\nu, \\ h'_{M+1} &= -\Delta x h_M C_\nu\,. \end{cases} \tag{13.5}$$

If we can prove that an arbitrary sample $I_j + e_j$ from the summed signal is a linear combination of the preceding $M+1$ samples, with the coefficients given by Equation 13.5, we have proved that $I(x) + e(x)$ is predictable, and have even found the prediction coefficients.

The right-hand side of the forward prediction equation (Equation 13.1) now reads as

$$\begin{aligned}
& \Delta x h'_1 \left(I_{j-1} + e_{j-1}\right) + \Delta x h'_2 \left(I_{j-2} + e_{j-2}\right) + \cdots \\
& + \Delta x h'_M \left(I_{j-M} + e_{j-M}\right) + \Delta x h'_{M+1} \left(I_{j-M-1} + e_{j-M-1}\right) \\
= \quad & \Delta x \left(h_1 + C_\nu\right) \left(I_{j-1} + e_{j-1}\right) + \Delta x \left(h_2 - \Delta x h_1 C_\nu\right) \left(I_{j-2} + e_{j-2}\right) + \cdots \\
& + \Delta x \left(h_M - \Delta x h_{M-1} C_\nu\right) \left(I_{j-M} + e_{j-M}\right) - (\Delta x)^2 h_M C_\nu \left(I_{j-M-1} + e_{j-M-1}\right) \\
= \quad & \left(\Delta x h_1 I_{j-1} + \Delta x h_2 I_{j-2} + \cdots + \Delta x h_M I_{j-M}\right) + \left(\Delta x C_\nu e_{j-1}\right) \\
& + \Delta x C_\nu \left[I_{j-1} - \Delta x h_1 I_{j-2} - \cdots - \Delta x h_{M-1} I_{j-M} - \Delta x h_M I_{j-M-1}\right] \\
& + \Delta x h_1 \left[e_{j-1} - \Delta x C_\nu e_{j-2}\right] + \Delta x h_2 \left[e_{j-2} - \Delta x C_\nu e_{j-3}\right] + \cdots \\
& + \Delta x h_M \left[e_{j-M} - \Delta x C_\nu e_{j-M-1}\right].
\end{aligned}$$

It follows immediately from Equation 13.4 that all expressions in square brackets vanish. Likewise, we see that the two expressions in parentheses in the final summation are equal to I_j and e_j. Thus, the above equation simplifies to

$$\begin{aligned}
& \Delta x h'_1 \left(I_{j-1} + e_{j-1}\right) + \Delta x h'_2 \left(I_{j-2} + e_{j-2}\right) + \cdots \\
& + \Delta x h'_M \left(I_{j-M} + e_{j-M}\right) + \Delta x h'_{M+1} \left(I_{j-M-1} + e_{j-M-1}\right) \\
= \quad & I_j + e_j.
\end{aligned}$$

Thus, the value of the summed function $I(x) + e(x)$ at any point x can indeed be calculated as a linear combination of its values at the points $x - \Delta x, x - 2\Delta x, \ldots, x - (M+1)\Delta x$.

Now we remember that all sinusoidal waves can be stated as linear combinations of two exponential waves with opposite wavenumbers. Namely,

$$\begin{cases}
\cos(2\pi\nu x) & = \dfrac{1}{2} e^{i2\pi\nu x} + \dfrac{1}{2} e^{i2\pi(-\nu)x}, \\[2ex]
\sin(2\pi\nu x) & = \dfrac{1}{2i} e^{i2\pi\nu x} - \dfrac{1}{2i} e^{i2\pi(-\nu)x} = \dfrac{1}{2} e^{i2\pi\nu x - i\pi/2} - \dfrac{1}{2} e^{i2\pi(-\nu)x - i\pi/2}.
\end{cases} \tag{13.6}$$

Therefore, one sinusoidal wave can be extrapolated by linear prediction using two coefficients h_l. Further, since a linear combination of n sinusoids is a linear combination of $2n$ exponential waves, it can be extrapolated using $2n$ h_l-coefficients. This result is very important to us, because after FSD every spectral line corresponds to one sinusoid in the signal domain. (In the absence of phase errors, this sinusoid is always a cosine wave.) The necessary h_l-coefficients, when the wavenumbers ν_k are known, can be found for example as follows: First set $M = 1$ and $h_1 = C_{\nu_1} = \exp(i2\pi\nu_1\Delta x)/\Delta x$. Then increase M by one and update the h_l-coefficients according to Equation 13.5, with C_ν calculated from the current ν-value, until all the ν-values have been taken into account. This procedure gives so-called theoretical impulse response, which we shall consider afresh later.

It is also important to note that Equation 13.5 does not include the amplitude A and the phase ϕ of the new exponential wave at all. This means that *once we have constructed the h_l-coefficients, they are able to extrapolate any linear combination of the exponential waves with the given wavenumbers, irrespective of the amplitudes and phases of these constituent waves*.

Above we have discussed only extrapolation by forward linear prediction, as defined in Equation 13.1. When the signal is predicted backward, as in Equation 13.2, the situation is identical to first mirroring the signal and after that predicting it forward. In the mirroring operation every unshifted exponential wave $e(x) = \exp(i2\pi\nu x)$ is changed to the complex conjugated wave $\exp(-i2\pi\nu x) = e^*(x)$. Thus, if $I(x)$ is a linear combination of unshifted exponential waves, it is changed to $I^*(x)$, and complex conjugating the forward prediction equation and substituting $-j$ for j gives the backward prediction equation, with the h_l-coefficients complex conjugated. In the special case that the combination of exponential waves is symmetric, or in other words, a linear combination of cosine waves, it is unaffected by complex conjugation (mirroring). Then, comparing Equations 13.1 and 13.2, we see that $h_l^* = h_l$ for every j, that is, the h_l-coefficients are real. Since shifting the cosine waves does not affect their predictability, we can state the more general result that any linear combination of shifted cosine waves can be extrapolated by real h_l-coefficients.

13.3 Extrapolation of decaying waves

Thus far the wavenumber ν has always been a real quantity. In linear prediction theory this means that the constituent waves $A\exp(i2\pi\nu x + i\phi)$ have constant amplitudes. Nevertheless, in deriving Equations 13.3–13.5, real ν-values were not assumed at any stage. Thus we may substitute

$$\nu = \nu_0 + i\frac{\alpha}{2\pi}, \tag{13.7}$$

and Equations 13.3–13.5 stay valid. The complex wavenumber of Equation 13.7 represents an exponentially decaying wave, since it gives

$$Ae^{i2\pi\nu x} = Ae^{-\alpha x}e^{i2\pi\nu_0 x}.$$

Thus, a linear combination of M exponentially decaying (or growing) exponential waves can also be extrapolated using M h_l-coefficients. Then, however, the same coefficients are not necessarily capable of backward extrapolation.

In addition to exponential decay, there are certain other amplitude functions that preserve the predictability. These include polynomial amplitudes. Search of all possible types of predictable functions, however, goes beyond the scope of this book. Besides, in resolution enhancement, which is the most important spectroscopic application of linear prediction, the constituent waves are made nondecaying by FSD. This can be done, because in a relatively narrow area most spectra (excluding NMR) can be assumed to have constant line shape, which means that all the constituent waves decay in a similar manner. It would also be unwise to extrapolate a decaying signal very far, since before long it would die out.

13.4 Predictability condition in the spectral domain

According to the above reasoning, if the signal is a linear combination of a finite number of exponential waves, it is predictable and thereby can be extrapolated. Since the inverse Fourier transform of an exponential wave is a Dirac's delta function shifted to its wavenumber ν (see Table 1.1), a signal is predictable if its spectrum consists of a finite number of shifted delta functions. Let us now assume that the signal indeed is a linear combination of M different exponential waves, and that we have found the coefficients $h_1, \ldots, h_M$ able to extrapolate it. Remember that these coefficients are then able to extrapolate any linear combination of these waves. Thus, if ν_k is the wavenumber of one of the constituent waves, then the wave $e(x) = \exp(i2\pi\nu_k x)$ alone can be extrapolated by the h_l-coefficients. Especially the value of $e(x)$ in the origin can be predicted forward as a linear combination of M preceding samples, or

$$e_0 = e(0) = \Delta x h_1 e_{-1} + \Delta x h_2 e_{-2} + \cdots + \Delta x h_M e_{-M},$$

or

$$1 = \Delta x h_1 e^{-i2\pi\nu_k \Delta x} + \Delta x h_2 e^{-i2\pi\nu_k 2\Delta x} + \cdots + \Delta x h_M e^{-i2\pi\nu_k M\Delta x}. \tag{13.8}$$

Comparing the right-hand side of this equation with Equation 3.3 reveals that it is, in fact, $H(\nu_k)$, or the inverse discrete Fourier transform of the sequence $h_1, \ldots, h_M$ computed at the point ν_k. (Note that $h_l = 0$ for $j < 1$ and $j > M$.)

The above reasoning is also valid to the inverse direction: If $H(\nu_k) = 1$, then clearly Equation 13.8 holds, and $e(0)$ can be predicted forward by $h_1, \ldots, h_M$. Multiplying both sides of Equation 13.8 by $\exp(i2\pi\nu_k x)$ shows that $e(x)$ can be predicted at an arbitrary point x as well. We thus have the *predictability condition*

$$\boxed{H(\nu_k) = 1.} \tag{13.9}$$

When this condition is valid, the coefficients $h_1, \ldots, h_M$ can be used to extrapolate an exponential wave with the wavenumber ν_k. We call $H(\nu)$ the transfer function.

Note that if the constituent waves are nondecaying (have constant amplitudes), then the predictability condition of Equation 13.9 also implies that the signal can be predicted backward. This can be seen by complex conjugating Equation 13.8, giving

$$1 = \Delta x h_1^* e^{i2\pi\nu_k \Delta x} + \Delta x h_2^* e^{i2\pi\nu_k 2\Delta x} + \cdots + \Delta x h_M^* e^{i2\pi\nu_k M\Delta x}.$$

Multiplying this equation by $\exp(i2\pi\nu_k x)$ we obtain the backward one-point prediction defined by Equation 13.2.

An alternative derivation of the condition of Equation 13.9 would be to apply the discrete convolution theorem derived earlier in Example 3.3. Since the forward prediction of Equation 13.1 in fact is the discrete convolution

$$\boldsymbol{h} * \boldsymbol{I}, \tag{13.10}$$

the discrete convolution theorem gives

$$E(\nu) = \mathcal{F}^{-1}\{\boldsymbol{h}\}E(\nu) = H(\nu)E(\nu). \tag{13.11}$$

Note that we have interpreted the signal vector $\boldsymbol{I}$ infinitely long, and $\boldsymbol{h}$ has also been zero-padded from both ends to infinity. Therefore, according to Equation 3.20, their spectra $E(\nu)$ and $H(\nu)$ are continuous-parameter functions, since $\Delta\nu = 0$. Equation 13.11 implies that $H(\nu) = 1$ anywhere the spectrum differs from zero, which is just at the wavenumbers of the exponential waves contained in the signal. If the number of the constituent waves is M, then the number of h_l-coefficients required to fulfill the condition in Equation 13.11 at the wavenumbers of them all is naturally M.

In Figure 13.1 there is an example of a Fourier self-deconvolved spectrum $E(\nu)$, consisting only of a set of shifted Dirac's delta functions. Above it, there is inverse Fourier transform $H(\nu)$ of such impulse response coefficients that can extrapolate the signal $\mathcal{F}\{E(\nu)\}$. As may be seen, $H(\nu) = 1$ anywhere $E(\nu)$ differs from zero.

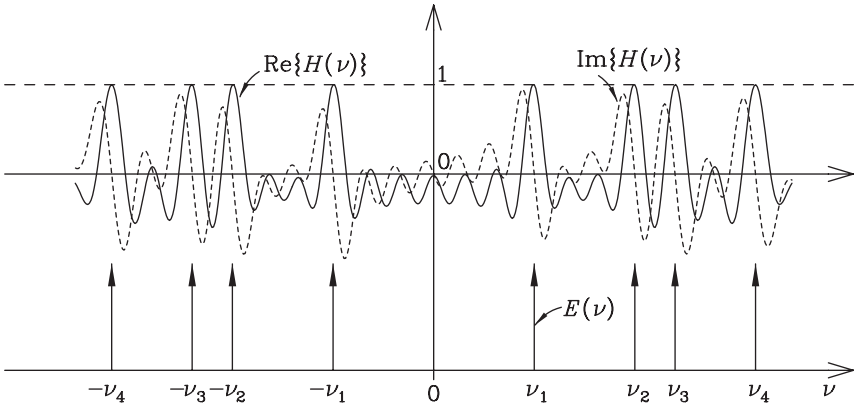


Figure 13.1: A Fourier self-deconvolved line spectrum $E(\nu)$ and (inverse) Fourier transform of a sequence of h_l-coefficients able to extrapolate its Fourier transform. Then, the transfer function $H(\nu_k) = 1$ for every ν_k.

13.5 Theoretical impulse response

Comparing Equation 13.10 to Equations 5.19 or 11.13 suggests calling the vector

$$\boldsymbol{h} = (h_1, \ldots, h_M) \tag{13.12}$$

impulse response. This appellation is correct in one-point extrapolation. In longer extrapolation the name impulse response can be misleading, however, due to error accumulation. From now on we shall in any case use the name impulse response and call the prediction coefficients h_l by the name impulse response coefficients.

By *theoretical impulse response* we mean such a sequence of impulse response coefficients $h_1, \ldots, h_M$ that has been designed to predict correctly exponential (or sinusoidal) waves with

some predetermined wavenumbers. If the length M of the impulse response is chosen equal to the number of the exponential waves (or twice the number of sinusoids), then the coefficients are determined uniquely. There are different ways of finding these coefficients. We have already come across one of them, based on adding the wavenumbers one by one into the impulse response by repeated application of the summation formula of Equation 13.5.

Alternatively, the impulse response coefficients can be solved all at once. This approach leads to a set of linear equations. The M equations can be formed by writing the predictability condition, Equation 13.9, for each of the M wavenumbers ν_k that should be predictable. This gives M complex equations

$$\Delta x \sum_{l=1}^{M} h_l e^{-i2\pi\nu_k l\Delta x} = 1, \quad k = 1, \ldots, M \tag{13.13}$$

to solve the M unknown coefficients $h_1, \ldots, h_M$. Another strategy for forming the set of equations would be to use a continuous trial solution $h(x)$ with M adjustable parameters, and take the h_l-coefficients by sampling it. This approach was used in the original version of theoretical impulse response. A set of linear equations is obtained by stating $h(x)$ as a linear combination of the same exponential waves that should be predictable, or

$$h(x) = \sum_k \left[C_k^+ e^{i2\pi\nu_k x} + C_k^- e^{-i2\pi\nu_k x} \right]. \tag{13.14}$$

Substituting $h_l = h(l\Delta x)$ in Equation 13.13 leads to a set of linear equations for solving the $C_k^{\pm}$-coefficients. Note that the set of equations obtained is always complex. Equation 13.14 is intended for sinusoidal (possibly real) signal, since it requires predictability at symmetric wavenumber pairs $\pm\nu_k$.

Although calculating the $C_k^{\pm}$-coefficients and after that sampling the continuous impulse response $h(x)$ may seem unnecessarily complicated, this procedure has one important advantage: the number of the impulse response coefficients is no longer forced to equal the number of the wavenumbers ν_k. Instead, we can sample $h(x)$ much further, since the h_l-coefficients are not solved directly from the Equations 13.13, wherefore their number may differ from the number of equations. This is an important detail, since the frequency tuning method, presented in Section 13.12, has turned out to behave properly only if M is as large as possible. For solving the $C_k^{\pm}$-coefficients, see Problem 7.

13.6 Matrix method impulse responses

Given a set of measured, and preferably Fourier self-deconvolved, signal samples, the impulse response coefficients can also be determined by requiring that they predict correctly M selected, known samples. If these samples are successive, this requirement in the case of forward prediction can be written

$$I'_j = I_j, \quad j = J + M, \ldots, J + 2M - 1. \tag{13.15}$$

Inserting I'_j from Equation 13.1 leads to a set of M equations linear in the h_l-coefficients. Since this set can be dressed as a matrix equation, the sequence of h_l-coefficients it gives is called *matrix method impulse response*. Above I_J is the lowest-index sample used in the calculations, since the first one of the equations reads

$$\Delta x h_1 I_{J+M-1} + \cdots + \Delta x h_M I_J = I_{J+M}.$$

In spectroscopy, detector saturation effects or spectral background distortions tend to affect significantly the first few interferogram samples, making the use of nonzero J worthwhile.

If, as it usually is, the signal is not exactly predictable by M impulse response coefficients, the resulting coefficients may behave badly in extrapolation. Then it is a good idea to use more than M equations in the set of Equation 13.15. Naturally, an exact solution no longer exists, but we can find the best fit, minimizing the square sum of the errors $I'_j - I_j$. Thus, if we denote by N the total number of samples utilized in the calculations ($N = 2M +$ the number of excess equations), we have the minimization problem

$$\sum_{j=J+M}^{J+N-1} (I'_j - I_j)^2 = \text{min!} \tag{13.16}$$

Note always that $N \geq 2M$, since in addition to the samples predicted, at least M samples preceding them must also be known in order to compute the estimates I'_j.

In some cases, especially if the signal contains very nearby frequencies and M is small, the solution of the minimization problem, Equation 13.16, can be very sensitive to noise, since the absolute values of the h_l-coefficients obtained are very large. It may then be wise somehow to prevent the coefficients from obtaining too large absolute values. This can be done by adding a small penalty in the above functional, punishing according to the size of the impulse response. Using a suitably scaled square sum of the impulse response coefficients as the penalty functional yields the minimization problem

$$\sum_{j=J+M}^{J+N-1} (I'_j - I_j)^2 + \varrho \sum_{j=1}^{M} (\Delta x h_l)^2 = \text{min!} \tag{13.17}$$

The parameter ϱ can be used to adjust the relative importance of the size of the impulse response. Normally, a surprisingly small ϱ, such as 10^{-10}, is enough to bring down the excessive sizes of the h_l-coefficients. The effect of using nonzero ϱ has been illustrated by a simulated example in Figure 13.2. Instead of the h_l-coefficients, their inverse discrete Fourier

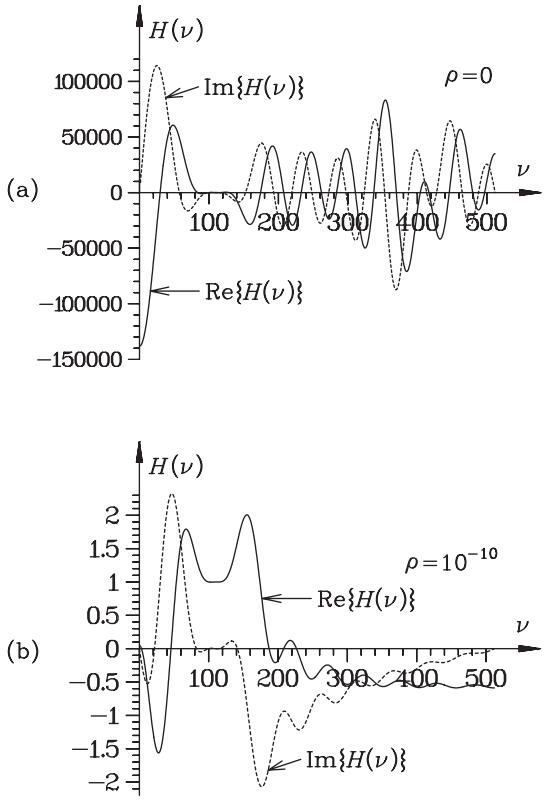


Figure 13.2: Even though $H(\nu)$ in plot (a) fulfills the predictability condition at the positions ν_k of the spectral lines, the corresponding h_l-coefficients behave badly when extrapolating a noisy signal. However, according to plot (b), a very small penalty coefficient ρ in the matrix method is enough to improve $H(\nu)$ considerably, without significantly affecting its values at the wavenumbers ν_k. [8]

transform $H(\nu)$ is drawn, since it is more illustrative. (Remember the predictability condition, Equation 13.9.) The signal used to compute the impulse response coefficients contained three noiseless cosine waves with the wavenumbers 100, 110, and 115, and the parameter values used were $M = 20$ and $N = 50$. As we can see from plot (a), without the ϱ-penalty $H(\nu)$ oscillates vigorously. If the signal contains noise at the wavenumbers where $|H(\nu)|$ is very large, the extrapolation can be totally spoiled. On the other hand, as may be seen from plot (b), as small penalty coefficient as 10^{-10} is enough to force $|H(\nu)|$ down, without considerably affecting the validity of the predictability condition.

Let us now assume that we want to obtain M impulse response coefficients by fitting the forward prediction condition as well as possible to $N - M$ signal samples. We then need to

know the N signal samples $I_J,\ldots,I_{J+N-1}$ in our attempt to solve the system of equations

$$\left\{\begin{array}{rcl} I_{J+M} & = & \Delta x h_1 I_{J+M-1} + \Delta x h_2 I_{J+M-2} + \cdots + \Delta x h_M I_J, \\ I_{J+M+1} & = & \Delta x h_1 I_{J+M} + \Delta x h_2 I_{J+M-1} + \cdots + \Delta x h_M I_{J+1}, \\ & \vdots & \\ I_{J+N-1} & = & \Delta x h_1 I_{J+N-2} + \Delta x h_2 I_{J+N-3} + \cdots + \Delta x h_M I_{J+N-M-1}\ . \end{array}\right.$$

This set of linear equations can be stated in matrix form as

$$\Delta x \begin{pmatrix} I_{J+M-1} & I_{J+M-2} & \cdots & I_J \\ I_{J+M} & I_{J+M-1} & \cdots & I_{J+1} \\ \vdots & \vdots & & \vdots \\ I_{J+N-2} & I_{J+N-3} & \cdots & I_{J+N-M-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{pmatrix} = \begin{pmatrix} I_{J+M} \\ I_{J+M+1} \\ \vdots \\ I_{J+N-1} \end{pmatrix}, \quad (13.18)$$

or in short

$$\Delta x \mathbb{I}\, \boldsymbol{h} = \boldsymbol{I}. \quad (13.19)$$

The solution of the least squares problem of Equation 13.16 then is

$$\boldsymbol{h} = \frac{1}{\Delta x} (\mathbb{I}^{\mathrm{H}} \mathbb{I})^{-1}\, \mathbb{I}^{\mathrm{H}} \boldsymbol{I}. \quad (13.20)$$

In the above, superscript H represents conjugate transposition; that is, the matrix is transposed and in case its elements are complex they also are complex conjugated.

When the size of the impulse response matters, we have to solve the minimization problem of Equation 13.17, whose solution is

$$\boldsymbol{h} = \frac{1}{\Delta x} \mathbb{B}_\varrho^{-1} \boldsymbol{y}\,, \quad (13.21)$$

where

$$\left\{\begin{array}{rcl} \mathbb{B}_\varrho & = & \mathbb{I}^{\mathrm{H}} \mathbb{I} + \varrho \mathbb{I}, \\ \boldsymbol{y} & = & \mathbb{I}^{\mathrm{H}} \boldsymbol{I}\,. \end{array}\right. \quad (13.22)$$

When the impulse response is determined by fitting the backward prediction condition to known samples, Equation 13.18 is replaced by the equation

$$\Delta x \begin{pmatrix} I^*_{J+N-M} & I^*_{J+N-M+1} & \cdots & I^*_{J+N-1} \\ I^*_{J+N-M-1} & I^*_{J+N-M} & \cdots & I^*_{J+N-2} \\ \vdots & \vdots & & \vdots \\ I^*_{J+1} & I^*_{J+2} & \cdots & I^*_{J+M} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{pmatrix} = \begin{pmatrix} I^*_{J+N-M-1} \\ I^*_{J+N-M-2} \\ \vdots \\ I^*_J \end{pmatrix}. \quad (13.23)$$

Note that the backward prediction equations have been complex conjugated so that impulse response coefficients appear nonconjugated. Now, if the matrix $\mathbb{I}$ and vector $\boldsymbol{I}$ are taken from Equation 13.23 instead of Equation 13.18, then Equations 13.20–13.22 remain unchanged.

13.7 Burg's impulse response

We call *Burg's impulse response* such a sequence of impulse response coefficients that is obtained by using *Burg's formula* (see below). Since Burg's formula is meaningful only in connection with so-called *Levinson–Durbin recursion*, we present it first. To this end let us assume that $I(x)$ is predictable and can be extrapolated to both directions without errors using the $m-1$ coefficients $h_1^{(m-1)}, \ldots, h_{m-1}^{(m-1)}$. We can thus insert $M = m-1$ and $I'_j = I_j$ in the backward prediction Equation 13.2 and rewrite it as

$$\Delta x \left(-\frac{1}{\Delta x} I_j + h_1^{(m-1)*} I_{j+1} + \cdots + h_{m-2}^{(m-1)*} I_{j+m-2} + h_{m-1}^{(m-1)*} I_{j+m-1} \right) = 0$$

or, substituting $j-m$ for j and reorganizing,

$$\Delta x \left(h_{m-1}^{(m-1)*} I_{j-1} + h_{m-2}^{(m-1)*} I_{j-2} + \cdots + h_1^{(m-1)*} I_{j-(m-1)} - \frac{1}{\Delta x} I_{j-m} \right) = 0.$$

The latter equation means that an attempt to predict the signal sample I_j forward by using the m coefficients $h_{m-1}^{(m-1)*}, h_{m-2}^{(m-1)*}, \ldots, h_1^{(m-1)*}, -\frac{1}{\Delta x}$ produces zero. Therefore, adding or subtracting these coefficients, multiplied by an arbitrary constant $\Delta x r$, to the original impulse response with $m-1$ coefficients, produces a new impulse response with m coefficients $h_1^{(m)}, \ldots, h_m^{(m)}$ defined by the equations

$$\left\{ \begin{array}{rcl} h_1^{(m)} & = & h_1^{(m-1)} - \Delta x r h_{m-1}^{(m-1)*}, \\ h_2^{(m)} & = & h_2^{(m-1)} - \Delta x r h_{m-2}^{(m-1)*}, \\ & \vdots & \\ h_{m-1}^{(m)} & = & h_{m-1}^{(m-1)} - \Delta x r h_1^{(m-1)*}, \\ h_m^{(m)} & = & r\,. \end{array} \right. \tag{13.24}$$

This procedure for increasing the length of the impulse response by one is known as Levinson–Durbin recursion. The arbitrary coefficient r, in turn, is known as the *reflection coefficient*. As the bottommost equation shows, it equals the last coefficient in the new, extended impulse response.

If the impulse response coefficients are able to extrapolate the signal without errors, then the Levinson–Durbin recursion, Equation 13.24 presented above, yields another, longer series of coefficients, that is capable of the same. The normal use of the recursion is to start from only one impulse response coefficient and add the number of coefficients one by one, in the hope of eventually obtaining an acceptable impulse response. Since very few signals can be predicted acceptably with only one coefficient, the coefficients are very bad initially. The choice of how the reflection coefficients are computed is then crucial to the outcome of the iteration.

In computing the reflection coefficients, Burg's formula has proven very effective. It reads

$$r = h_m^{(m)} = \frac{2}{\Delta x} \frac{\sum_{n=m}^{N-1} e_n^{(m-1)} b_{n-1}^{(m-1)*}}{\sum_{n=m}^{N-1} \left[|e_n^{(m-1)}|^2 + |b_{n-1}^{(m-1)}|^2 \right]}, \tag{13.25}$$

where again N is the total number of signal samples utilized in the computations. Now, however, it is enough that $N > M$. The numbers $e_n^{(m-1)}$ and $b_n^{(m-1)}$ are the prediction residuals when predicting I_n forward or I_{n-m+1} backward by using the impulse response coefficients $h_1^{(m-1)}, \ldots, h_{m-1}^{(m-1)}$. They are also easily obtained recursively from $e_n^{(m-2)}$ and $b_{n-1}^{(m-2)}$ as

$$\begin{cases} e_n^{(m-1)} &= I_n - I'^{(m-1)}_n = I_n - \Delta x \sum_{l=1}^{m-1} h_l^{(m-1)} I_{n-l} \\ &= e_n^{(m-2)} - \Delta x h_{m-1}^{(m-1)} b_{n-1}^{(m-2)}, \\ b_n^{(m-1)} &= I_{n-m+1} - I'^{(m-1)}_{n-m+1} = I_{n-m+1} - \Delta x \sum_{l=1}^{m-1} h_l^{(m-1)*} I_{n-m+1+l} \\ &= b_{n-1}^{(m-2)} - \Delta x h_{m-1}^{(m-1)*} e_n^{(m-2)}. \end{cases} \tag{13.26}$$

The Levinson–Durbin recursion and Burg's formula should be applied repeatedly, until an impulse response with the desired length M is obtained. The procedure can be described as follows:

Step 1: Decide the length M of the desired impulse response, and the number $N > M$ of signal samples used in computing it. Set $m = 1$ and $e_n^{(0)} = b_n^{(0)} = I_n$, $n = 0, \ldots, N-1$.

Step 2: Compute $h_m^{(m)}$ from the Burg's formula, Equation 13.25. If $m > 1$, find the other components $h_1^{(m)}, \ldots, h_{m-1}^{(m)}$ by applying Levinson–Durbin recursion, given by Equation 13.24.

Step 3: If $m = M$, accept the latest impulse response coefficients and stop. Otherwise increase m by 1, compute $e_n^{(m-1)}$ and $b_n^{(m-1)}$ with $n = m-1, \ldots, N-1$ from Equation 13.26, and return to Step 2.

13.8 The q-curve

Above, impulse response coefficients given by Burg's formula and the matrix methods were derived by fitting the prediction equations, Equations 13.1 and 13.2, into measured signal samples. In matrix methods this fit behaves "normally" in the sense that the more samples are fitted, the better will the obtained coefficients perform in extrapolating the signal, if M is fixed. However, for Burg's impulse response coefficients this is not quite true. Besides, in both Burg's method and matrix methods only one-point extrapolation was used in the fitting

in order to get equations linear in the h_l-coefficients. This does not guarantee the success of longer extrapolations, where already predicted samples are used again to predict subsequent samples. Thus, some method to evaluate the ability of given impulse response coefficients to extrapolate a given signal would be very useful. A natural test would be to use the coefficients actually to extrapolate a known piece of the signal, and to compare the output with the correct signal samples by computing the square sum of the differences. In order to avoid dependence on simple scaling of the signal, this square sum should be proportioned to the square sum of the correct samples. This kind of test is the quality factor, or q-factor.

A very useful version of the q-factor, applicable to real spectra (whereupon the signal is conjugate symmetric, or $I_{-j} = I_j^*$), is to extrapolate the signal backward across the origin to negative x-values. This has been demonstrated for a real signal in Figure 13.3. By using this approach, the actual measured samples $I_0, I_1, \ldots$ need not be extrapolated, and therefore remain free to be used in the calculation of the h_l-coefficients. The exact definition is

$$q = 1 - \sqrt{\frac{\sum_{j=J}^{N_0-1} |\hat{I}_{-j} - I_j^*|^2}{\sum_{j=J}^{N_0-1} |I_j|^2}}. \tag{13.27}$$

Here N_0 is some fixed index that determines the last sample tested. It does not need to be the same as the number N of samples used in computing the impulse response coefficients. The extrapolated sample estimates are denoted as $\hat{I}_{-j}$ to make a difference with the one-point extrapolations I'_j, computed from correct samples. The q-factor is a function of M (number of impulse response coefficients) and N (number of known samples used to compute the coefficients), or $q = q(M, N)$. For that reason, it can be used to optimize these numbers. Usually the difference $N - M$ is fixed, and either $q(N)$ or $q(M)$ is drawn, giving the *q-curve*. The better are the coefficients $h_1, \ldots, h_M$, the larger is the q-factor, the maximal value being one.

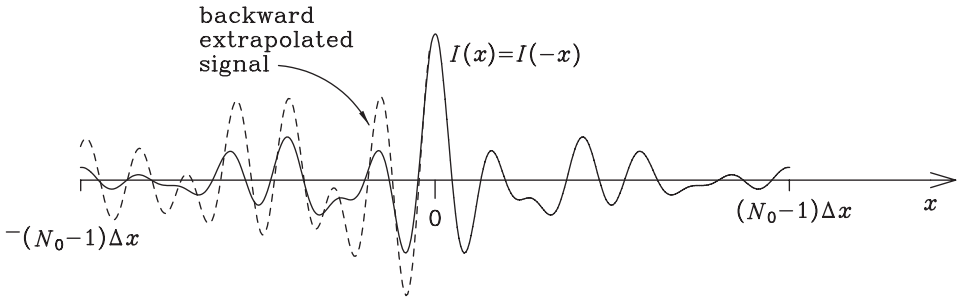


Figure 13.3: The q-factor is computed by extrapolating the signal backward across the origin and comparing the extrapolation with the mirror image of the original signal.

13.9 Spectral line narrowing by signal extrapolation

In spectroscopy a problem encountered again and again is the overlapping of spectral lines that are too wide compared with their interspaces. Therefore, an ability to narrow spectral lines artificially would be extremely useful. For a line spectrum, such a line narrowing is equivalent to resolution enhancement. Let us now ponder what such a line narrowing means in the signal domain. We assume that all the spectral lines have (at least approximately) identical functional line shape, with only line heights and positions varying.

In fact, line narrowing is the inverse of smoothing. In smoothing, all the spectral details are spread so that every sharp detail is replaced by a continuous distribution concentrating in the vicinity of the original detail. The purpose of smoothing is to suppress high-frequency spectral noise, but an inevitable side effect is that the spectral lines become wider. This kind of operation is clearly convolution by a function $W(\nu)$ that decays away from the origin. Therefore, in the signal domain smoothing appears as simple multiplication (apodization) by an also decaying truncation function. Thus spectral smoothing damps the signal, or makes it narrower, and the inverse operation is therefore to enhance the signal far from the origin. The same conclusion can also be arrived at from the similarity theorem (see Equation 2.9): When the spectrum is narrowed by a coefficient a, the signal is widened by the same coefficient. (See also Problem 6.10.)

We have already encountered one method of widening the signal, the FSD. However, FSD is not able to increase the width of the signal unlimitedly. This limitation of FSD comes, of course, from the fact that it cannot extend the signal beyond its point of truncation (see again Figure 12.6). Signal extrapolation, however, offers us a means to overcome this limitation.

Let us assume that at ν_k there is a spectral line with area A_k. The original, recorded line shape function, common to all the spectral lines, is assumed to be $W_\mathrm{R}(\nu)$. The partial spectrum containing only the line at ν_k is therefore

$$E_k(\nu) = A_k W_\mathrm{R}(\nu - \nu_k),$$

and the corresponding partial signal is the constituent wave

$$I_k(x) = \mathcal{F}\{E_k(\nu)\} = A_k \mathcal{F}\{W_\mathrm{R}(\nu - \nu_k)\} = A_k e^{i2\pi\nu_k x} \mathcal{F}\{W_\mathrm{R}(\nu)\} = A_k e^{i2\pi\nu_k x} w_\mathrm{R}(x),$$

where $w_\mathrm{R}(x) = \mathcal{F}\{W_\mathrm{R}(\nu)\}$. In line narrowing we always have several overlapping spectral lines, and the partial spectrum under examination is a linear combination (compare Equation 12.1)

$$E(\nu) = \sum_k E_k(\nu) = \sum_k A_k W_\mathrm{R}(\nu - \nu_k), \tag{13.28}$$

corresponding to the partial signal

$$I(x) = w_\mathrm{R}(x) \sum_k A_k e^{i2\pi\nu_k x}. \tag{13.29}$$

Let us assume that the original line shape $W_\mathrm{R}(\nu)$ is known (for instance thanks to a singlet line near to the interesting spectral region), and only the number of lines, as well as their

heights and positions, are to be determined. It would in principle be possible to fit even the line shape by using a parametric line shape model, but this would make the problem very ill-behaved. Even slightest errors in the input samples would be enough to completely muddle the output. The most straightforward strategy of line narrowing would be to fit the model given by Equation 13.28 directly in the spectral domain. Our strategy, however, is to fit the model of Equation 13.29 in the signal domain. In the signal domain the amplitude function $w_{\mathrm{R}}(x)$ is separated to a common factor and it can be manipulated without knowing the coefficients A_k and wavenumbers ν_k, or even the number of spectral lines. This is what made FSD so easy and reliable a method. Likewise, when extrapolating the waves by using linear prediction, the coefficients A_k do not affect the impulse response coefficients, as we have found earlier. Line positions ν_k do affect the h_l-coefficients, anyway. However, it is possible to deduce from the line narrowed line shapes whether the corresponding waves $A_k e^{i2\pi\nu_k x}$ were extrapolated correctly by the impulse response coefficients used.

In Figure 13.4 there is a series of simulations of the line shape distortions when one of the wavenumbers that the h_l-coefficients can extrapolate, marked by ν_{G}, is varied in the vicinity of a true line position $\nu_k = 250$. The extrapolated signal is so apodized that the output line shape should be symmetric Besselian (see Figure 12.6 E). As we can see, when $\nu_{\mathrm{G}} \neq \nu_k$, the output line shape is distorted in such a way that it has deeper sidelobe on the side that is away from the true line position. In case of a multiline spectrum, this kind of behavior is valid for each line separately. This information contained in the line shape distortions will be utilized in the frequency tuning method, to be presented in Section 13.12.

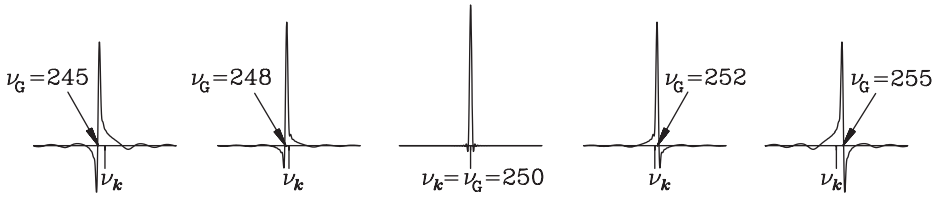


Figure 13.4: Distortions in the symmetry of a narrowed spectral line can be used to search the exact line position. [9]

13.10 Imperfect impulse response

Now we will briefly consider the realistic situation where the predictability condition, Equation 13.9, or equivalently, Equation 13.8, is not valid for every wavenumber ν_k present in the signal. We begin by an impulse response with only one nonzero coefficient h_1. If we write $\Delta x h_1$ in the form

$$\Delta x h_1 = |\Delta x h_1|\, e^{i2\pi\nu_0\Delta x} = \exp\left(i2\pi\left[\nu_0 - i\ln\left(|\Delta x h_1|\right)/(2\pi\,\Delta x)\right]\Delta x\right),$$

a comparison with Equation 13.3 immediately reveals that this impulse response coefficient is able to extrapolate correctly only an exponential wave with the complex wavenumber

$$\nu_G = \nu_0 - i \ln(\Delta x\,|h_1|)/(2\pi\,\Delta x).$$

(If $\Delta x\,|h_1| \neq 1$, this wave is either decaying or growing.) Now, if h_1 is used to extrapolate a given signal, it interprets the last known sample I_{N-1} as a sample from an exponential wave constant$\times \exp(i2\pi\,\nu_G x)$ and extrapolates it correspondingly, quite regardless of the behavior of the signal prior to the sample I_{N-1}.

In Figure 13.5 we see a simulated example of what happens when the imaginary part of the wavenumber ν_G, which the h_l-coefficients are able to extrapolate, is varied. The signal extrapolated is a cosine wave, but it can also be interpreted as the real part of a single nondecaying exponential wave. In the bottom row, ν_G coincides with the wavenumber ν_k of the cosine wave. In the midmost row, $\mathrm{Re}\{\nu_G\} = \nu_k$, but $\mathrm{Im}\{\nu_G\} > 0$, and on the top row $\mathrm{Re}\{\nu_G\} = \nu_k$, but $\mathrm{Im}\{\nu_G\} < 0$. Before taking inverse Fourier transform, the extrapolated wave was apodized by squared parable to obtain a smoother, Besselian line shape (see Figure 12.6 E). Note that when computing a theoretical impulse response, there is no danger that ν_G would contain an unwanted imaginary part. However, especially when the signal is extrapolated by a matrix method impulse response, it often happens that some partial waves decay or increase.

Given M nonzero impulse response coefficients, the complex polynomial of the complex variable $\exp(-i2\pi\,\nu\Delta x)$,

$$1 - \Delta x \sum_{l=1}^{M} h_l \left(e^{-i2\pi\,\nu\Delta x}\right)^l,$$

is of degree M and thereby has M distinct roots $\exp(-i2\pi\,\nu_{Gk}\Delta x)$, excluding the very unusual cases of multiple roots. According to Equation 13.8, the wavenumbers $\nu_{G1}, \ldots, \nu_{GM}$ corresponding to these roots are such that the h_l-coefficients are able to extrapolate them without errors. When the extrapolation starts after the last reliable known signal sample I_{N-1}, these impulse response coefficients need the preceding M measured samples $I_{N-1}, \ldots, I_{N-M}$, and interpret them as samples from the group of waves

$$\sum_{k=1}^{M} C_k e^{i2\pi\,\nu_{Gk}x}, \tag{13.30}$$

where C_j are complex constants. Thus, beginning from the sample I_{N-M}, the extrapolated signal can be interpreted as a linear combination of waves with the wavenumbers $\nu_{G1}, \ldots, \nu_{GM}$ only. If $N = M$, there are no prior samples left to question this interpretation. However, if $N > M$, the signal samples $I_0, \ldots, I_{N-M-1}$ are not used in computing the extrapolated samples, and if there were originally other wavenumbers present in the signal than $\nu_{G1}, \ldots, \nu_{GM}$, this situation is also visible in the line narrowed spectrum. This information is exploited in the frequency tuning method to improve the impulse response coefficients, until the wavenumbers $\nu_{G1}, \ldots, \nu_{GM}$ coincide with the true wavenumbers ν_k contained in the signal.

The above reasoning about the information content in the signal can also be performed in the spectral domain as follows. Let us assume that we have N reliable signal samples. Then

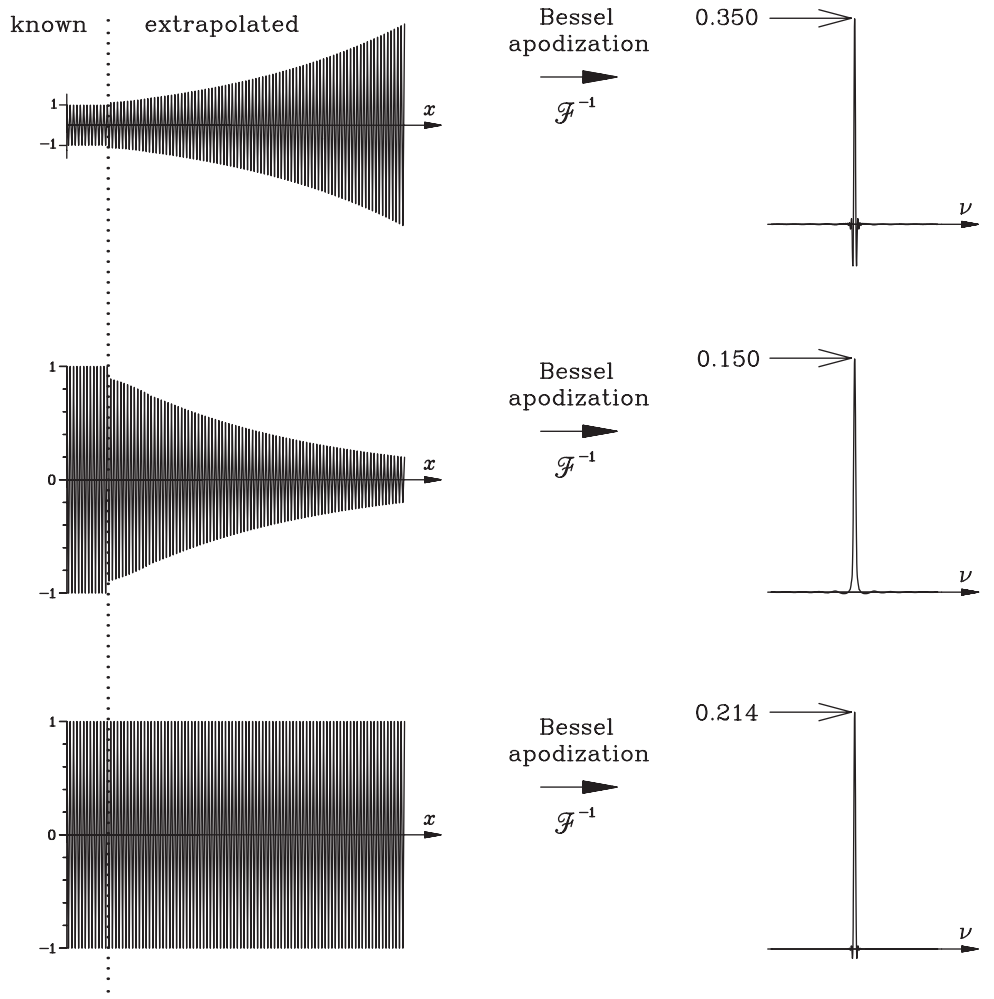


Figure 13.5: Impulse response coefficients extrapolate any signal as if it were one of the signals the coefficients are able to extrapolate correctly. In case of a complicated signal, the correctness of the extrapolation has to be deduced from the output line shape or by computing the q-factor. [9]

the corresponding spectrum also contains N independent spectral samples $E_0, \ldots, E_{N-1}$. Now, since we assume the functional form of the spectral line shape fixed, every line has only two parameters associated in it, for example, height and position. When the impulse response coefficients are given, they will always extrapolate the signal as a linear combination of only those wavenumbers ν_{Gk} that they are able to extrapolate. Thus the line positions in the output spectrum are fixed, and only the heights of the lines can be used to explain the spectrum. Therefore, in general, N spectral lines are needed to explain completely N arbitrary spectral

samples. If, however, we are able to select the h_l-coefficients freely, then the wavenumbers ν_{Gk} they are able to extrapolate can also be selected freely by computing a suitable theoretical impulse response. Then both the line positions and heights can be adjusted, and at most $N/2$ (round fractions up) lines of a given shape are needed to explain completely a sequence of N spectral samples. However, if the maximal number of spectral lines is used in explaining the measured spectral samples, there is no residual spectrum left over to check the reliability of the line narrowed spectrum.

In practice we can assume that our signal has no phase errors, or that the C_j-coefficients in the signal of Equation 13.30, corresponding to the Fourier self-deconvolved spectrum, are real. Therefore, if the extrapolated signal is apodized (multiplied) by a symmetric real function $w(x)$, the ensuing, narrow line shape $W(\nu)$ is real and symmetric. This information is also utilized in the frequency tuning method (see also Figure 13.4).

Figure 13.6 shows a simulated, noiseless example of how the distortions in the line narrowed spectrum can be utilized to improve the impulse response, and thereby the line narrowing itself. Plot (a) shows the starting point, the spectrum $E(\nu)$ after FSD. In the signal domain only 10 samples were retained, so that the line shape is $19\Delta x\,\mathrm{sinc}(\pi 19\Delta x\nu)$. The number of spectral lines is three, their heights are equal, and their true positions are marked by vertical dashed lines. The black pins, in turn, indicate the wavenumbers ν_{Gk} that the impulse response coefficients are able to extrapolate. In order that these wavenumbers could be adjusted (tuned) freely, the theoretical impulse response was always used. Its length was always maximal, or $M = 10$, because this choice has proved to most effectively reveal unattended lines in the line-narrowed spectrum. In plot (b) we see the outcome of the line narrowing, when the first line coincides with one of the predictable wavenumbers, but there is only one predictable wavenumber between the positions of the second and third lines. As we can see, the second and third lines are represented by only one line, standing on a broad and shallow hump. The shape of the line is clearly asymmetric, and the missing line has caused slight asymmetries even to the first line, even though its wavenumber is predictable. In plot (c), the two predictable wavenumbers have been tuned so that the line shapes are perfectly symmetric. The missing line is still visible as a hump. Finally, plot (d) shows the outcome when one new predictable wavenumber has been added somewhere inside the hump, and all the three predictable wavenumbers have been tuned until all the spectral lines are perfectly symmetric. The result is an output spectrum with narrowed lines at exactly correct positions and with correct relative heights. Since the length of the signal after extrapolation was 15 times its starting length, the lines are narrowed by approximately the same coefficient. The exact narrowing coefficient depends on the output line shape $W(\nu)$ selected. Here Besselian $W(\nu)$ was used. The operations displayed in Figure 13.6 are similar to those performed in the frequency tuning method.

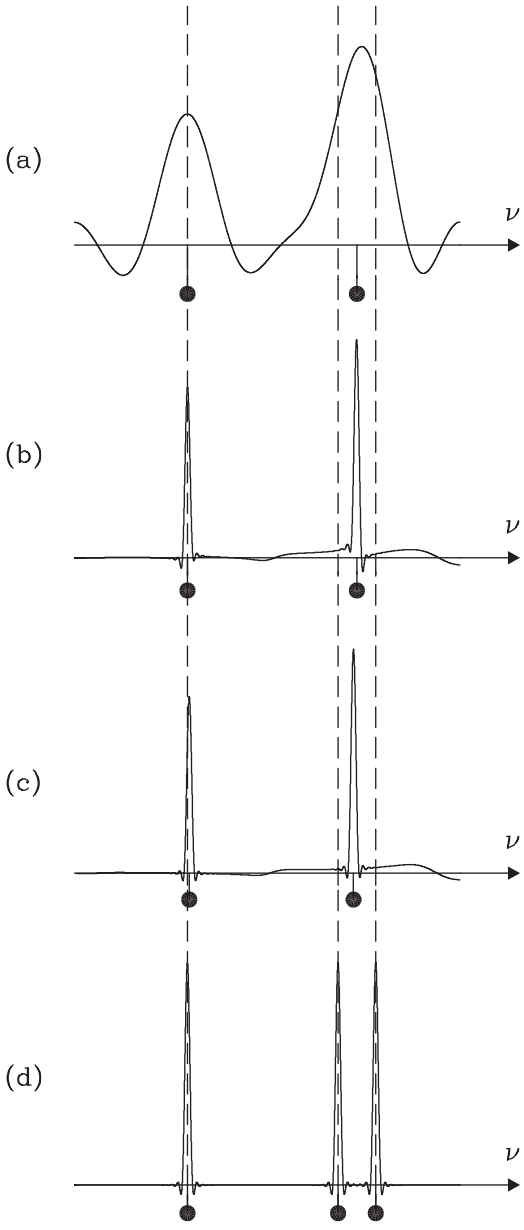


Figure 13.6: When using theoretical impulse response, the wavenumbers it can extrapolate (black pins) can be improved according to the spectral distortions, until they coincide with the true line positions (vertical dashed lines).

13.11 The LOMEP line narrowing method

The principle of the LOMEP (Line shape Optimized Maximum Entropy linear Prediction) line narrowing method is illustrated in Figure 13.7 by showing what happens in the signal domain to one constituent wave $I(x)$ (corresponding to a symmetric pair of spectral lines at $\pm\nu_k$). We assume that the original line shape function $W_{\mathrm{R}}(\nu)$, and thereby the amplitude function $w_{\mathrm{R}}(x)$ of the constituent waves, is known. In addition we assume that the signal is real and symmetric. Thus, it is enough to consider only the positive halves of the signal ($x \geq 0$) and spectrum ($\nu \geq 0$). The LOMEP advances in the following steps:

Step 1: Separate the interesting part of the spectrum. Extend it symmetrically to the negative ν-values to make the corresponding partial signal real. Fourier transform this partial spectrum to obtain the partial signal.

Step 2: Divide the partial signal by $w_{\mathrm{R}}(x)$. (This is *deapodization* in the signal domain, or FSD in the spectral domain.) Now, every constituent wave is a nondecaying cosine wave, as displayed in Figure 13.7, and therefore can be extrapolated, if the number of the impulse response coefficients is at least twice the number of spectral lines. Since the last deapodized signal samples are usually very noisy, you should now also select the index of the first signal sample to be extrapolated.

Step 3: Fix $N = M + 1$, and compute the q-factor as a function of M or N. Every q-value $q(M)$ computed requires the computation of a sequence of M new impulse response coefficients, and an extrapolation to obtain the estimates $\hat{I}_{-j}$ needed in calculating the q-factor. The impulse response coefficients are computed by using Levinson–Durbin recursion and Burg's formula.

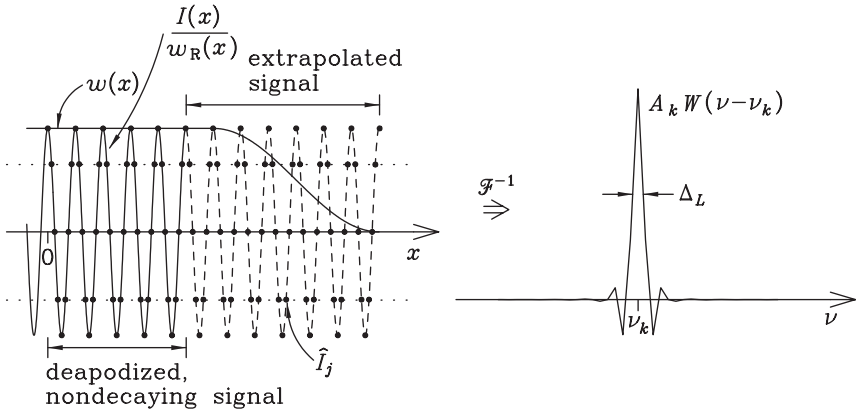


Figure 13.7: Spectral line narrowing by signal extrapolation. After the original line shape function $W_{\mathrm{R}}(\nu)$ has been removed by FSD, one spectral line corresponds to one nondecaying cosine wave, which can be extrapolated. Apodizing the extrapolated wave by $w(x)$ leads to a new, narrower line shape $W(\nu)$.

Step 4: Accept the impulse response coefficients that provided the maximal q-factor. Use them to extrapolate the signal as far as you wish.

Step 5: Extend the extrapolated partial signal symmetrically to negative x-values. Then apodize (multiply) it by some symmetric, smoothly decaying amplitude function $w(x)$.

Step 6: Compute inverse Fourier transform of the extrapolated and apodized partial signal. This gives the spectrum with the new line shape $W(\nu) = \mathcal{F}^{-1}\{w(x)\}$. The FWHM of this new line shape – marked as Δ_L in Figure 13.7 – is smaller than the original.

In Figure 13.8 the algorithm, excluding the optimization of the impulse response in Step 3, has been demonstrated by a noiseless simulated partial spectrum with four lines.

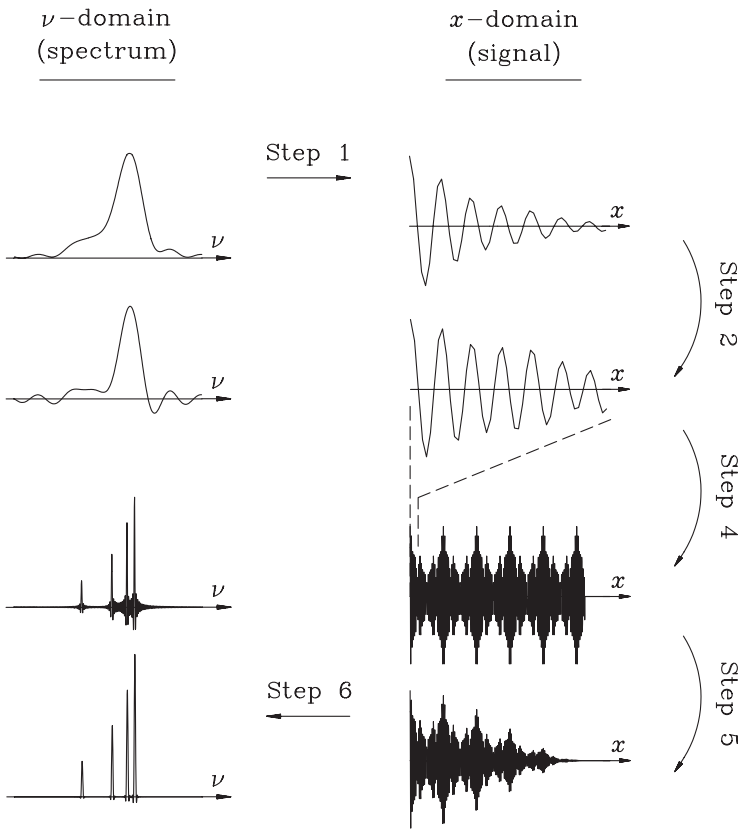


Figure 13.8: Principle of spectral line narrowing by signal extrapolation. The step numbers refer to the LOMEP algorithm. Because of symmetry, only the positive halves of the x- and ν-axes are drawn.

Note that often the spectral line width is proportional to the line position, so that the spectral portion examined should be relatively narrow to minimize the differences in line widths. If the original line shape is asymmetric, then in Step 1 the symmetrization of the spectrum should not be done, since in Step 2 FSD is replaced by Fourier complex self-deconvolution (FCSD). To compensate this, the possible imaginary part of the deapodized partial signal is zeroed in the end of Step 2.

In separating the partial spectrum, whose lines are to be narrowed, you should select it so that the derivatives of the spectrum at the ends of this spectral segment are as near to zero as possible. The purpose is to avoid spurious peaks when symmetrizing the partial spectrum in Step 1.

13.12 Frequency tuning method

The frequency tuning method, or the gulf tuning method, is an iterative line narrowing procedure that is designed to make maximal use of the information contained in the distortions in an incorrectly line narrowed spectrum. Like LOMEP, this method requires that after FSD the partial signal consists of perfect cosine waves without phase errors. Then, if the extrapolated and apodized signal is extended symmetrically to negative x-values before computing inverse Fourier transform, the output line shape should be perfectly symmetric. Asymmetries in the line shapes are a token of discrepancies between true line positions and the wavenumbers that the impulse response is able to extrapolate (see Figure 13.4). Likewise, missing waves, not extrapolated at all, can be detected from the output spectrum, if the number of original spectral samples N is greater than the number M of the impulse response coefficients (see Figure 13.6). Thus, monitoring the line narrowed spectrum continuously and improving the h_l-coefficients respectively provides us with a means to improve iteratively the quality of the line narrowing.

While in LOMEP the signal extrapolation was carried out by using impulse response coefficients given by the Levinson–Durbin recursion and Burg's formula, the frequency tuning method utilizes the theoretical impulse response. Thus, the line positions can also be fitted explicitly. The theoretical impulse response should always be as long as possible, which means that M should equal the number of reliable original signal samples. In this way, the missing lines are most clearly visible in the spectrum.

To begin with, trial line positions are placed at spectral positions, where there is at least one nearby line for certain. A theoretical impulse response is then constructed, able to extrapolate just these wavenumbers. These impulse response coefficients are then used to enhance the spectral resolution by extrapolating the deapodized (partial) signal, like in LOMEP. Now the output line shapes reveal whether the trial line positions coincide with the true line positions. The trial line positions (known as *gulfs*) are then tuned, until they coincide with the true ones, which situation is revealed by perfectly symmetric line shapes. Thus, instead of maximizing a single q-factor, every wavenumber is fitted separately. This feature makes the frequency tuning method insensitive to large numbers of spectral lines.

When such wavenumbers have been found that all the lines in the line narrowed spectrum are perfectly symmetric, the output spectrum should be searched through for missing spectral lines. If the original spectrum contains lines that have not yet been taken into account, they leave small humps in the output spectrum. New trial line positions are then added at the most prominent humps, and new impulse response coefficients are computed that are able to extrapolate these wavenumbers in addition to the old ones.

It is clear from what is mentioned above, that the frequency tuning method is a highly graphical and interactive of its nature. The frequency tuning itself can be mostly automatized, but after each tuning the user has to decide where there are unattended lines, and set the starting values of their wavenumbers. To help the user analyze the output spectrum, the suitably scaled residual spectrum, containing only the background with the spectral lines deprived, should also be drawn. The residual spectrum is obtained by subtracting a line with the predetermined shape $W(\nu)$ at each trial line position $\nu_{\mathrm{G}k}$. The height of the line subtracted should equal the value of the line narrowed spectrum at $\nu_{\mathrm{G}k}$. Thus, the value of the residual spectrum is zero at each line position. In Figure 13.9 the frequency tuning method has been applied to a simulated noise-free test spectrum with four lines. The residual spectrum, in a suitable scale, has been drawn by a dashed line where useful. New gulfs are added and tuned, until all the lines are symmetric and the background (residual) spectrum is smooth or, in case of a noiseless signal, disappears.

The frequency tuning procedure can be described by the step-by-step algorithm given below. Note that Steps 2 and 3 are practically identical to Steps 1 and 2 in LOMEP above. Vector $\boldsymbol{\nu}$ is used to store the current estimates $\nu_{\mathrm{G}k}$ of the wavenumbers ν_k contained in the interesting piece of spectrum. Note that always $\nu_{\mathrm{G}k} > 0$. In every extrapolation such a theoretical impulse response is always used that is able to extrapolate cosine waves with all the wavenumbers in $\boldsymbol{\nu}$. This means that the h_l-coefficients must be able to extrapolate all the pairs of wavenumbers $\pm\nu_{\mathrm{G}k}$.

Step 1: Separate the interesting piece of spectrum. This piece is called the partial spectrum. Choose the first estimates $\nu_{\mathrm{G}k}$ for spectral line positions ν_k and store them in vector $\boldsymbol{\nu}$. Do not define too many $\nu_{\mathrm{G}k}$ at once. They are easy to add later. Note that often the spectral line width is dependent on the line position, so that the partial spectrum should be relatively narrow to minimize the differences in line widths.

Step 2: If the original line shape $W_{\mathrm{R}}(\nu)$ is symmetric, extend the partial spectrum symmetrically to the negative ν-values. Fourier transform the partial spectrum to obtain the corresponding partial signal.

Step 3: Deapodize (divide) the partial signal by $w_{\mathrm{R}}(x) = \mathcal{F}\{W_{\mathrm{R}}(\nu)\}$. If $w_{\mathrm{R}}(x)$ and the partial signal were complex, the imaginary part of the deapodized partial signal should be zeroed. Now every constituent wave is a nondecaying cosine wave and can be extrapolated, if the number of the impulse response coefficients is at least twice the number of spectral lines. Since the last deapodized signal samples are usually very noisy, you should now also select the index of the first signal sample to be extrapolated.

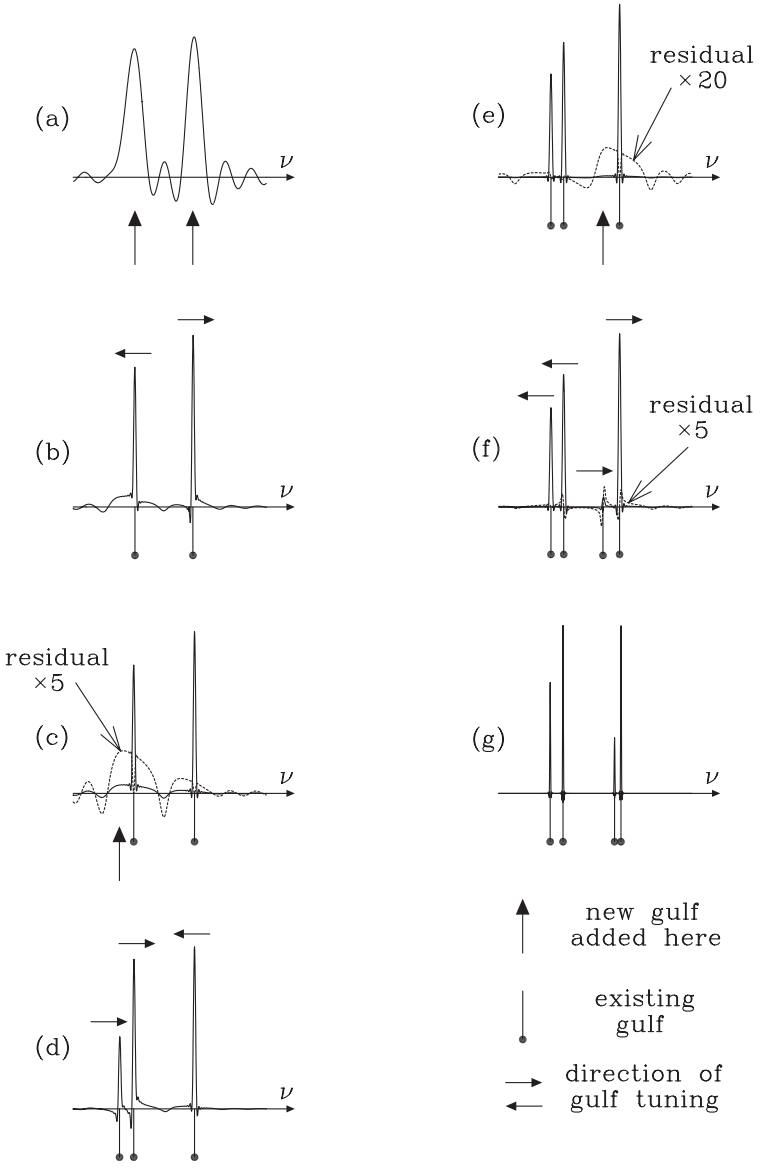


Figure 13.9: The frequency tuning method consists of repeated cycles (a) – (g), where new gulfs are added to the most prominent humps and then tuned, until the narrowed spectral lines are perfectly symmetric. The cycles are terminated when there are no more humps left. A gulf wavenumber ν_G is such a wavenumber that the transfer function $H(\nu_G) = H(-\nu_G) = 1$.

Step 4: By using the model of Equation 13.14, compute the theoretical impulse response coefficients h_l, able to extrapolate all the wavenumbers stored in vector $\boldsymbol{\nu}$. Irrespective of the number of these wavenumbers, choose M to be equal to the number of reliable samples in the partial signal.

Step 5: Extrapolate the deapodized partial signal by using the current h_l-coefficients. Extend the extrapolated partial signal symmetrically to negative x-values. Then apodize (multiply) it by some symmetric, smoothly decaying amplitude function $w(x)$.

Step 6: Compute inverse Fourier transform of the extrapolated and apodized partial signal.

Step 7: If all the spectral lines (or the residual spectrum at every trial line position $\nu_{\mathrm{G}k}$) are symmetric, go to Step 9. Otherwise proceed to Step 8.

Step 8: Tune each wavenumber $\nu_{\mathrm{G}k}$ away from the deeper sidelobe of the corresponding spectral line. Equivalently, calculate the derivative of the output spectrum at each trial line position $\nu_{\mathrm{G}k}$. (This can be done rapidly by use of the derivative theorem and discrete Fourier transform.) If at $\nu_{\mathrm{G}k}$ the derivative of the spectrum is positive, increase $\nu_{\mathrm{G}k}$, and vice versa. By using this tuning rule it is easy to make a computer program to tune the wavenumbers automatically. Return to Step 4.

Step 9: If the residual spectrum is smooth, stop. The output spectrum is the final line narrowed spectrum. Otherwise proceed to Step 10.

Step 10: Find approximately the centers of the most prominent humps in the spectrum. Add their positions in vector $\boldsymbol{\nu}$. Return to Step 4.

In Figures 13.10 and 13.11 we see the outcome of applying the frequency tuning method to an experimental $^{\mathrm{P}}\mathrm{Q}_9$ branch of the ν_5 rovibrational band of $^{12}\mathrm{CH}_3\mathrm{I}$. The original Q branch is plotted on the top of Figure 13.10, and also the singlet used to determine $w_{\mathrm{R}}(x)$ is indicated. The signal amplitude function $w_{\mathrm{R}}(x)$ was searched by Fourier transforming the singlet alone, and fitting a parametric line shape model to the envelope of the transform. The line narrowed Q branch is plotted at the bottom of Figure 13.10. The signal was extrapolated to 30 times its original length. In Figure 13.11 there is an enlarged image of the line narrowed Q branch. The new line shape is Besselian, because the apodization function $w(x)$ used was a squared parable (see Figure 12.6 E).

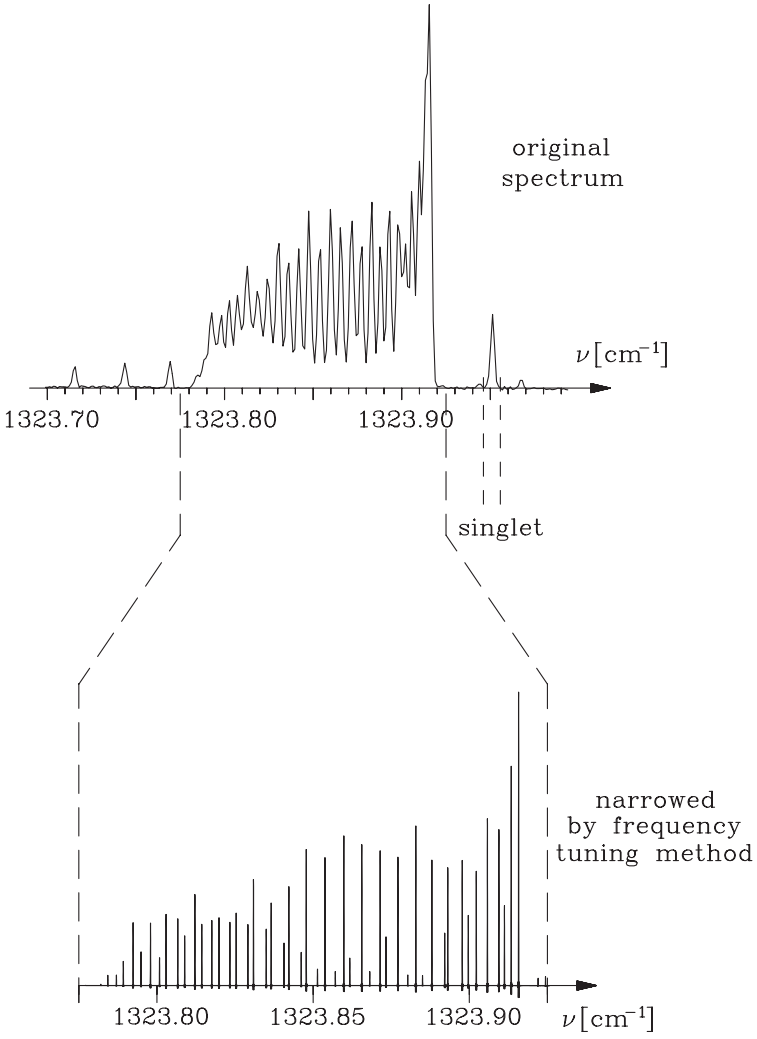


Figure 13.10: The frequency tuning method applied to a piece of a high-resolution FT-IR spectrum of $^{12}CH_3I$. The input line shape was taken from the singlet line indicated in the topmost plot.

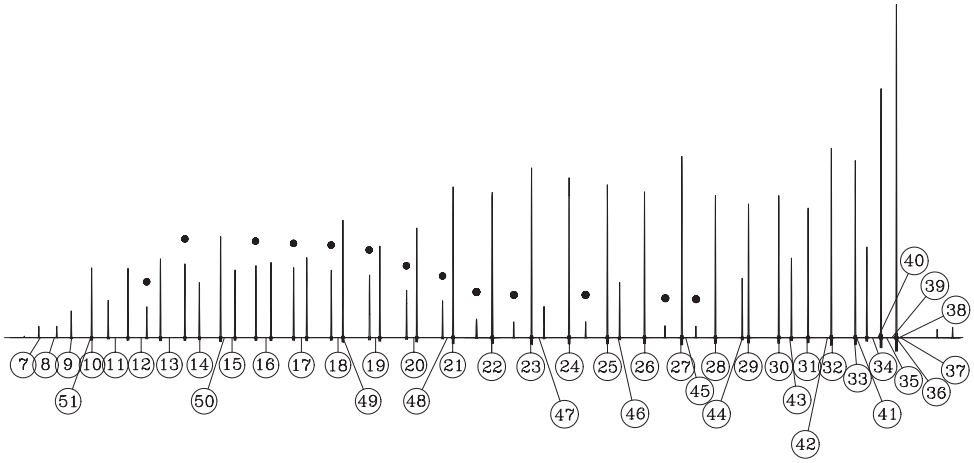


Figure 13.11: An enlarged plot of the line narrowed Q branch in Fig. 13.10. The theoretical line positions, obtained from literature, are indicated by pins containing the corresponding J quantum number, and black dots denote unidentified lines.

13.13 Other applications

In addition to line narrowing, there are also other spectroscopic applications of linear prediction. For the time being, these applications have been tested only very superficially, and therefore only the general principles are presented here. Remember that even though in line narrowing the constituent waves were always made nondecaying by FSD, decaying waves can also be extrapolated if their amplitude curves are suitable.

In Figure 13.12 we see the principle of spectral noise filtering by using linear prediction in the signal domain. The noisy spectrum $E(\nu)$ (only a partial spectrum with one line is drawn here) is first Fourier transformed. Then the end of the signal $I(x)$, dominated by noise, is replaced by new signal samples extrapolated from the more reliable samples in the beginning of the signal. Finally, taking inverse Fourier transform gives the noise-filtered spectrum $E'(\nu)$. Note that the starting point of the extrapolation, marked as X_{extr} in the figure, can be selected much earlier than the optimal point of truncation X_i, if the noise filtering were made by simple signal truncation. This same principle can also be used to correct instrumental distortions, arising from signal truncation during registration, as illustrated in Figure 13.13.

It may be likewise useful to extrapolate shorter signal segments. Spectral background and interference originate from a relatively short distorted signal portion, as has been explained earlier. These distorted signal segments could be reproduced by extrapolating from the surrounding, better signal. It has transpired, however, that these distortions tend to reappear in the extrapolated signal. In NMR, backward one-point extrapolation may be used to predict the sample I_0, if it cannot be measured properly.

Outside spectroscopy, extrapolation by using linear prediction has been successfully applied also in regenerating damaged segments to audio signals. In this application, the com-

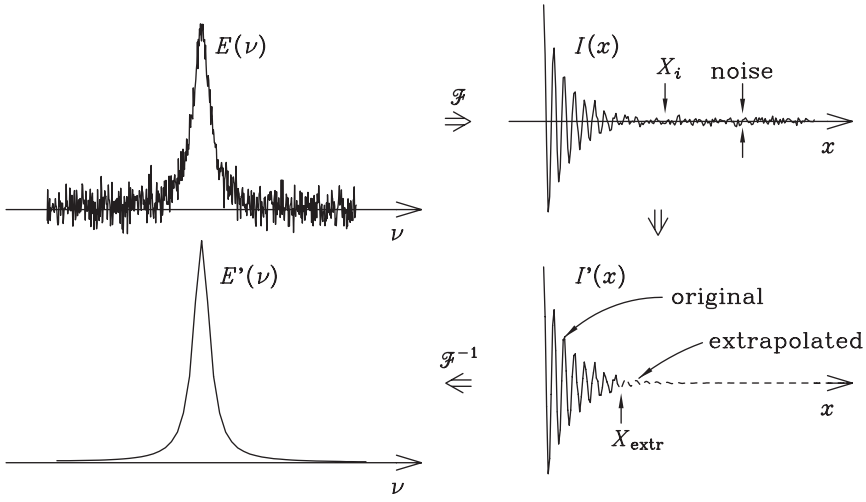


Figure 13.12: Noise filtering by signal extrapolation. In the signal domain, the noisy part is regenerated by extrapolating it from the first part, where the signal-to-noise ratio is better. More noise is removed this way than by simple truncation in the optimal point X_i.

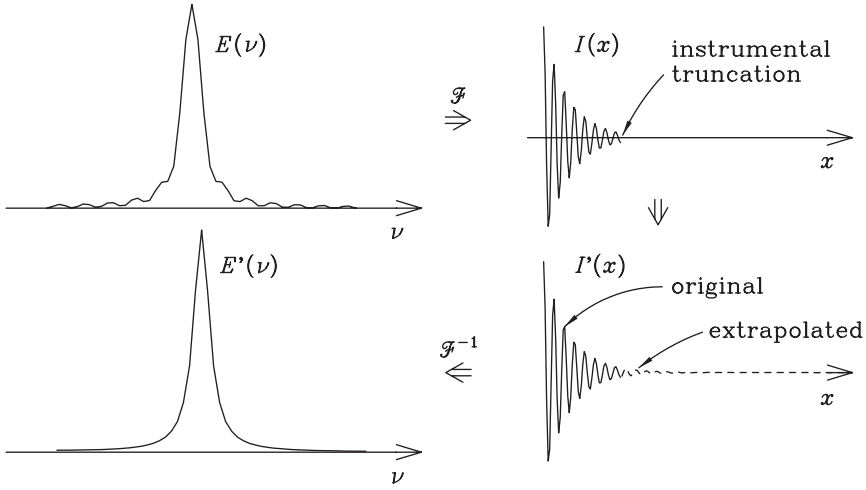


Figure 13.13: The same procedure that was used to noise filtering in Figure 13.12 can also be applied to the removal of oscillations due to device distortions.

bination of Levinson–Durbin recursion and Burg's formula has turned out to be by far the best method to construct the impulse response coefficients. Since the number of different frequencies in audio signals is very large, N and M must also be very large. For music signals, choosing M to be of the order of 1000 and fixing the ratio N/M near to 2 has

proven a good compromise, so that the N samples used to construct the h_l-coefficients are still stationary enough, and the M coefficients are able to extrapolate a sufficient amount of different frequencies. Naturally, the best result is obtained by extrapolating from both directions. The samples immediately preceding and immediately following the corrupted section are then used to compute the forward and backward extrapolating impulse response coefficients, respectively. The forward and backward extrapolated signals are combined so that up to the midpoint of the corrupted section the forward extrapolation dominates, and there the emphasis is rapidly but continuously switched to the backward extrapolation. As we may see from the experimental test shown in Figure 13.14, the extrapolation error is incredibly small, considering the large number of samples reconstructed. Moreover, the reconstructed section is very plausible to the human ear.

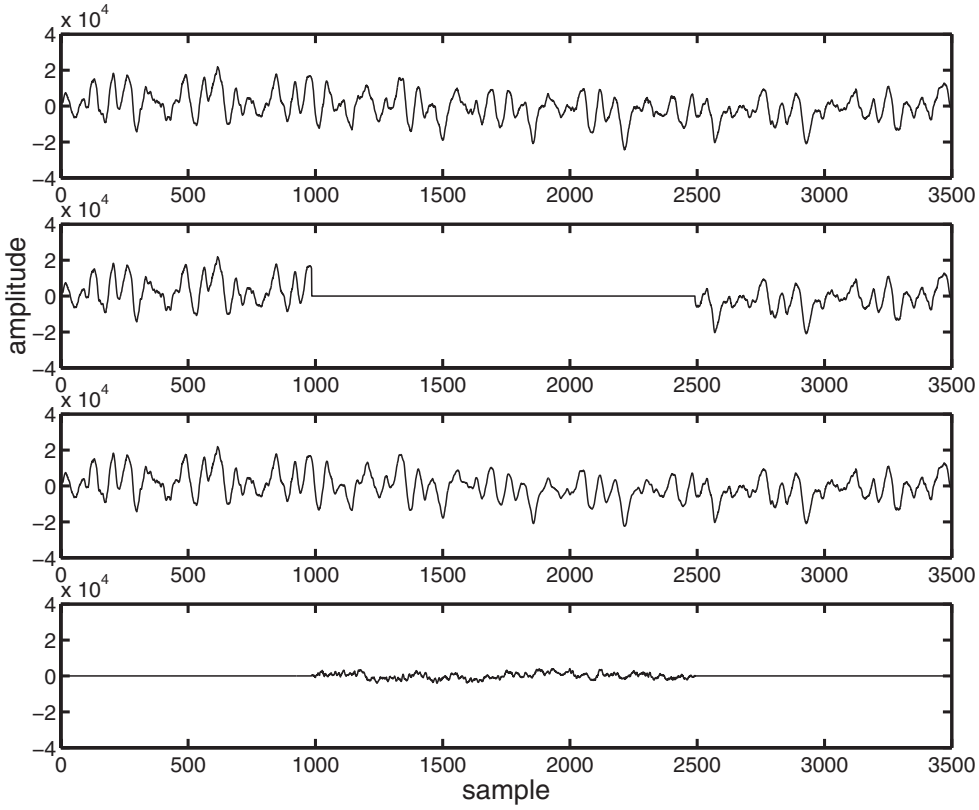


Figure 13.14: A piece of a signal from a jazz recording (top); the same signal after 1500 samples (0.034 seconds in time) have been lost; the missing samples have been extrapolated from both directions by Burg's impulse response with $M = 2000$ and $N = 5000$ (see the text); and the extrapolation residual, obtained by subtracting the extrapolated signal from the true signal (bottom). The numbers on the x-axis are the sample numbers.

13.14 Summary

As a short summary about signal extrapolation we state the following (all constants are real):

A signal that can be expressed in the form

$$I(x) = \sum_{k=1}^{M} A_k e^{i\phi_k} e^{-\alpha_k x} e^{i2\pi\nu_k x}$$

can be predicted, and thereby extrapolated, by M impulse response coefficients h_l. The same impulse response coefficients are applicable irrespective of the values of the amplitudes A_k and phases ϕ_k. If all the α_k are zero, then the same coefficients always apply to forward prediction (Equation 13.1) and backward prediction (Equation 13.2).

In the special case

$$I(x) = \sum_{k=1}^{n} A_k \cos(2\pi\nu_k x + \phi_k) \tag{13.31}$$

the same impulse response coefficients are always able to predict the signal both forward and backward. In addition, the impulse response coefficients are real. The minimal number of the impulse response coefficients needed is $2n$.

If the original signal is real and free of phase errors, so that the original line shape $W_R(\nu)$ is real and symmetric, then the interesting piece of spectrum (the partial spectrum) can be extrapolated symmetrically to negative ν-values before it is Fourier transformed. The ensuing partial signal, after division by $w_R(x)$, is then a real signal of the form given by Equation 13.31 with $\phi_k = 0$ for every k. The minimal number of impulse response coefficients needed to extrapolate it is twice the number of spectral lines with $\nu_k > 0$.

If the extrapolated cosine waves are apodized by a (real and) symmetric function $w(x)$, the line shape $W(\nu)$ after inverse Fourier transform is likewise (real and) symmetric. Any asymmetries are a token of incorrect impulse response coefficients. This characteristic is applied in the frequency tuning method to improve iteratively the h_l-coefficients.

When M (the number of h_l-coefficients) is smaller than the number of known signal samples, then the spectral lines, whose constituent waves were not extrapolated, are visible as humps in the line narrowed spectrum.

Example 13.1: Function ax^2 (a is a real constant) has been sampled by using sampling interval Δx. (There is not inevitably a sampling point in the origin.) Show that these samples can be extrapolated by three impulse response coefficients h_1, h_2, h_3, and find the values of these coefficients.

Solution. We require that the forward predicted estimate, given by Equation 13.1, is errorless at an arbitrary point x, or (coefficient a has been cancelled)

$$\begin{aligned} x^2 &= \Delta x h_1 (x - \Delta x)^2 + \Delta x h_2 (x - 2\Delta x)^2 + \Delta x h_3 (x - 3\Delta x)^2 \\ &= (\Delta x h_1 + \Delta x h_2 + \Delta x h_3)\, x^2 - 2\Delta x\, (\Delta x h_1 + 2\Delta x h_2 + 3\Delta x h_3)\, x \\ &\quad + (\Delta x)^2\, (\Delta x h_1 + 4\Delta x h_2 + 9\Delta x h_3)\,. \end{aligned}$$

This requirement is fulfilled if the coefficient of x^2 is 1, the coefficient of x is 0, and the constant term is 0, or

$$\begin{cases} \Delta x h_1 + \Delta x h_2 + \Delta x h_3 &= 1, \\ \Delta x h_1 + 2\Delta x h_2 + 3\Delta x h_3 &= 0, \\ \Delta x h_1 + 4\Delta x h_2 + 9\Delta x h_3 &= 0\,. \end{cases}$$

These conditions have the unique solution

$$\begin{cases} h_1 &= 3/\Delta x, \\ h_2 &= -3/\Delta x, \\ h_3 &= 1/\Delta x\,. \end{cases}$$

Problems

1. Certain M impulse response coefficients $h_1, \ldots, h_M$ are able to predict any sample I_j from a signal $I(x)$ as a linear combination of the samples $I_{j-1}, \ldots, I_{j-M}$ as

$$I_j = \Delta x \sum_{l=1}^{M} h_l I_{j-l}.$$

Find new coefficients $h'_3, \ldots, h'_{M+2}$ that are able to predict signal samples I_j from samples $I_{j-3}, \ldots, I_{j-M-2}$ as

$$I_j = \Delta x \sum_{l=3}^{M+2} h'_l I_{j-l}.$$

2. Let us define the sequence of samples I_j so that

$$I_j = a k_1^j + b k_2^j,$$

where a, b, k_1, and k_2 are real numbers.

(a) Show that the sequence can be extrapolated by two impulse response coefficients h_1 and h_2, and calculate $h_1 \Delta x$ and $h_2 \Delta x$ as functions of k_1 and k_2.

(b) Compute the numerical values of $h_1 \Delta x$ and $h_2 \Delta x$, when $k_1 = 0.8$ and $k_2 = 0.8^{-1} = 1.25$. Finally, test the coefficients by extrapolating the sequence

$$x_j = 3 \times (0.8)^j - 2 \times (0.8)^{-j}.$$

Are your coefficients also able to extrapolate this sequence backwards?

3. (a) Construct such impulse response coefficients that are able to extrapolate the Fibonacci sequence 1, 1, 2, 3, 5, 8, 13, ..., where each sample is the sum of two preceding samples. Since the sampling interval is not defined now, it is enough to find the products $\Delta x h_l$.

   (b) Are your coefficients also able to extrapolate this sequence backwards? If not, find the coefficients $\Delta x h_l$ capable of backward extrapolation.

4. Certain impulse response coefficients can extrapolate any sinusoidal wave with the wavenumber ν_0. Calculate the q-factor, when these coefficients are used to extrapolate

   (a) cosine wave with the wavenumber ν_0,

   (b) sine wave with the wavenumber ν_0.

5. Let $M = 4$ and let the impulse response coefficients be

$$\begin{cases} h_1 &= 884.9270\ \text{cm}^{-1}, \\ h_2 &= -195.7739\ \text{cm}^{-1}, \\ h_3 &= h_1, \\ h_4 &= -2000.0000\ \text{cm}^{-1}. \end{cases}$$

   The signal domain sampling interval is $\Delta x = 0.5 \times 10^{-3}$ cm.

   Prove that these coefficients are able to extrapolate a nondecaying wave, whose wavenumber is $\pm 200\ \text{cm}^{-1}$ or $\pm 700\ \text{cm}^{-1}$.

   Hint: Apply the predictability condition of Equation 13.9.

6. The h_l-coefficients given in the preceding problem are used to extrapolate backwards a sequence of five known samples $I_0, \ldots, I_4$. The sampling interval is 0.5×10^{-3} cm, as in the previous problem. Calculate the q-factor with $J = 1$ and $N_0 = 5$, when

   (a) the sequence is the constant amplitude cosine wave $I_j = \cos(2\pi \nu j \Delta x)$, where $\nu = 700\ \text{cm}^{-1}$,

   (b) the sequence is the decaying cosine wave $I_j = \exp\left[-\alpha(j\Delta x)^2\right] \cos(2\pi \nu j \Delta x)$, where $\alpha = 100.0 \times 10^3\ \text{cm}^{-2}$ and $\nu = 700\ \text{cm}^{-1}$.

7. Let us calculate the theoretical impulse response coefficients by sampling the function

$$h(x) = \sum_k \left[C_k^+ \exp(i 2\pi \nu_k x) + C_k^- \exp(-i 2\pi \nu_k x) \right],$$

   where the summation is over all the wavenumbers ν_k present in the signal ($\nu_k > 0$). The actual coefficients are the samples

$$h_l = h(l \Delta x).$$

   Derive a linear set of equations for solving the numbers C_k^+ and C_k^-.

   Hint: Start by writing the predictability condition of Equation 13.9 for an arbitrary wavenumber ν_k present in the signal.

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Answers to problems

Chapter 1.

2. $h(-t)$ (both).
3. $e^{-\pi^2 t^2/\alpha}$ (both).
4. $e^{-2\pi\sigma|t|}$.
5. $I = \cos(2\pi f_0 T)\ T^2\ \mathrm{sinc}^2(\pi f_0 T)$.
6. $\dfrac{1}{\pi f}\dfrac{\sin[(1+1/N)\alpha/2]\sin(\alpha/2)}{N\sin[\alpha/(2N)]}$, where $\alpha = 2\pi f T$. $N \to \infty : T\,\mathrm{sinc}^2(\pi T f)$.
10. $\dfrac{2}{\pi}\displaystyle\int_0^\infty \frac{\sin(2\pi f\varepsilon)}{f}\,\mathrm{d}f = 1$ (from mathematical tables).
11. (a) 1. (b) $e^{\pm i 2\pi f_0 t}$. (c) $2\cos(2\pi f_0 t)$. (d) $\pm 2i\sin(2\pi f_0 t)$.

Chapter 2.

4. $\begin{cases} 2T - |t|, & |t| \le 2T, \\ 0, & |t| > 2T. \end{cases}$
7. $\frac{1}{2i}H(f - f_0) - \frac{1}{2i}H(f + f_0)$.
9. (a) $2T\,\mathrm{sinc}(\pi 2Tf) - T\,\mathrm{sinc}^2(\pi Tf)$. (b) $2\cos(2\pi Tf)\ T\ \mathrm{sinc}^2(\pi Tf)$.
   (c) $T\ [1 + 2\cos(2\pi Tf)]\ \mathrm{sinc}^2(\pi Tf) - 2T\,\mathrm{sinc}(\pi 2Tf)$.
10. $T\,\mathrm{sinc}^2(\pi Tf)$.
11. $-if e^{-\pi f^2}$.
12. $iA\pi f_1[\delta(f - f_1) - \delta(f + f_1)] + iA\pi f_2[\delta(f - f_2) - \delta(f + f_2)]$.
13. (a) $\frac{2}{3}\pi$. (b) $e^{-\pi a^2}$.
14. $\frac{3}{4}\pi$.
15. $\frac{\pi}{4a}$.

Chapter 3.

4. 5.14 ns, 10.29 ns, 15.43 ns.
5. $\mathcal{F}^{-1} : \Delta t\boldsymbol{W}^-$, where $W^-_{jk} = w_M^{-jk}$; $j, k = 0, \ldots, M-1$.
   $\mathcal{F}^{+1} : \Delta f\boldsymbol{W}^+$, where $W^+_{jk} = w_M^{jk}$; $j, k = 0, \ldots, M-1$.

Chapter 4.

2. $H_k = \begin{cases} \frac{1}{2}(C_k + C^*_{N-k}) + \frac{w_{2N}^{-k}}{2i}(C_k - C^*_{N-k}), & k = 1, \ldots, N-1, \\ \frac{1}{2}(C_{k-N} + C^*_{2N-k}) + \frac{w_{2N}^{-k}}{2i}(C_{k-N} - C^*_{2N-k}), & k = N+1, \ldots, 2N-1, \\ \frac{1}{2}(C_0 + C_0^*) + \frac{1}{2i}(C_0 - C_0^*), & k = 0, \\ \frac{1}{2}(C_0 + C_0^*) - \frac{1}{2i}(C_0 - C_0^*), & k = N. \end{cases}$

3. (a) $H_k = 1$. (b) $H_k = e^{-i\pi k}$. (c) $H_k = \cos(\pi pk/N)$. (d) $H_k = -i\sin(\pi pk/N)$.

Chapter 5.

3. (a) e^{-st_0}, (b) $\frac{1}{s}$, (c) $b\delta(t) - abe^{-at}$.
4. $\frac{1}{a-s} + \frac{1}{a+s}$, if $|\mathrm{Re}\{s\}| < a$.
5. (a) $\frac{s}{s^2+a^2}$. (b) $\frac{a}{s^2+a^2}$.
6. $h(t) = t + \sin t$.
7. $f(t) = \sin(2t)$, $g(t) = \sin(2t) + \cos(2t)$.
8. $y(t) = \dfrac{Kp}{p^2 - \omega_0^2}\left[\dfrac{1}{\omega_0}\,\sin(\omega_0 t) - \dfrac{1}{p}\,\sin(pt)\right]$.

   In resonance: $y(t) = \dfrac{K}{2\omega_0^2}\left[\sin(\omega_0 t) - \omega_0 t\cos(\omega_0 t)\right]$.

10. $g(t) = \begin{cases} \frac{1}{2}e^{2t}, & t < 0, \\ -\frac{1}{2}e^{-2t}, & t > 0. \end{cases}$
11. $h_{\mathrm{out}}(t) = \begin{cases} \frac{\beta_1}{\beta_1 - \beta_0}\,A\,(e^{-\beta_0 t} - e^{-\beta_1 t}), & t \geq 0, \\ 0, & t < 0. \end{cases}$
12. $g(t) = \delta(t) - \frac{R}{L}\,e^{-\frac{R}{L}t}$ at $t \geq 0$, and $g(t) = 0$ otherwise.
13. $g(t) = te^{-Ct}$ at $t \geq 0$, and $g(t) = 0$ otherwise.

Chapter 6.

1. $x = \frac{1}{2\Delta\nu}$.
3. $E'(\nu) = \frac{1}{2}I_0 \displaystyle\sum_{n=-\infty}^{\infty} J_n(2\pi\nu_0\varepsilon)\left[\delta(\nu - \nu_0 + \frac{n}{K}) + \delta(\nu + \nu_0 - \frac{n}{K})\right]$.
4. $2N = 32768$.
7. $x = \frac{\ln 10}{2\pi\sigma}$.
8. $E(\nu) \approx \dfrac{2\pi}{\nu_0}\left[1 - \dfrac{\sqrt{1 - (\nu/\nu_0)^2}}{\sin(\alpha_{\max})}\right]$.
9. $L = \dfrac{1.21\,f^2}{\nu_0 r^2}$.
10. (a) $2\,\ln 2/\pi$, (b) $4\,\ln 2/\pi$, (c) ≈ 1.2067.
11. (a) $0.8858929413789046806\ldots/T$, (b) $T \approx 8.8589\ \mu$s.
12. (a) $T = 885.9$ ns, (b) $\Delta t = 2.381$ ns.

Chapter 9.

1. $\theta = \arcsin(\frac{\lambda}{L+p}) \approx 0.005273$ rad.
4. $I_{\text{out}}(\theta) = a^2 \operatorname{sinc}^2(\pi a\nu \sin\theta)\, \dfrac{\sin^2(N\pi 18a\nu \sin\theta)}{\sin^2(\pi 18a\nu \sin\theta)}\, \dfrac{\sin^2(10\pi a\nu \sin\theta)}{\sin^2(\pi 2a\nu \sin\theta)}$.
6. (a) $8/9 \approx 89\,\%$.
7. (a) $I_{\text{out}}(s, p) = D^2 \operatorname{sinc}^2(\pi Ds)\, \dfrac{\sin^2(\pi Nds)}{\sin^2(\pi ds)}\, h^2 \operatorname{sinc}^2(\pi hp)$.
8. $I_{\text{out}}(s, p) = \delta^2(s, p) + X^2 \operatorname{sinc}^2(\pi Xs) Y^2 \operatorname{sinc}^2(\pi Yp)$.
9. $E_{\text{out}}(s, p) = 4R^2 \operatorname{sinc}(2\pi Rs) \operatorname{sinc}(2\pi Rp) - \pi R^2 \dfrac{J_1(2\pi R\sqrt{s^2+p^2})}{\pi R\sqrt{s^2+p^2}}$.
11. $I_{\text{out}}(s, p) = \frac{L^2}{4\pi^2 s^2} \{\operatorname{sinc}^2[\pi L(p+s)] + \operatorname{sinc}^2[\pi L(p-s)]\}$
$-\frac{L^2}{2\pi^2 s^2} \cos(2\pi Ls) \operatorname{sinc}[\pi L(p+s)] \operatorname{sinc}[\pi L(p-s)]$.
12. $D = 2f\lambda/d \approx 2.5$ mm.
13. $I_{\text{out}} \approx \delta^2(s, p) + \frac{1}{4C^2} \sin^2\left(\frac{\pi^2\nu^2}{Cf^2} r'^2\right)$, where $r' = \sqrt{x'^2 + y'^2}$.

Chapter 10.

1. $\Delta x = \pi/C$.
3. $H^{(k)}(0) = \begin{cases} 0, & k \text{ is odd}, \\ i^k \frac{C^k}{k+1}, & k \text{ is even}. \end{cases}$
4. $\langle x\rangle = \mu$, standard deviation$= \sigma$ (this is the normal distribution).
7. (a) $\sigma = \sqrt{2}$. (b) $\sigma = 1/(\sqrt{2}\pi a)$.
8. $2pq/(p^2+q^2)^2$.
9. $7/8$.

Chapter 11.

2. (a) $\Delta t \approx 179\ \mu$s. (b) $\Delta t = 625\ \mu$s.
3. $G_k = \dfrac{\sin[\pi(2n+1)k/M]}{(2n+1)\sin(\pi k/M)}$.
5. $G_k = 1 - 2\sin(\pi k/M)\sin(3\pi k/M)$.
7. $1/3$.
8. $\dfrac{D - 2/T_0}{D} = 1 - \dfrac{8}{p(m+1)}$.
9. (a) $\Delta t = 31.25\ \mu$s. (b) $g(t) = \delta(t) - \cos[2\pi\ (2.5\text{ kHz})\, t]\ \operatorname{sinc}[\pi\ (3\text{ kHz})\, t]\ (6\text{ kHz})$, where $\delta(t) = 1/\Delta t$ in the origin, zero otherwise. (c) 641 data.
11. $\Delta(0) \approx -0.0226$.
13. $\Delta(0) \approx -0.1790$.
14. $Q_0 \approx 3.091$.
16. (a) $N' = \sqrt{\frac{T}{3T_0}}\, N$.
17. $(S/N)' \approx 1580$.
19. $2N = 40\,000$.
20. Three data.
21. $x = \pm 2nd$.

Chapter 12.

1. $\frac{I}{I_{\mathrm{R}}} = \begin{cases} \dfrac{1}{\pi\sigma}\, e^{2\pi\sigma|t|}(1-\dfrac{|t|}{T}), & |t| \le T, \\ 0, & |t| > T. \end{cases}$
2. (a) $e^{2\pi\sigma|t|}\ \mathrm{sinc}^2(\frac{2\pi\sigma}{K}\,t)$. (b) $\frac{2}{K\pi}\, e^{2\pi\sigma|t|}\ \mathrm{sinc}^2(\frac{2\pi\sigma}{K}\,t)$.
3. $Q_0 \approx K\sqrt{1.2067\pi K_0(K-1)/K}\,\left[e^{1.2067\pi K_0(K-1)/K}-1\right]^{-1/2}$.
5. $\dfrac{B\pi}{2p}\,\dfrac{1}{\cos(\frac{q\pi+i2\pi^2 t}{2p})}$, $B = \dfrac{2p}{\pi}\,\cos(\dfrac{q\pi}{2p})$.
6. (a) $A_1\delta(\nu-\nu_1)+A_2\delta(\nu-\nu_2)+A_3\delta(\nu-\nu_3)$. (b) $A_1 e^{i2\pi\nu_1 x}+A_2 e^{i2\pi\nu_2 x}+A_3 e^{i2\pi\nu_3 x}$.
7. $\delta\left[x-(s-L/2)\right]-\delta\left[x-(s+L/2)\right]$.
8. $\frac{1}{4}$.
9. (a) ≈ 0.44626. (b) ≈ 1.881.
10. $\frac{2L}{\nu}\,\cos(2\pi\nu L)-\frac{1}{\pi\nu^2}\,\sin(2\pi\nu L)$.
11. Second: $\frac{1}{L}\,\delta\left[x-(s-L)\right]-\frac{2}{L}\,\delta(x-s)+\frac{1}{L}\,\delta\left[x-(s+L)\right]$.

Chapter 13.

1. $\begin{cases} h_l' &= \left(\Delta x h_1^2+h_2\right)\Delta x h_{l-2}+\Delta x h_1 h_{l-1}+h_l;\ l=3,\ldots,M, \\ h_{M+1}' &= \left(\Delta x h_1^2+h_2\right)\Delta x h_{M-1}+\Delta x h_1 h_M, \\ h_{M+2}' &= \left(\Delta x h_1^2+h_2\right)\Delta x h_M\ . \end{cases}$
2. (a) $\begin{cases} h_1\Delta x &= k_1+k_2, \\ h_2\Delta x &= -k_1k_2\,. \end{cases}$ (b) $\begin{cases} h_1\Delta x &= 2.05, \\ h_2\Delta x &= -1\,. \end{cases}$
   These coefficients can extrapolate to both directions.
3. (a) $M=2,\ \Delta x h_1 = \Delta x h_2 = 1$.
   (b) Above coefficients can extrapolate only forward.
   Backward extrapolating coefficients are $\begin{cases} \Delta x h_1 &= -1, \\ \Delta x h_2 &= 1\,. \end{cases}$
4. (a) 1. (b) -1 (q-factor assumes symmetric signal).
6. (a) 1. (b) 0.7115266.
7. $\begin{cases} \Delta x\sum\limits_k\left[C_k^+ S\left(\beta_{k1}^+\right)+C_k^- S\left(\beta_{k1}^-\right)\right] &= 1, \\ \Delta x\sum\limits_k\left[C_k^+ S\left(-\beta_{k1}^-\right)+C_k^- S\left(-\beta_{k1}^+\right)\right] &= 1, \\ \Delta x\sum\limits_k\left[C_k^+ S\left(\beta_{k2}^+\right)+C_k^- S\left(\beta_{k2}^-\right)\right] &= 1, \\ \Delta x\sum\limits_k\left[C_k^+ S\left(-\beta_{k2}^-\right)+C_k^- S\left(-\beta_{k2}^+\right)\right] &= 1, \\ \qquad\qquad\vdots & \vdots\ , \end{cases}$
   where $S\left(\beta_{kn}^{\pm}\right) = \sum\limits_{l=1}^{M}\exp(i2\beta_{kn}^{\pm}l) = \begin{cases} M, & \beta_{kn}^{\pm}=0, \\ \frac{\exp(i2M\beta_{kn}^{\pm})-1}{1-\exp(-i2\beta_{kn}^{\pm})}, & \beta_{kn}^{\pm}\neq 0\,, \end{cases}$
   and $\begin{cases} \beta_{kn}^+ &= \pi\,(\nu_k-\nu_n)\,\Delta x, \\ \beta_{kn}^- &= \pi\,(-\nu_k-\nu_n)\,\Delta x\,. \end{cases}$

Fourier Transforms in Spectroscopy. J. Kauppinen, J. Partanen

ISBNs: 3-527-40289-6 (Hardcover); 3-527-60029-9 (Electronic)

Bibliography

References

[1] P. Saarinen and J. Kauppinen, "Multicomponent Analysis of FT-IR Spectra," Appl. Spectrosc. **45**, (6), 953–963 (1991).

[2] J. Kauppinen, "Working Resolution of 0.010 cm^{-1} between 20 cm^{-1} and 1200 cm^{-1} by a Fourier Spectrometer," Appl. Opt. **18**, (11), 1788–1796 (1979).

[3] J. Kauppinen and V.-M. Horneman, "Large Aperture Cube Corner Interferometer with a Resolution of 0.001 cm^{-1}," Appl. Opt. **30**, (18), 2575–2578 (1991).

[4] J. Kauppinen, K. Jolma, and V.-M. Horneman, "New wave-number calibration tables for H_2O, CO_2, and OCS lines between 500 and 900 cm^{-1}," Appl. Optics. **21**, (18), 3332–3336 (1982).

[5] J. K. Kauppinen, D. J. Moffatt, H. H. Mantsch, and D. J. Cameron, "Smoothing of spectral data in the Fourier domain," Appl. Optics. **21**, (10), 1866–1872 (1982).

[6] J. K. Kauppinen, D. J. Moffatt, D. G. Cameron, and H. H. Mantsch, "Noise in Fourier-Self-Deconvolution," Appl. Opt. **20**, (10), 1866–1879 (1981).

[7] A. Losev, "On a Model Line shape for Asymmetric Spectral Peaks," Appl. Spectrosc. **48**, (10), 1289–1290 (1994).

[8] P. E. Saarinen, J. K. Kauppinen, and J. O. Partanen, "New Method for Spectral Line Shape Fitting and Critique on the Voigt Line Shape Model," Appl. Spectrosc. **49**, (10), 1438–1453 (1995).

[9] P. E. Saarinen, "Spectral Line Narrowing by Use of the Theoretical Impulse Response," Appl. Spectrosc. **51**, (2), 188–200 (1997).

Further reading

R. M. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, Inc., USA, 1965).

E. O. Brigham, *The Fast Fourier Transform* (Prentice-Hall, New Jersey, 1974).

R. D. Stuart, *Fourier Analysis* (Methuen, London, 1961).

P. F. Panter, *Modulation, Noise, and Spectral Analysis: Applied to information transmission* (McGraw-Hill, Inc., New York, 1965).

R. J. Bell, *Introductory Fourier Transform Spectroscopy* (Academic Press, Inc., New York, 1972).

J. W. Goodman, *Introduction to Fourier Optics,* second ed. (The McGraw-Hill Companies, Inc., New York, 1996).

M. H. Hayes, *Digital Signal Processing* in Schaum's outline series (The McGraw-Hill Companies, Inc., New York, 1999).

J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (John Wiley and Sons, New York, 1978).

G. Guelachvili and K. N. Rao, *Handbook of Infrared Standards with Spectral Maps and Transition Assigments between 3 and 2 600 μm* (Academic Press, Inc., Orlando, Florida, 1986).

A. G. Marshall and F. D. Verdun, *Fourier Transforms in NMR, Optical, and Mass Spectrometry* (Elsevier Science Publishers B.V., Amsterdam, 1990).

E. Hecht, *Optics* (Addison Wesley Longman, Inc., Reading, Massachusetts, 1998).

P. A. M. Dirac, *The Principles of Quantum Mechanics,* fourth ed. (Oxford University Press, 1959), Chapter 15. The δ function, pp. 58–61.

J. K. Kauppinen, D. J. Moffatt, H. H. Mantsch, and D. G. Cameron, "Fourier Self-Deconvolution: A Method for Resolving Intrinsically Overlapped Bands," Appl. Spectrosc. **35**, (3), 271–276 (1981).

J. K. Kauppinen, "Fourier Self-Deconvolution in Spectroscopy," in *Spectrometric Techniques,* G. A. Vanasse, ed., Vol. III (Academic Press, Inc., New York, 1983), Chapter 4, pp. 199–232.

J. K. Kauppinen, P. E. Saarinen, and M. R. Hollberg, "Linear prediction in spectroscopy," J. Mol. Struct. **324**, 61–74 (1994).

S. Haykin and S. Kesler, "Prediction-Error Filtering and Maximum-Entropy Spectral Estimation," in *Nonlinear Methods of Spectral Analysis,* second ed., S. Haykin, ed., in Topics in Applied Physics (Springer-Verlag, Berlin, 1983), Chapter 2, pp. 9–72.

J. K. Kauppinen, D. J. Moffatt, M. R. Hollberg, and H. H. Mantsch, "A New Line-Narrowing Procedure Based on Fourier Self-Deconvolution, Maximum Entropy, and Linear Prediction," Appl. Spectrosc. **45**, (3), 411–416 (1991).

J. K. Kauppinen, D. J. Moffatt, M. R. Hollberg, and H. H. Mantsch, "Characteristic of the LOMEP Line-Narrowing Method," Appl. Spectrosc. **45**, (9), 1516–1521 (1991).

J. K. Kauppinen, D. J. Moffatt, and H. H. Mantsch, "Nonlinearity of the maximum entropy method in resolution enhancement," Can. J. Chem. **70**, (12), 2887–2894 (1992).

J. K. Kauppinen and E. K. Saario, "What is Wrong with MEM?," Appl. Spectrosc. **47**, (8), 1123–1127 (1993).

J. K. Kauppinen and P. E. Saarinen, "True linear prediction by use of a theoretical impulse response," J. Opt. Soc. Am. B **11**, (9), 1631–1638 (1994).

P. E. Saarinen, "The Spectral Line Narrowing Problem and a Computer Program Based on the Gulf Tuning Method," Appl. Spectrosc. **52**, (12), 1569–1582 (1998).