Self-Study Manual on Optical Radiation Measurements: Part I—Concepts, Chapters 7, 8, and 9

Fred E. Nicodemus, Editor

Radiometric Physics Division
Center for Radiation Research
National Measurement Laboratory
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, Secretary

Jordan J. Baruch, Assistant Secretary for Science and Technology

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued June 1979

PREFACE

Preparation of the NBS Self-Study Manual on Optical Radiation Measurements is a major part of our effort to meet the needs of the National Measurement System with respect to the measurement of optical electromagnetic radiation. Needs exist for measurements with uncertainties of about one per cent, but the uncertainties actually achieved are often 5 to 10 per cent, or even more. These relatively poor accuracies, as compared to those in many other types of physical measurements, result to a great extent from the multidimensional character of optical radiation; radiant power is distributed and may vary with position, direction, wavelength, time, and polarization. In addition, many of those who make measurements of optical radiation have little or no training in optics and are limited in the amount of time they can spend to make up this deficiency. Moreover, there are few schools that offer courses in radiometry, and there are almost no adequate texts or references dealing with this entire subject.

The idea of producing a self-study manual at NBS to try to fill some of this void was developed by one of us (HJK) in the latter part of 1973. Detailed planning got under way in the summer of 1974 when a full-time editor (FEN) was appointed. NBS Tech. Note 910-1 with the first three chapters appeared in March 1976; TN 910-3 with chapter 6 came out in June 1977; and TN 910-2 with chapters 4 and 5 in February 1978. This Tech Note 910-4 includes chapters 7, 8, and 9. Three more chapters, 10, 11, and 12, are planned for publication by September 30, 1979. New budgetary limitations mean that no funds for continuing the Manual will be available after that date although, unfortunately, substantially more is needed to cover just the basic fundamentals of radiometry in Part I--Concepts. (The detailed outline, with chapter headings and topics to be covered in future chapters under a plan that had been developed when we were expecting to continue publishing about three chapters a year at least until September 1981, are given later in this preface.) Various possibilities for continuing to work on the Manual with outside support are being explored.

Our aim has been to provide a comprehensive tutorial treatment that is complete enough for self instruction. That is the meaning of the phrase "self-study"; the Manual does not contain explicitly programmed learning steps as that phrase often connotes. In addition, through detailed, yet concise, chapter summaries, the Manual is designed to serve also as a convenient reference source. Those already familiar with a topic should turn immediately to the summary at the end of the appropriate chapter. They can determine from that summary what, if any, of the body of the chapter they want to read for more details.

The material in the Manual is presented at the level of a college graduate in science or engineering but, even for those with facility in college mathematics and a first course in physics, it's not at all easy reading in spite of our best efforts at clarity and simplicity. This is an unavoidable result of the primary aim ("to make one-per-cent measurements commonplace") coupled with the fact that it must serve the needs of so many different fields, including astronomy and astrophysics, mechanical heat-transfer engineering, illumination engineering, photometry, meteorology, photo-biology and photo-chemistry, optical pyrometry, remote sensing, military infrared applications, etc.

Apparently, it is very difficult for those who have not been directly involved to realize the full implications of that situation. Each of us tends to think of radiometry and radiometric measurements in terms of his own immediate experience and requirements. What he wants is a set of simple carefully designed procedures, with appropriate cautions concerning likely sources of error, for making his particular measurements. However, the next reader wants the same thing, but for *his*, entirely different measurements. The desired radiometric quantities to be measured are different, the instrumentation is different, the ambient conditions are different -- the possible ways in which significant differences may exist, in terms of the radiation parameters (position, direction, spectrum, time or frequency of modulation or fluctuation, and polarization) as well as instrumental and environmental parameters, are so numerous that any attempt to cover them all with a "cookbook" treatment of specific measurement procedures would be impossibly unwieldy and could never be completed within any conceivable budget limitations short of utopia. The only way in which we can hope to effectively assist *every* reader who needs to make one-per-cent measurements is to provide him with material which, *with sufficient effort on his*

part, will give him enough of an understanding and grasp of basic general principles and techniques to solve his own particular problems. That's why we have concentrated on the basic material of Part I--Concepts and why it is not easy reading. The problems of radiometry in general just aren't simple ones, even when limited to classical radiometry in the domain of geometrical optics (ray optics). Although we have kept within that limitation so far, we recognize the increasingly urgent need for similar treatment of the problems of coherent radiation with strong interference and diffraction effects that do not follow the laws of ray optics. The basic theoretical work underlying the radiometry of coherent radiation is just now being developed at a much more sophisticated level and we had planned, as part of the proposed three-year program, to study that material and, if possible, use it to work out and present a chapter on the effects of coherence in radiometric measurements to go along with the rest of this Manual. But it is not at all clear that sufficient progress has been made in the basic work to make this feasible yet. If production of the Manual continues for a few more years, we will try to prepare such a chapter.

The basic approach and focal point of the treatment in this Manual is the measurement equation, first introduced in detail in Chapter 5. We believe that every measurement problem should be addressed with an equation relating the quantity desired to the data obtained through a detailed characterization of the instruments used and the radiation field observed, in terms of all of the relevant parameters. Those parameters always include the radiation parameters (listed above), as well as environmental and instrumental parameters peculiar to each measurement configuration. The objective of the Manual is to develop the basic concepts required so that the reader will be able to use this measurement-equation approach. It is our belief that this is the only way that uncertainties in the measurement of optical radiation can generally be limited to one, or at most a few, per cent.

The overall plan for the Manual organized it into three Parts:

Part I--Concepts:

Step-by-step buildup of the *measurement equation* in terms of the radiation parameters, the properties and characteristics of sources, optical paths, and receivers, and the environmental and instrumental parameters. Useful quantities are defined and discussed and their relevance to various applications in many different fields (photometry, heat-transfer engineering, astronomy, photo-biology, etc.) is indicated. However, discussions of actual devices and measurement situations in this Part are mainly for purposes of illustrating concepts and basic principles.

Part II--Instrumentation:

Descriptions, properties, and other pertinent data concerning typical instruments, devices, and components involved in common measurement situations. Included is material dealing with sources, detectors, filters, atmospheric paths, choppers (and other types of optical modulators), prisms, gratings, polarizers, radiometers, photometers, spectroradiometers, spectrophotometers, etc.

Part III--Applications:

Measurement techniques for achieving a desired level of, or improving, the accuracy of a measurement. Included will be a very wide variety of examples of environmental and instrumental parameters with discussion of their effects and how to deal with them. This is where we deal with the real measurements in the real world. The examples will also be drawn from the widest possible applications in illumination engineering, radiative heat transfer, military infrared devices, remote sensing, meteorology, astronomy, photo-chemistry and photo-biology, etc.

-- That was our rather ambitious plan when we started out; limitations of support and available resources, particularly available authors, have determined how much, or how little, we could accomplish. So far, as indicated earlier, we have concentrated on Part I--Concepts. The chapters already published are:

1. Introduction,
2. Distribution of Optical Radiation with Respect to Position and Direction--Radiance,

3. Spectral Distribution of Optical Radiation,

4. More on the Distribution of Optical Radiation with Respect to Position and Direction,

5. An Introduction to the Measurement Equation,

6. Distribution of Optical Radiation with Respect to Polarization.

Chapters 1, 2, and 3 were in TN 910-1; 4 and 5 in TN 910-2; and 6 in TN 910-3. In the present Tech. Note 910-4, are three more chapters:

7. The Relative Spectral Responsivity and Slit-Scattering Function of a Spectroradiometer,

8. Deconvolution,

9. Physically Defining Measurement-Beam Geometry by Using Opaque Barriers.

Currently scheduled for completion in the fiscal year ending September 30, 1979, are three more chapters on the following topics:

Linearity Calibrations,

Physical Photometry,

Blackbody Radiation and Temperature Scales.

This is a departure from the topics listed for FY 79 in our earlier plan for activities up to September 1981. Even though that plan has been abandoned, we present it here to show what coverage we had considered as providing a reasonably complete and useful treatment of basics in Part I--Concepts and as a beginning of Part II--Instrumentation and Part III--Applications: in FY-79:

Part III chapter on Linearity Calibrations,

Part I chapter on Physical Photometry,

Part III chapter on Spectroradiometry of Continuous UV Spectra (including an appendix on Uncertainty),

study and decision on feasibility of a basic Part I chapter on Coherence in Radiometry vs. a more limited treatment of Diffraction Corrections to Throughput Calculations,

preparatory study and completion of detailed outline for a Part I chapter on Distribution of Optical Radiation with Respect to Time;

in FY-80:

Part I chapter on Distribution of Optical Radiation with Respect to Time,

Part I chapter on Spectrophotometry,

Part I chapter on Coherence in Radiometry or on Diffraction Corrections to Throughput Calculations,

Part III chapter on Spectroradiometry of Spectral Lines;

in FY-81:

Part I or II chapter on Detectors of Optical Radiation,

Part I chapter on Blackbody Radiation and Temperature Scales,

Part II chapter on Radiometric and Photometric Source Standards.

The plan then called for a reevaluation to decide whether to continue with more chapters, particularly those for Parts II and III.

Incidentally, in preparing material for the Manual, we have rediscovered the fact that the best way to learn something is to try to teach it to someone else. The exercise of preparing tutorial material for such wide general application has required us to analyze our own measurement activities in a different way that has broadened our understanding, resulting in improved methods and more accurate results. Also, note that all references to measurement accuracy or uncertainty in this preface are concerned not only with precision (relating to the repeatability of measurement results) but also with accuracy (relating to agreement with the "truth" which, while unknowable in the last analysis, is approximated by analysis and estimates based on the widest possible experience including agreement with measurements of the same quantities by others, particularly when using different instrumentation and measurement methods).

We are indebted to a great many individuals for invaluable "feedback" that has helped us to put these chapters together more effectively. We renew our invitation to all readers to submit comments, criticisms, and suggestions. In particular, we would welcome illustrative examples and problems from as widely different areas of application as possible.

We particularly acknowledge the inputs from members of CORM (the Council for Optical Radiation Measurements), especially the CORM Coordinators. We are also grateful to Donald A. McSparron, Joseph C. Richmond, John B. Shumaker, Albert T. Hattenburg, and Diana Nyyssonen, all of NBS, for helpful discussions and editorial assistance.

We are especially grateful to Mrs. Betty Castle for the skillful and conscientious effort that produced the excellent typing and layout of this difficult text. We also want to thank Paul M. Beachley for his capable help with the figures.

Fred E. Nicodemus, Editor

Henry J. Kostkowski, Program Director

April 1979

Contents

List of Figures

List of Tables

SELF-STUDY MANUAL on OPTICAL RADIATION MEASUREMENTS

Part I. Concepts

This is the fourth in a series of Technical Notes (910-) entitled "Self-Study Manual on Optical Radiation Measurements". It contains Chapters 7, 8, and 9 of this Manual. Additional chapters will continue to be published, similarly, as they are completed. The Manual is a comprehensive tutorial treatment of the measurement of incoherent radiation that is complete enough for self instruction. Detailed chapter summaries make it also a convenient authoritative reference source.

The manner in which the spectral responsivity of a spectroradiometer containing a monochromator varies with wavelength is treated in Chapter 7. The nature and characteristics of this function, how it is determined, and the effects it has on spectroradiometric measurements are discussed in detail. The traditional, indirect determination, involving the slit-scattering function, is shown to be highly inaccurate except for wavelengths in the central portion of the function. An introduction to the measurement of spectral line radiation is also presented.

Deconvolution, discussed in Chapter 8, is the numerical process of recovering an improved spectral distribution from spectroradiometric measurements inevitably smeared spectrally by the spectral-responsivity function of the radiometer. A simple, iterative technique, which is extensively used, will be completely satisfactory in almost all radiometric situations. A more sophisticated technique, which is founded on a sounder theoretical basis, should succeed in many of the few remaining instances where the simpler technique fails. Examples of computer programs for the two techniques are given in an appendix. Both are applied to a simple illustrative numerical example with brief mention of some of the sources of difficulty and the limitations of deconvolution. Finally, we touch on the question of errors and describe a couple of useful measures of accuracy.

The physical definition of measurement beams, including the integration limits in the measurement equation, is treated in Chapter 9 in terms of geometrical-optics quantities and concepts. The usual definitions of geometrical optics are extended to situations where defining the measurement beam may not involve the imaging of any source. The importance of aperture pairs or equivalent aperture/aperture-image pairs, in defining measurement beams without vignetting, is developed. Also discussed are entrance pupils as receiving apertures in radiometry (their positions and dimensions), diffusers, vignetting, and the effects of beam geometry on overall instrument radiance and irradiance responsivities. The treatment is in terms of simple, ideal, geometrical optics, with sharp image and shadow boundaries, and only passing mention is made of perturbations caused by imperfect imaging, aberrations, scattering, and diffraction.

Key Words: Aperture/aperture-image pair; aperture pair; calibration; convolution; deconvolution; geometrical-optics radiometry; inversion; measurement-beam geometry; relative spectral responsivity; slit-scattering function; slit-scattering function corrections; spectral line radiometry; spectroradiometer characterizations; spectroradiometry; vignetting.

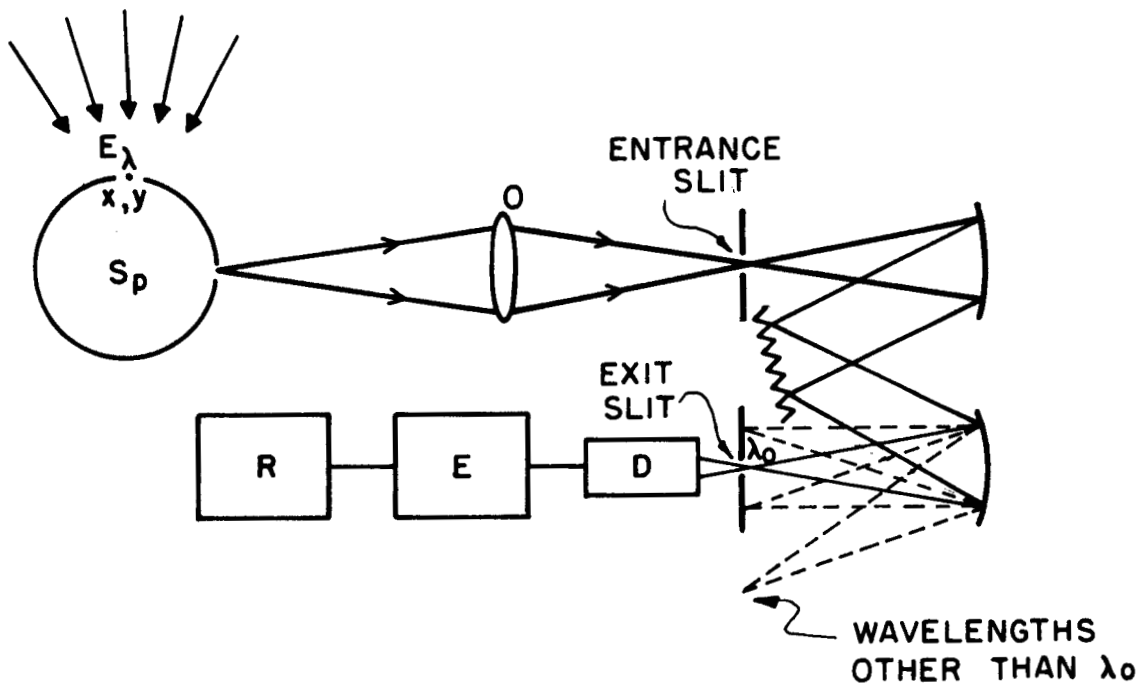


Figure 7.1. Schematic diagram of a typical spectroradiometer used for spectral-irradiance measurements. Monochromator is shown set for a wavelength λ_o. Note that, for simplicity, dispersion is shown as taking place at the second focusing (curved) mirror rather than at the ruled surface of the diffraction grating.

$E_\lambda(x,y)$ -- Spectral irradiance at the point x,y in the receiving aperture

S_p -- Averaging sphere

O -- Focusing optics

D -- Detector

E -- Electronics

R -- Readout

Chapter 7. The Relative Spectral Responsivity and Slit-Scattering Function of a Spectroradiometer

by Henry J. Kostkowski

In this CHAPTER. We are concerned with how the spectral responsivity of a spectroradiometer containing a monochromator varies with wavelength. The need for this information and some typical relative spectral responsivities are presented. The relationship between the spectral responsivity and the slit-scattering function, involving the responsivity factor, is developed, and a method for experimentally determining these quantities is outlined. Examples are given of the types of spectral distortions produced by the responsivity function, but details on how to correct for these distortions are postponed until Chapter 8. Finally the radiometric measurement of spectral-line radiation, which also requires spectral-responsivity information, is introduced.

NEED for the RELATIVE SPECTRAL RESPONSIVITY in SPECTRAL MEASUREMENTS.

When one is interested in making spectral radiometric measurements, particularly at many different wavelengths, a spectroradiometer containing a monochromator is usually used. If the quantity to be measured is spectral irradiance, a suitable instrument is the one used in problem 2 of Chapter 5 [7.1][1] and shown schematically in figure 7.1. The measurement equation for spectral irradiance when using such an instrument was shown, in Chapter 5, to be [eq. (5.23)]

$$S(A,\Delta\lambda,\lambda_o) = \int_{\Delta\lambda}\int_{A} E_\lambda(x,y,\lambda)\cdot R_\Phi(x,y,\lambda_o,\lambda)\cdot dA\cdot d\lambda \ [S], \tag{7.1}$$

where $S(A,\Delta\lambda,\lambda_o)$ and $R_\Phi(x,y,\lambda_o,\lambda)$ are the output signal and spectral-flux responsivity function[2], respectively, of the instrument for a wavelength setting of λ_o, A is the area of the receiving aperture, and $\Delta\lambda$ is the wavelength interval for which the value of R_Φ is not zero. For simplicity in the present discussion, we assume that the spectral irradiance E_λ and the responsivity R_Φ are each uniform throughout A, and we focus attention on just the spectral considerations for any given A by writing

$$S(\Delta\lambda,\lambda_o) = A\cdot\int_{\Delta\lambda} E_\lambda(\lambda)\cdot R_\Phi(\lambda_o,\lambda)\cdot d\lambda \ [S], \tag{7.1a}$$

---

[1]Figures in brackets indicate literature references listed at the end of *this chapter*.

[2]The word "function" is added to "spectral-flux responsivity" or "spectral-irradiance responsivity" when we wish to emphasize the spectral shape of these quantities. In general $R(x,y,\lambda_o,\lambda)$ may have a different spectral shape; i.e., be a different function of wavelength λ, for each wavelength setting λ_o of the instrument and for each point x,y. For brevity we will often refer to a term such as "spectral-flux responsivity function" as simply the responsivity function.

or

$$S(\Delta\lambda,\lambda_o) = \int_{\Delta\lambda} E_\lambda(\lambda)\cdot R_E(\lambda_o,\lambda)\cdot d\lambda \ [S] \tag{7.1b}$$

where $R_E(\lambda_o,\lambda) = A\cdot R_\Phi(\lambda_o,\lambda)$ $[S\cdot W^{-1}\cdot m^2]$ is the spectral-irradiance responsivity first introduced in Chapter 5 {eq. (5.49) [7.1]}.

In problem 2 of Chapter 5, we assumed that $\Delta\lambda$ was sufficiently small so that E_λ changed negligibly over this wavelength range. In this case

$$S(\Delta\lambda,\lambda_o) = E_\lambda(\lambda_o)\cdot\int_{\Delta\lambda} R_E(\lambda_o,\lambda)\cdot d\lambda \ [S]. \tag{7.2}$$

We also assumed that a similar equation would be valid when calibrating the instrument with a standard of spectral irradiance resulting in

$$S^s(\Delta\lambda,\lambda_o) = E^s_\lambda(\lambda_o)\cdot\int_{\Delta\lambda} R_E(\lambda_o,\lambda)\cdot d\lambda \ [S] \tag{7.3}$$

where the superscript s refers to the standard. Combining eqs. (7.2) and (7.3) we obtain

$$E_\lambda(\lambda_o) = \frac{S(\Delta\lambda,\lambda_o)}{S^s(\Delta\lambda,\lambda_o)}\cdot E^s_\lambda(\lambda_o) \ [W\cdot m^{-2}\cdot nm^{-1}]. \tag{7.4}$$

This is the simplest and most frequently used relation for spectral measurements when utilizing a spectroradiometer containing a monochromator. However, when $\Delta\lambda$ cannot be made sufficiently small for eq. (7.4) to be valid, we must utilize eq. (7.1b) directly in determining E_λ.

There are two ways of obtaining a solution for $E_\lambda(\lambda)$ from eq. (7.1b). When the relative spectral distributions of $E_\lambda(\lambda)$ and $R_E(\lambda_o,\lambda)$ are known, the solution can be obtained in a manner similar to that used in problem 3 of Chapter 5. Then, instead of eq. (7.4), the equation for E_λ is

$$E_\lambda(\lambda_o) = \frac{S(\Delta\lambda,\lambda_o)}{S^s(\Delta\lambda,\lambda_o)}\cdot\frac{\int_{\Delta\lambda} E^s_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda}{\int_{\Delta\lambda} e_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda}\cdot e_\lambda(\lambda_o) \ [W\cdot m^{-2}\ nm^{-1}] \tag{7.5}$$

where $r(\lambda_o,\lambda) = R_E(\lambda_o,\lambda)/K$ [dimensionless], $e_\lambda(\lambda) = E_\lambda(\lambda)/K_1$ [dimensionless], and the arbitrary factors $K[S\cdot W^{-1}\cdot m^2]$ and $K_1[W\cdot m^{-2}]$ are constants with respect to wavelength. The quantity $r(\lambda_o,\lambda)$ is called the relative responsivity function.

If the relative spectral irradiance $e_\lambda(\lambda)$ is not known, it is still possible to obtain $E_\lambda(\lambda_o)$. The technique for doing this utilizes eq. (7.1b) and is referred to as a

deconvolution. Now however, measurements are required at a number of wavelengths rather than at just the wavelength λ_o at which E_λ is desired. Also, the relative responsivity function must be known *at each of these wavelengths*. Details relative to the deconvolution technique will be presented in Chapter 8.

Even if eq. (7.4) is thought to be sufficiently accurate for obtaining E_λ in a particular application, it is highly desirable to confirm this with one of the two above described methods. Thus for reliable measurements of spectral irradiance or any other spectral quantity, the corresponding relative spectral responsivity is required.

TYPICAL and IDEALIZED RELATIVE RESPONSIVITY FUNCTIONS. The relative responsivity function $r(\lambda_o,\lambda)$ of a spectroradiometer may be obtained by irradiating the instrument, while it is set on a wavelength λ_o, with a succession of monochromatic beams having the same irradiance[1] but different wavelengths, which are distributed over a range large enough to cover the entire responsivity function. A plot of the resulting output signal against wavelength is the relative responsivity function $r(\lambda_o,\lambda)$. Figure 7.2 is an example of such a plot for the spectroradiometer in figure 7.1 when the width of the image of the entrance slit is equal to the width of the exit slit. In general, this function will be different for different wavelength settings λ_o.

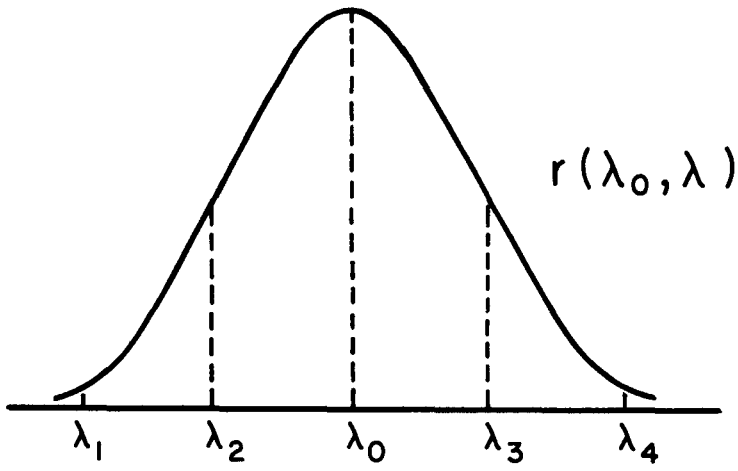


Figure 7.2. Relative responsivity function of a spectroradiometer containing a monochromator with the image of the entrance slit having the same width as the exit slit. The wavelength setting is λ_o. Wavelengths shown correspond to those in figure 7.3.

[1]We assume that the responsivity does not vary with the magnitude of the irradiance; that is, the instrument is "linear". Otherwise, the relative spectral responsivity determined in this manner would be valid only for the irradiance level used in the determination.

In order to understand why the typical responsivity function has the spectral shape shown in figure 7.2, let us examine closely the relative positions of the exit slit and the images of the entrance slit for a few of the monochromatic beams with wavelengths close to λ_o. The images we are talking about are located at the intersections, in the plane of the exit slit, of the optical rays shown as dashed lines in figure 7.1. Another view of these images, looking in a direction perpendicular to the plane of the exit slit, is shown in figure 7.3. Here the exit slit is represented by solid vertical lines and the monochromatic images of the entrance slit, each of different wavelength, by dashed lines. The cross hatching represents the radiant flux in the monochromatic image.

As can be seen in figure 7.3, when the wavelength of the monochromatic beam is λ_o, the wavelength setting of the instrument, flux fills the exit slit resulting in a signal from the instrument corresponding to the peak value shown in figure 7.2. When the monochromatic beam has a wavelength different than but close to λ_o, the flux in this image will only partially fill the exit slit. In figure 7.3, flux associated with wavelengths λ_2 and λ_3 fills half the exit slit and gives rise to about half the output signal or relative responsivity. The spectral propagance of the spectroradiometer between the entrance aperture and the exit slit, and the spectral responsivity of the detector, usually change only slightly over a small wavelength range such as λ_o-λ_2, and therefore, the output signal is approximately proportional to the portion of the exit slit filled with flux. In figure 7.3 it appears that the output signals for the monochromatic beams of wavelengths λ_1 and λ_4 should be zero. However, due to imperfect imaging, diffraction, and scattering, a small amount of flux of these wavelengths does fall on the exit slit and produces the output signal indicated in figure 7.2. Even when the image of the entrance slit is positioned far from the exit slit, as illustrated in figure 7.4b, some flux associated with this wavelength is scattered and diffracted onto the exit slit resulting in a small but significant output signal. That is, the relative responsivity function is not zero even for wavelengths such as λ_5 in figure 7.4a.

In the limit of perfect imagery and no scattering and when the instrument dispersion, detector responsivity, and propagance do not vary with wavelength over the wavelength range indicated in figure 7.3, the plot of the output signal against wavelength of the monochromatic beams would be triangular in shape. That is, the relative responsivity function would have the triangular shape shown in figure 7.5.

When the exit slit and the image of the entrance slit do not have the same width, the relative responsivity function has a spectral shape similar to that in figure 7.6 rather than that in figure 7.2. The sloping top is due to small variations with respect to wavelength in the detector responsivity and instrument propagance. Under the ideal conditions referred to in the previous paragraph, the spectral shape of this relative responsivity function would be trapezoidal. This can be confirmed by studying figures 7.7 and 7.8.

The width of the relative responsivity function is often of interest. It is called the spectral slit width with the symbol $\Delta\lambda_w$, as shown in figures 7.5, 7.7 and 7.8, and is

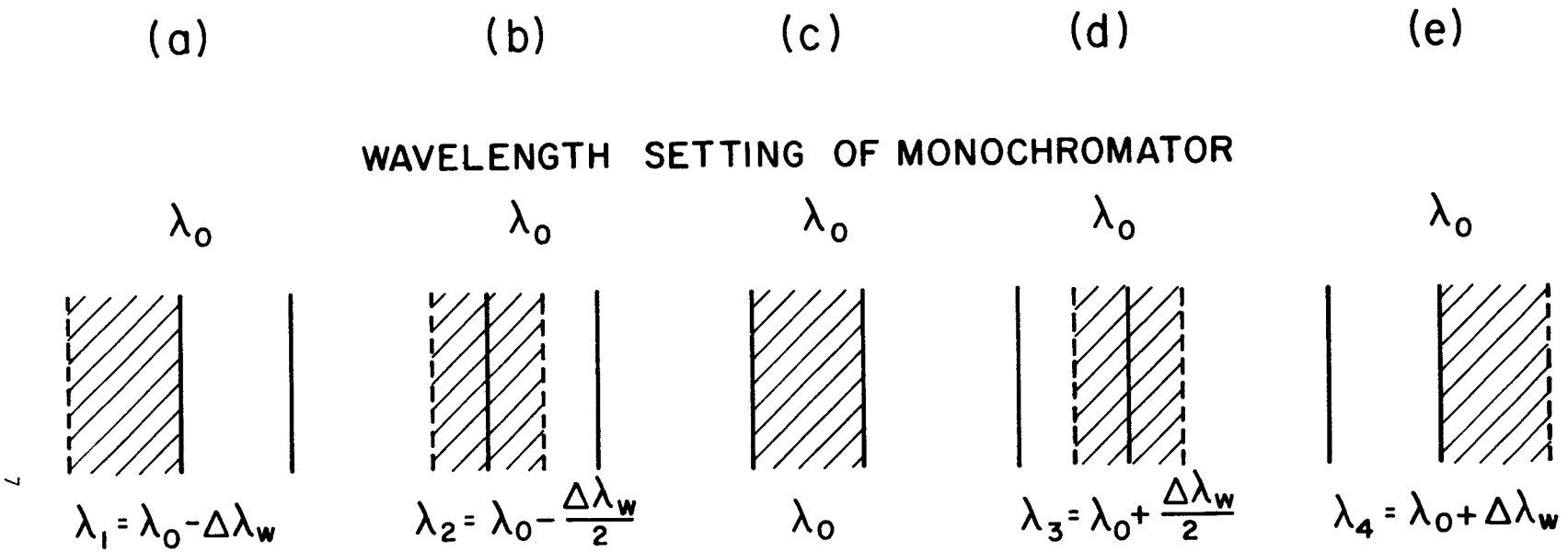


Figure 7.3. Relative positions of exit slit and image of entrance slit, showing why the responsivity function has a spectral shape approximating a triangle. The figure depicts a direct determination of the relative responsivity function using a series of monochromatic beams each having the same irradiance but a different wavelength. Dashed vertical lines represent the image of the entrance slit, solid vertical lines the equal sized exit slit, and crosshatching the monochromatic radiant flux of different wavelengths. The wavelength interval $\Delta\lambda_w$ is the nominal spectral slit width.

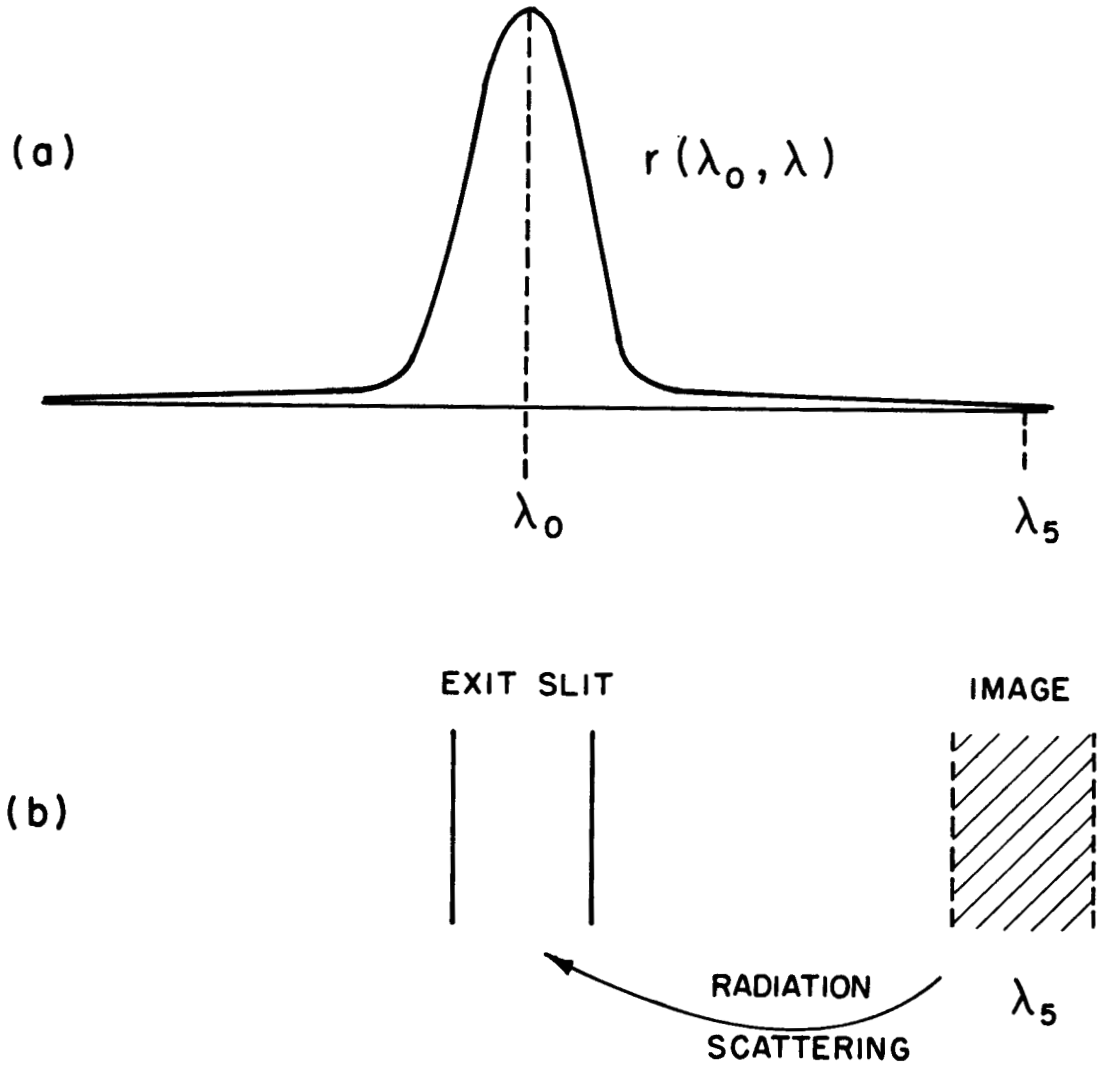


Figure 7.4. Illustration of radiation scattering giving rise to the wings of the relative responsivity function.

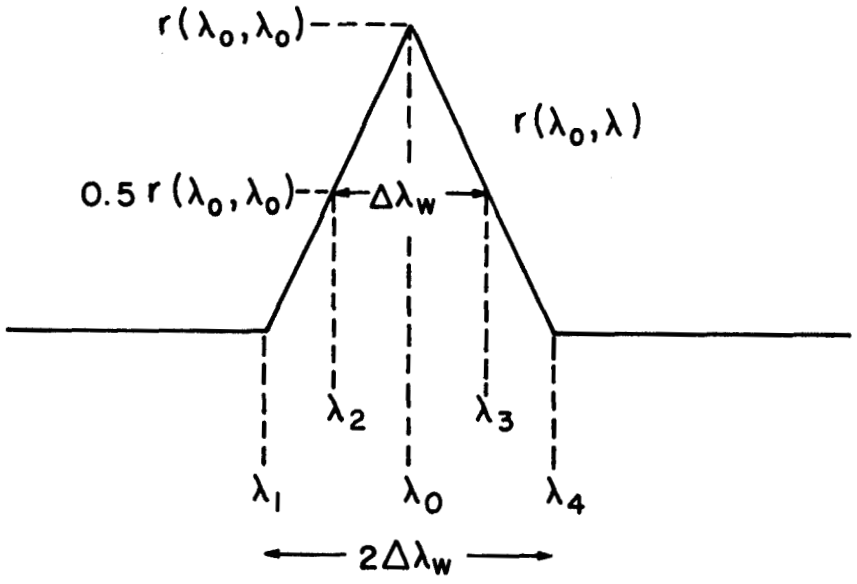


Figure 7.5. Triangular relative responsivity function with a spectral slit width of $\Delta\lambda_w$. The wavelengths correspond to those in figure 7.3.

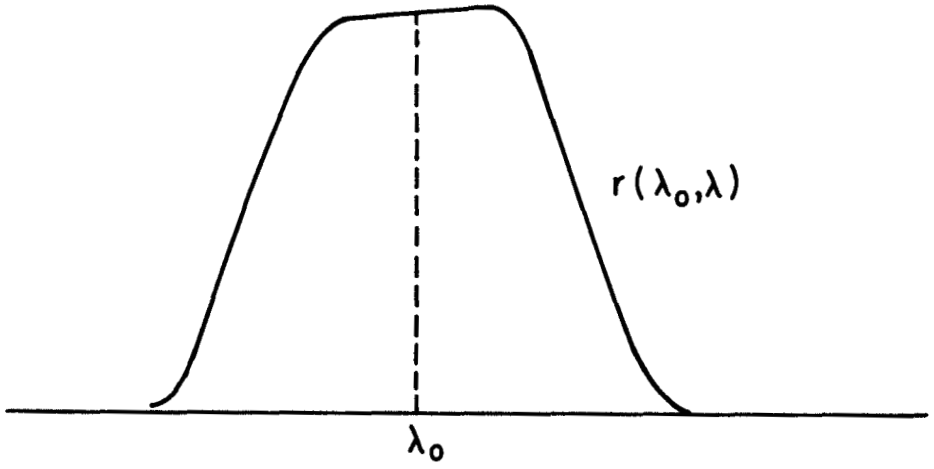


Figure 7.6. Relative responsivity function of a spectroradiometer containing a monochromator with the image of the entrance slit having a different width than the exit slit. Sloping top is due to change in detector responsivity and/or instrument propagance over this wavelength interval. Wavelength setting of spectroradiometer is λ_o.

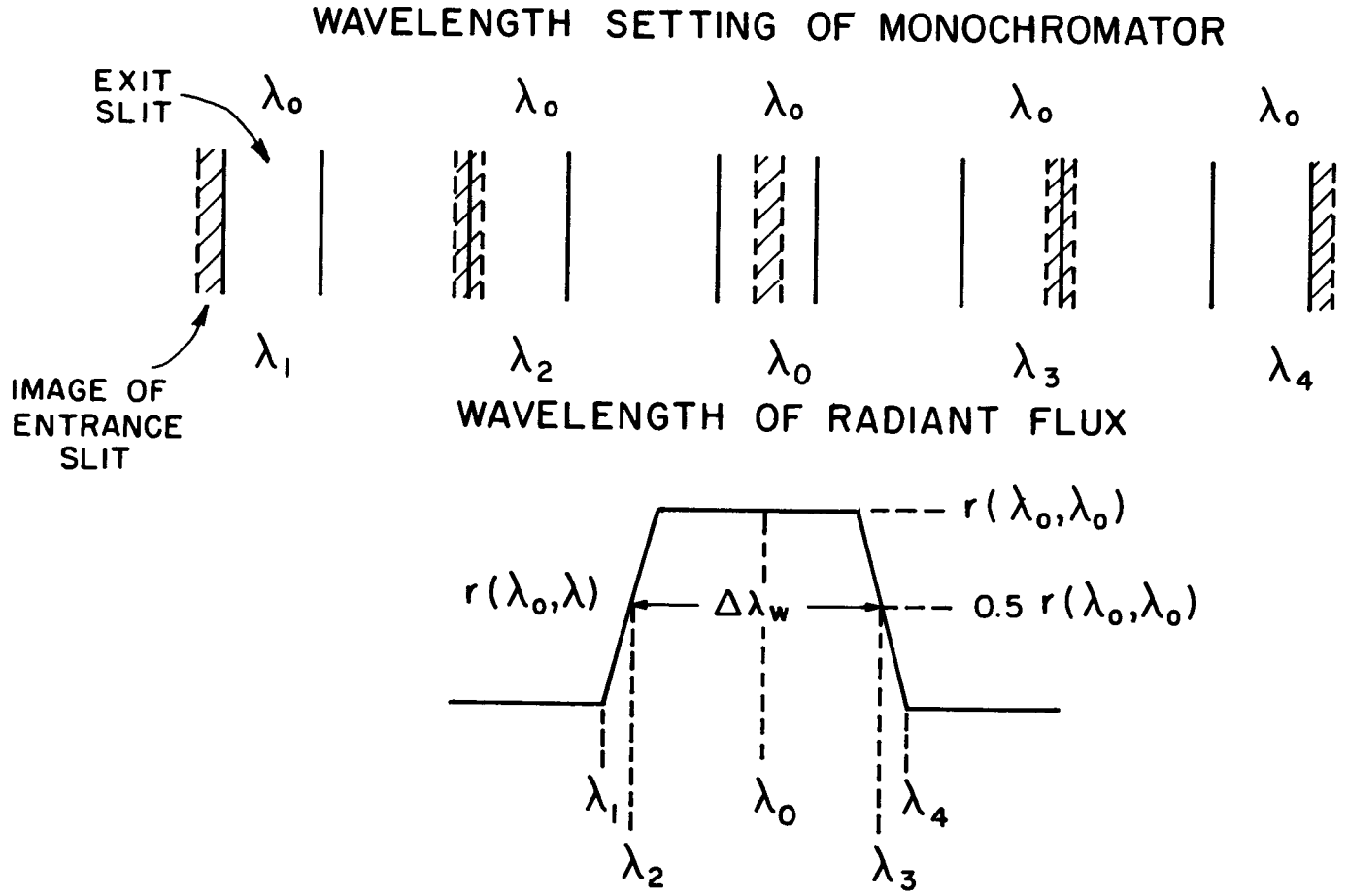


Figure 7.7. Trapezoidal relative responsivity function with entrance-slit image that is narrow compared to the exit slit. Spectral slit width is $\Delta\lambda_w$.

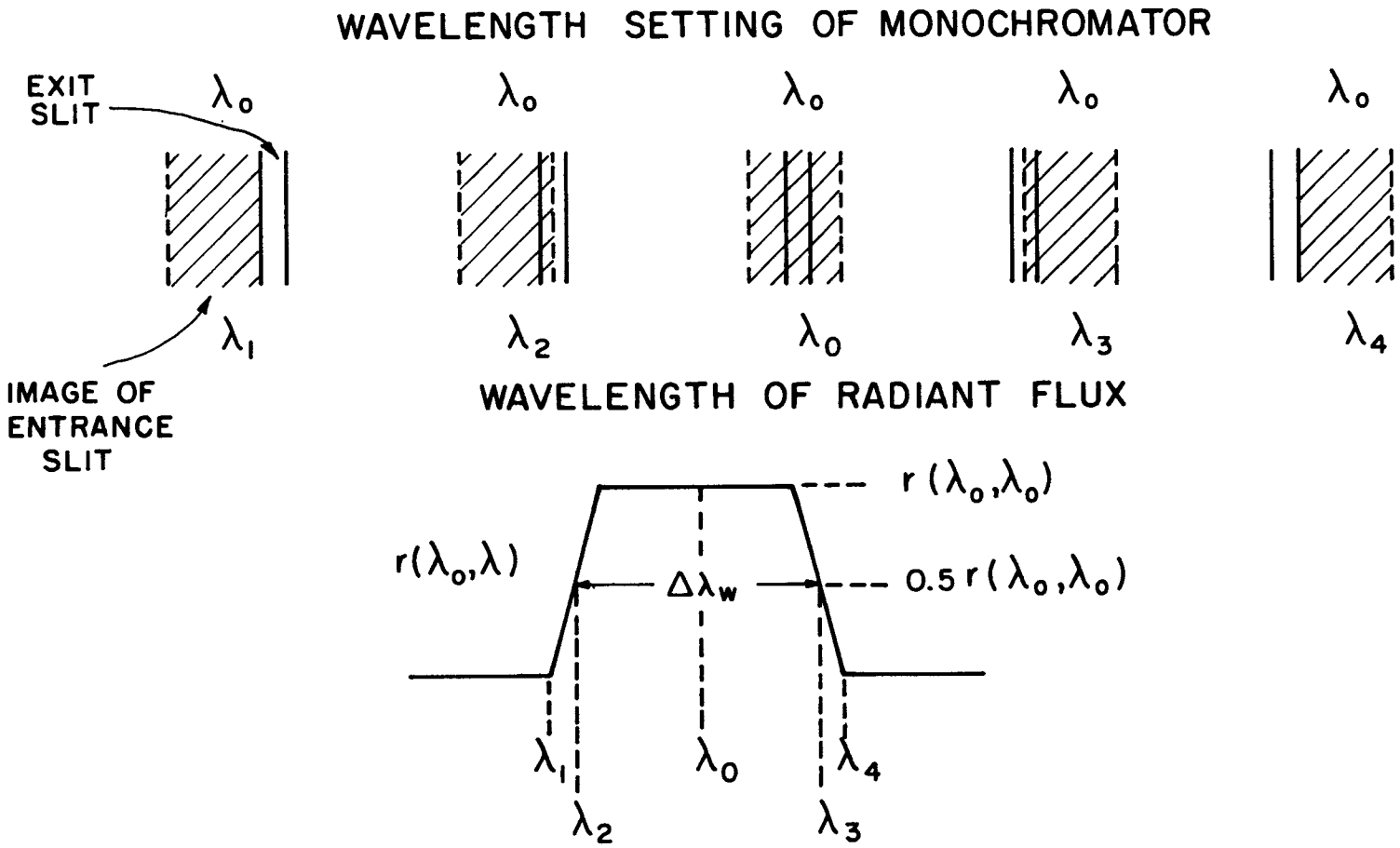


Figure 7.8. Trapezoidal relative responsivity function with entrance-slit image that is wide compared to the exit slit. Spectral slit width is $\Delta\lambda_w$.

defined as either (1) the width of the function at half its peak height or (2) as the *ratio* of the integral of the function with respect to wavelength to the peak height of the function; i.e., $\int_{-\infty}^{\infty} r(\lambda_o,\lambda)d\lambda / r(\lambda_o,\lambda_o)$. For the idealized functions, triangular or trapezoidal, the two definitions result in the same value. For the triangular shape

$$\Delta\lambda_w \equiv \Delta x \Big/ \frac{dx}{d\lambda} \text{ [nm]} \tag{7.6}$$

where Δx is the width of the exit slit, and $dx/d\lambda$ is the linear dispersion of the instrument. For the trapezoidal responsivity function

$$\Delta\lambda_w \equiv \Delta x_\ell \Big/ \frac{dx}{d\lambda} \text{ [nm]} \tag{7.7}$$

where Δx_ℓ is either the width of the entrance slit image or the exit slit, whichever is larger. The width of the base $\Delta\lambda_B$ of the trapezoidal responsivity function is

$$\Delta\lambda_B = \left(\Delta x_{\substack{\text{image of} \\ \text{ent. slit}}} + \Delta x_{\substack{\text{exit} \\ \text{slit}}}\right) \Big/ \frac{dx}{d\lambda} \text{ [nm]} \tag{7.8}$$

and the width of the flat top $\Delta\lambda_T$ of this function is

$$\Delta\lambda_T = \left|\Delta x_{\substack{\text{image of} \\ \text{ent. slit}}} - \Delta x_{\substack{\text{exit} \\ \text{slit}}}\right| \Big/ \frac{dx}{d\lambda} \text{ [nm]}. \tag{7.9}$$

These various equations may be confirmed by studying figures 7.3, 7.5, 7.7 and 7.8. From eqs. (7.8) and (7.9), one sees that as one of the slits becomes very narrow relative to the other, the relative responsivity plot approaches a rectangle.

DIRECT DETERMINATION of RELATIVE RESPONSIVITY FUNCTION using SPECTRALLY TUNABLE MONOCHROMATIC BEAMS. The relative responsivity function may be obtained experimentally by the technique described in the preceding section; that is, irradiating the spectroradiometer with a monochromatic beam whose wavelength can be varied over the wavelength range for which the magnitude of the responsivity function is not zero and whose irradiance at the receiving aperture of the instrument can be kept constant or can be measured as the wavelength is changed. The validity of this technique is easily confirmed starting with the spectral measurement equation,

$$S(\Delta\lambda,\lambda_o) = \int_{\Delta\lambda} E_\lambda(\lambda)\cdot R_E(\lambda_o,\lambda)\cdot d\lambda \text{ [S]} \tag{7.1b}$$

or

$$S(\Delta\lambda,\lambda_o) = K\int_{\Delta\lambda} E_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda \text{ [S]} \tag{7.10}$$

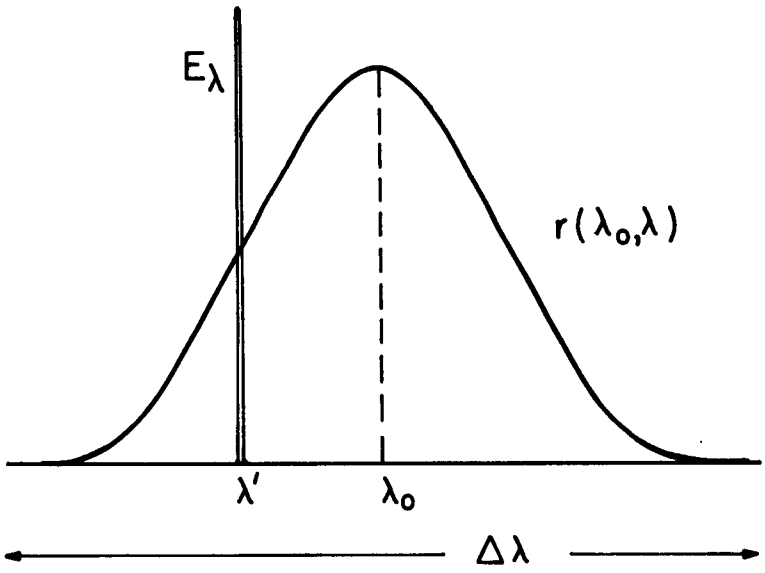


Figure 7.9. Illustration of why the output signal of a spectroradiometer viewing a narrow spectral line is proportional to the relative responsivity function $r(\lambda_o,\lambda')$, where λ_o is the wavelength setting of the instrument and λ' is the wavelength of the line. Varying λ' throughout the wavelength range $\Delta\lambda$ while the irradiance of the beam is unchanged, results in a determination of the relative responsivity function for the instrument wavelength setting λ_o.

where $r(\lambda_o,\lambda)$ is the relative responsivity function and K is a constant with respect to wavelength, both introduced in connection with eq. (7.5). Figure 7.9 shows a typical relative responsivity function and the spectral irradiance of a monochromatic beam, represented by a very narrow spectral line at λ'. The spectral line should be sufficiently narrow so that $r(\lambda_o,\lambda)$ may be treated as constant, reducing eq. (7.10) to

$$S(\Delta\lambda'_{line},\lambda_o) = K\cdot r(\lambda_o,\lambda')\cdot\int_{\Delta\lambda'_{line}} E_\lambda(\lambda)\cdot d\lambda \ [S] \qquad (7.11)$$

where the integral in eq. (7.11) is the irradiance of the monochromatic beam at the receiving aperture of the radiometer. Thus the output signal is proportional to the relative responsivity function for the wavelength setting of the spectroradiometer λ_o and evaluated at the wavelength of the monochromatic beam λ'. Varying λ' throughout the wavelength

range[1] $\Delta\lambda$ while the irradiance of the beam is unchanged, results in a determination of the relative responsivity function for the instrument wavelength setting λ_o.

This direct method of determining the relative responsivity function has been used infrequently because an easy-to-use, spectrally tunable, monochromatic beam has not been available. The only tunable monochromatic incoherent beam currently available is that emerging from the exit slit of a monochromator. The entrance slit of the monochromator is irradiated with the image of a high radiance tungsten ribbon lamp or carbon arc and the exiting beam is monitored with a detector calibrated for relative spectral flux responsivity. This is certainly not a convenient, easy-to-use source. Alignment of the components is critical, the source is not readily portable, and it is not as spectrally narrow or as high in irradiance as is often desired for relative responsivity measurements. However, as a result of the recent availability of small, high-throughput double monochromators and of silicon detector standards of spectral responsivity with 1.5-5% uncertainty, this direct method, using a tunable incoherent beam, is more appealing now than it was a few years ago.

A coherent source that can be used in the direct method is the tunable dye laser. This monochromatic beam can be obtained with a much narrower spectral band width and with higher irradiance than that from a monochromator. However, it is less portable, not available at all wavelengths, and significantly more expensive. In addition, in using it, one must take steps [7.2] to insure that the resulting relative responsivity function is not affected by the coherence of the beam. Another possible problem is that some lasers seem to include radiation at wavelengths other than the primary wavelength with irradiances of about 10^{-6} that of the primary spectral line. However, this interfering radiation can be measured and removed by a filter or dispersive device when it presents a problem. In spite of these limitations, the tunable dye laser appears to be an attractive device for this application.

INDIRECT DETERMINATION of RELATIVE RESPONSIVITY FUNCTIONS using FIXED WAVELENGTH MONOCHROMATIC BEAMS -- SLIT-SCATTERING FUNCTION and RESPONSIVITY FACTOR.

Most measurements of the relative responsivity functions of spectroradiometers containing monochromators have been made using an indirect method where a fixed-wavelength, monochromatic beam and a standard source of continuous radiation are spectrally scanned with the monochromator. This indirect method is simpler to carry out and requires less expensive apparatus than the direct method, but certain conditions must exist for the resulting relative responsivity function to be accurate. In this section these conditions and two new quantities, the slit-scattering function and the responsivity factor, that are required in the indirect method, are discussed in considerable detail.

An insight into why and when the simpler indirect method works can be obtained from examining, again, the relative positions of the exit slit and the image of the entrance slit

---

[1]Note that $\Delta\lambda$ is the wavelength interval in which $r(\lambda_o,\lambda) \neq 0$; $\Delta\lambda'$ is the much narrower interval in which $E_{\lambda_{line}} \neq 0$; $\Delta\lambda' << \Delta\lambda$.

for this method and for the direct method presented in the last section. Figures 7.3 and 7.4 give these positions for the direct method where the spectroradiometer is set on the wavelength λ_o and output signals are obtained for an incident monochromatic beam whose wavelength is changed throughout $\Delta\lambda$, the wavelength range for which the responsivity function has a significant value. We showed that when the irradiance of these beams was constant the output signals were proportional to the relative responsivity function. Figure 7.10 gives the slit positions for the indirect method where the wavelength of the monochromatic beam is fixed at λ_o and the wavelength setting of the spectroradiometer is changed (scanned) throughout $\Delta\lambda$. The relative positions of the exit slits and the images of the entrance slits in figures 7.3 and 7.10 are the same. Also, even though the *wavelengths* associated with the corresponding "look-alike" figures are different, except for case c, the wavelength differences are, respectively, the same. This is the case because it was assumed, in preparing figure 7.10, that the dispersion of the instrument was the same for all the wavelengths indicated. This, then, is one of the required conditions for the indirect method to apply. Another condition, which is not apparent from the two figures, is that the optical aberrations, scattering and diffraction must be the same for the look-alike diagrams in figures 7.3 and 7.10. Even if these conditions hold, however, the relative output signals corresponding to each of the two figures may be different. This is because the spectral propagance of the monochromator and of the fore optics and the spectral responsivity of the detector will not, in general, be the same for all the wavelengths indicated. However, when the above conditions hold, these factors depend only on the wavelength of the flux and not on the wavelength setting of the monochromator.

All the above conditions that are required for the indirect method to apply can be incorporated into one mathematical statement. This is that the relative responsivity function can be written as a product of two functions, one depending only on the difference between the wavelength setting of the monochromator and the wavelength of the flux and the other function depending only on the wavelength of the flux; that is,

$$r(\lambda_o,\lambda) = z(\lambda_o-\lambda)\cdot r^f(\lambda) \text{ [dimensionless]}, \tag{7.12}$$

where we call z the slit-scattering function[1,2] and r^f the responsivity factor. The

[1]The slit-scattering function in eq. (7.12) has often been referred to as simply the slit function. However, we are not continuing this practice because the CIE committee TC-1.2 on Photometry and Radiometry has proposed that the name "slit function" be used for the "relative spectral transmittance of a *monochromator* as a function of wavelength for a given setting of wavelength and slits". In addition, our new name for $z(\lambda_o-\lambda)$ is more descriptive because, although the central portion of this function is determined primarily by the slit widths and dispersion of the monochromator, the wings of the function are determined primarily by internal radiation scattering.

(See page 17 for Footnote [2])

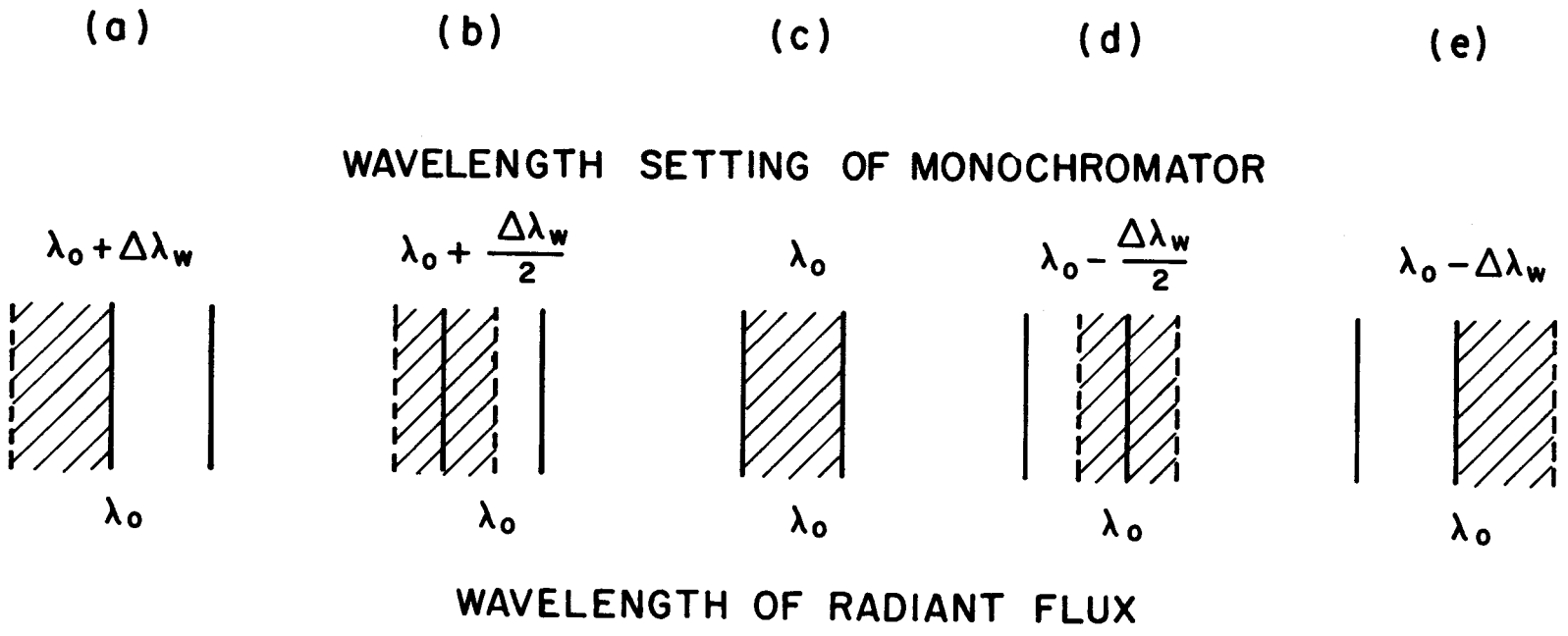


Figure 7.10. Relative positions of exit slit and image of entrance slit when a spectroradiometer spectrally scans a monochromatic beam of wavelength λ_o for determining the slit-scattering function at this wavelength. Image of entrance slit is represented by dashed vertical lines, the equal sized exit slit by solid vertical lines and the λ_o monochromatic radiant flux by crosshatching.

[2]Sometimes one sees the argument of the slit-scattering function in eq. (7.12) written $\lambda-\lambda_o$ rather than $\lambda_o-\lambda$. Physically, this difference is unimportant. We have adopted $\lambda_o-\lambda$ because this is the form in which convolution integrals are normally expressed in an extensive literature on deconvolution which is cited in Chapter 8. A minor disadvantage of this form is that it establishes the rather unappealing sign convention by which the argument of the slit function is negative for the long wavelength portion of the function and vice versa. However, we felt that the extensive use of this form in various mathematical theorems and in the deconvolution literature outweighed this slight disadvantage.

essence of the indirect method and eq. (7.12) is that the relative responsivity function for a particular wavelength setting of a monochromator λ_o can be obtained by varying λ_o rather than by varying λ as in the conventional direct method. Since the value of the slit-scattering function z, when the $\lambda_o-\lambda$ dependence in eq. (7.12) is valid, depends only on the difference between λ_o and λ, it is immaterial in determining this function whether one varies λ_o or λ. The advantage of varying λ_o is that then $r^f(\lambda)$ is constant, and therefore a spectral scan of a single, fixed-wavelength, monochromatic beam (constant λ) can be used to obtain z. The explicit relationship between the slit-scattering function $z(\lambda_o-\lambda)$ and the output signal of a spectroradiometer when scanning[1] a narrow, isolated spectral line can be derived by starting with the spectral portion of the measurement equation

$$S(\Delta\lambda,\lambda_o) = \int E_\lambda(\lambda)\cdot R_E(\lambda_o,\lambda)\cdot d\lambda \ [S]. \tag{7.1b}$$

Assuming that $R_E(\lambda_o,\lambda) = K\cdot r(\lambda_o,\lambda) = K\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)$ $[S\cdot W^{-1}\cdot m^2]$, we obtain

$$S(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)\cdot d\lambda \ [S]. \tag{7.13}$$

Letting $E_\lambda(\lambda)$ be the spectral irradiance of a narrow spectral line of wavelength λ_{line} and designating the wavelength setting of the spectroradiometer as λ', as indicated in figure 7.11,

$$S(\Delta\lambda_{line},\lambda') = K\cdot\int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot z(\lambda'-\lambda)\cdot r^f(\lambda)\cdot d\lambda \ [S] \tag{7.14}$$

where $\Delta\lambda_{line}$ is the wavelength range for which the spectral irradiance of the line is significant. If both $z(\lambda'-\lambda)$ and $r^f(\lambda)$ change negligibly over $\Delta\lambda_{line}$,

[1]The terms "scanning" and "spectral scan" are consistently used here for an instrumental scan, where the wavelength setting of the instrument is varied or "scanned".

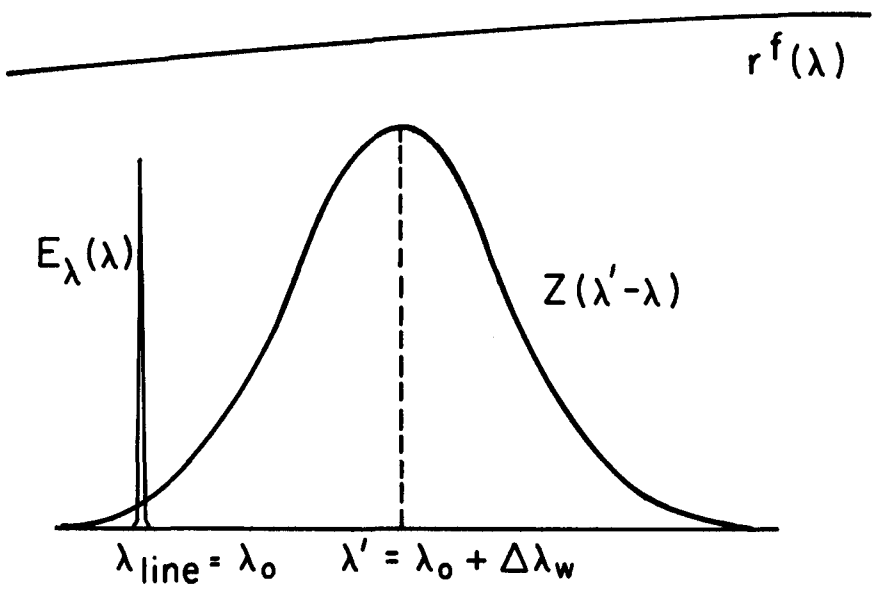


Figure 7.11. Spectral plots of a narrow spectral line, a typical slit-scattering function, and a responsivity factor, indicating why a spectral scan of such a line with the spectroradiometer results in the mirror image of the slit-scattering function and not the slit-scattering function itself. Note that the values shown for λ_{line} and λ' are those for the situation depicted in figure 7.10(a).

$$S(\Delta\lambda_{line},\lambda') = K\cdot z(\lambda'-\lambda_{line})\cdot r^f(\lambda_{line})\cdot\int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot d\lambda \ [S] \tag{7.15}$$

and

$$z(\lambda'-\lambda_{line}) = C\cdot S(\Delta\lambda_{line},\lambda') \ [\text{dimensionless}] \tag{7.16}$$

where C is a constant for a given λ_{line}. Since one usually is interested only in relative values of z, the constant C is frequently set equal to the reciprocal of the peak value of S, making the peak value of z unity. Varying the wavelength setting of the instrument λ' over a sufficiently wide range provides, from eq. (7.16), a determination of the slit-scattering function $z(\lambda'-\lambda_{line})$. Implicit in this determination is that the slit-scattering function does not change when the wavelength setting of the instrument, in this case λ', changes.

Note that when $\lambda' > \lambda_{line}$, as indicated in figure 7.11, the monochromator wavelength setting is in the long wavelength wing of the spectral line. However the output signal for this setting is proportional to the value of the slit-scattering function evaluated at λ_{line} which is in the short wavelength wing (where $(\lambda'-\lambda)$ is positive) of the slit-scattering function. This spectral mirror-image relationship between the scan of a spectral line (monochromatic beam) and the slit-scattering function is caused by the fact that in the indirect method the wavelength setting of the spectroradiometer is varied rather than the wavelength of the monochromatic beam as in the direct method. It might help to clarify this situation further if we compare figures 7.3a and 7.10a. Figure 7.3a depicts the direct determination of the relative responsivity function at a wavelength which is $\Delta\lambda_w$ less than the peak wavelength λ_o. For the equivalent of figure 7.3a in the indirect method, the wavelength setting must still be greater than the wavelength of the flux by $\Delta\lambda_w$. However, since the wavelength of the flux in the indirect method is fixed and in this case, as shown in figure 7.10a, equal to λ_o, the wavelength setting in figure 7.10a must be $\lambda_o+\Delta\lambda_w$. Moreover, in terms of the spectral scan of a spectral line representing the required monochromatic beam, the spectral position represented by figure 7.10a is in the long wavelength portion of the scan at a wavelength greater than the peak wavelength. Thus the long wavelength side or wing of the spectral scan of the monochromatic beam corresponds to the short wavelength side or wing of the slit-scattering function and vice versa.

In order to determine the relative responsivity function from eq. (7.12) one must determine the responsivity factor r^f as well as the slit-scattering function. It is easy to demonstrate how the responsivity factor can be obtained from a spectral scan of a spectral standard of continuous radiation. Starting again with the spectral portion of the measurement equation

$$S(\Delta\lambda,\lambda_o) = \int_{\Delta\lambda} E_\lambda(\lambda)\cdot R_E(\lambda_o,\lambda)\cdot d\lambda \ [S]. \tag{7.1b}$$

We assume that $R_E(\lambda_o,\lambda) = K\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)$ and $E_\lambda(\lambda) = E^s_\lambda(\lambda)$, the spectral irradiance of a standard of continuous radiation. Thus eq. (7.1b) becomes

$$S^s(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E^s_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)\cdot d\lambda \ [S], \tag{7.17}$$

and if $S^s(\Delta\lambda,\lambda_o)$ is determined for a sufficiently large range and number of monochromator wavelength settings λ_o, eq. (7.17) can be solved for $E^s_\lambda\cdot r^f$ by deconvolution (see Chapter 8). Since E^s_λ is known, this will also be a solution for the responsivity factor r^f.

Sometimes a sufficiently accurate solution for r^f can be obtained without a deconvolution. If $E^s_\lambda\cdot r^f$ is approximately linear over the wavelength range for which $z(\lambda_o-\lambda)$ is significant and z is approximately symmetrical with respect to λ_o, eq. (7.17) can be written as

$$S^s(\Delta\lambda,\lambda_o) \simeq K\cdot E^s_\lambda(\lambda_o)\cdot r^f(\lambda_o)\cdot\int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda \ [S] \tag{7.18}$$

and

$$r^f(\lambda_o) \simeq \frac{S^s(\Delta\lambda,\lambda_o)}{K\cdot E^s_\lambda(\lambda_o)\cdot\int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda} \ [\text{dimensionless}] \tag{7.19}$$

where all the wavelength-dependent quantities on the right-hand side of eq. (7.19) are known or measurable.

Obtaining the relative responsivity function by the indirect method, that is, via the slit-scattering function and responsivity factor, is significantly easier than by the direct method. Only a single wavelength monochromatic beam and a spectral standard of continuous radiation are required. The single-wavelength monochromatic beam can be obtained from an incoherent source having a strong, isolated, narrow spectral line such as the 253.7 [nm] line of mercury or from a single laser spectral line, provided coherence effects are adequately accounted for [7.2]. The major problem or limitation of the indirect method is the validity of the $\lambda_o-\lambda$ dependence in eq. (7.12). In the relatively small wavelength range corresponding to the central portion of the relative responsivity function (approximately $\lambda_o \pm \Delta\lambda_w$), eq. (7.12) is generally quite accurate. In addition, in this central region, the spectral line can usually have a very different wavelength than the λ_o for which the relative responsivity function is desired. The major requirement is that the monochromator dispersion at the two wavelengths not be significantly different. The situation is quite different, however, for the distant wings of the responsivity function. We have obtained spectral scans of monochromatic beams with wavelengths 90 [nm] apart where the spectral values differed by 100% at 50 [nm] from the peak. Of course, if eq. (7.12) had been valid over this wavelength range, the spectral shape of these scans would have been the same. The differences are mainly due to a difference in internal scattering. In addition, spectral scans of monochromatic beams sometimes exhibit spectral bumps and weak spectral lines 25 [nm] and more from the peak of the scan. We believe these are due, respectively, to multiple reflections from the sides and various parts of the instrument and to grating ghosts. The wavelengths of the grating ghosts have a clear-cut regularity and can be predicted from measurements of a few monochromatic beams with different wavelengths, but this does not appear to be the case with their amplitudes. Thus, in general, the indirect method is not very accurate for determining the distant wings of the relative responsivity function.

It appears that spectral scans of monochromatic beams using monochromators possessing high-quality holographic (photo-resist) gratings are much smoother than those with ruled gratings. With monochromators having these gratings, it may be possible to define a more general slit-scattering function and apply the indirect method even when the $\lambda_o-\lambda$ dependence of eq. (7.12) does not exist. Research directed at this possibility is being pursued by the author.

EXAMPLES of an EXPERIMENTALLY DETERMINED RESPONSIVITY FACTOR and of SLIT-SCATTERING FUNCTIONS. Since the responsivity factor of a spectroradiometer depends on the spectral propagance of the instrument and its detector responsivity, the spectral shape of this factor will vary greatly with instruments having different types of optical elements and detectors. However, a fairly common spectral shape for the responsivity factor of a spectroradiometer is shown in figure 7.12. This is a plot of the responsivity factor for the NBS spectroradiometer used to realize and maintain the NBS scale of spectral irradiance [7.3].

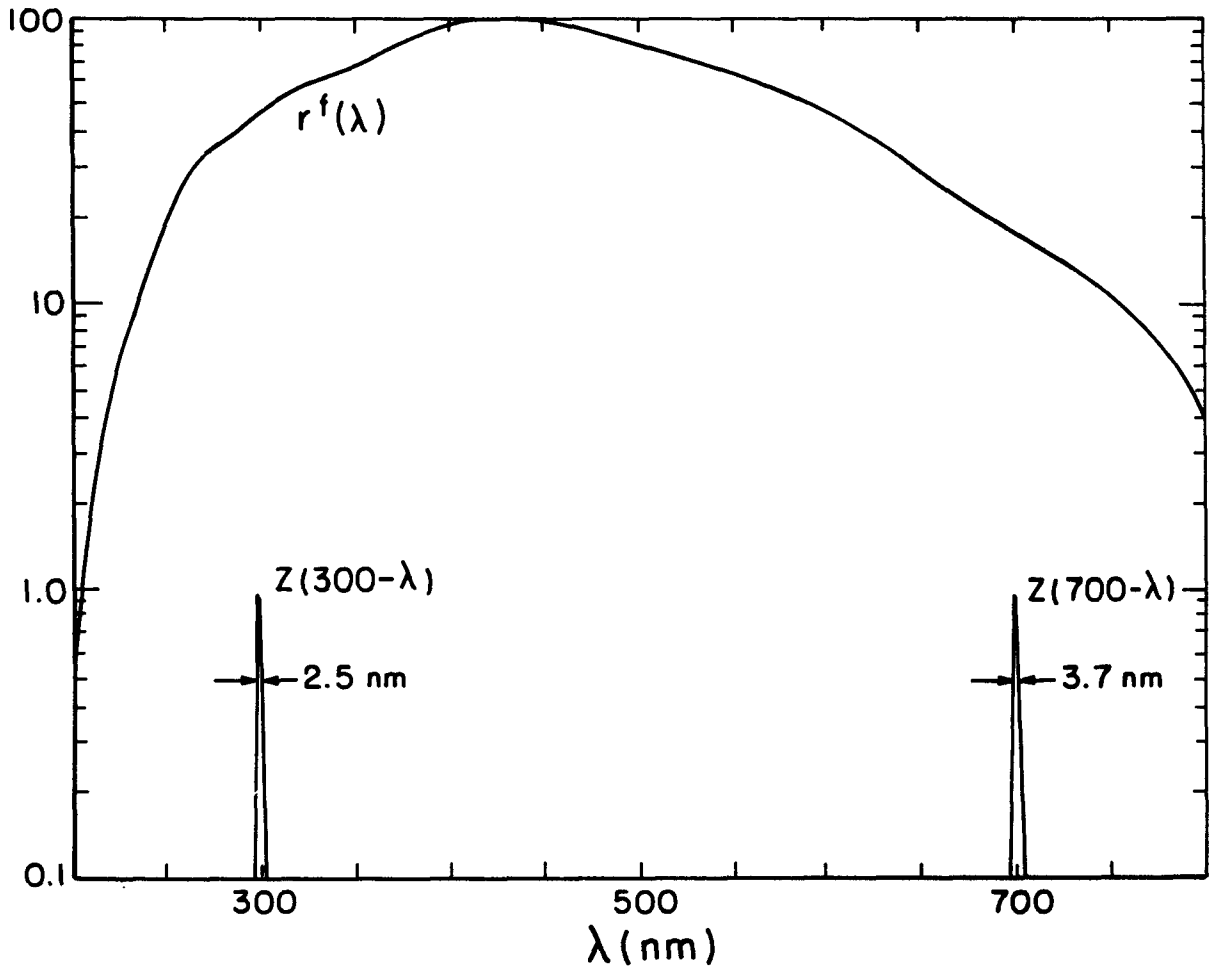


Figure 7.12. The responsivity factor and central portions of slit-scattering functions for a spectroradiometer consisting of a double monochromator with a quartz prism and grating, a photomultiplier detector with a tri-alkali cathode, and an averaging-sphere receiving element coated on the inside with $BaSO_4$. The slit widths are set at one millimeter.

The responsivity factor was determined with eq. (7.19) using data from a spectral scan of a deuterium lamp standard and a tungsten quartz halogen lamp standard and the manufacturer's published data on the dispersion of the instrument. This particular instrument is similar to the one shown in figure 7.1 except that the foreoptics utilizes mirrors and the dispersing portion of the instrument is a double monochromator containing a quartz prism and a grating. The inside surface of the averaging sphere in the NBS instrument is coated with barium sulphate and the detector is a photomultiplier with a tri-alkali cathode and quartz window. The rapid decrease in the responsivity factor below 300 [nm], that is shown in figure 7.12, is primarily due to the averaging sphere. The decrease at long wavelengths is largely due to the spectral responsivity of the detector. The relative spectral efficiency of the grating also contributes to the short- and long-wavelength falloffs.

Also shown in figure 7.12 for the NBS spectroradiometer are the central portions of the slit-scattering functions determined at 300 [nm] and 700 [nm] for a slit width of one millimeter. The dispersion of the double monochromator of this instrument, and therefore the spectral slit width of the instrument, varies by about 50% over this wavelength range, as indicated. An all-grating instrument would have a much smaller change in dispersion over this wavelength range and an all-prism instrument a much larger change.

Figure 7.13 shows the long wavelength portion of the slit-scattering functions at 254 [nm] for two single grating monochromators with a ruled and holographic grating, respectively, and for a double monochromator with two holographic gratings. These were obtained by spectrally scanning a low pressure mercury source between 200 [nm] and 254 [nm]. In the case of the double monochromator, a short wavelength cut-off filter was also used because radiation from the mercury source for other than the 254 [nm] line was also significant. There was continuum and line radiation below 250 [nm] that produced a signal ranging from 10 to 500 times that produced by the slit-scattering function; i.e., scattering of the 254 [nm] line. This situation was even worse in the long wavelength wing of the 254 [nm] mercury line; that is, in the short wavelength wing of the slit function which is not shown in the figure. The values of the 254 [nm] slit-scattering functions in the wings at about 50 [nm] from their peak are 4×10^{-5} and 10^{-8} for the single and double holographic grating instruments, and are representative values in the UV for good-quality instruments.

RESPONSIVITY-FUNCTION EFFECTS. The effects of the responsivity function on spectroradiometric measurements can be illustrated by performing calculations that simulate the measurement of a few characteristic spectral irradiance distributions. We use the simplified measurement equation

$$S(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda \ [S] \tag{7.10}$$

and the corresponding equation for a calibration measurement

$$S^s(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E^s_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda \ [S] \tag{7.10a}$$

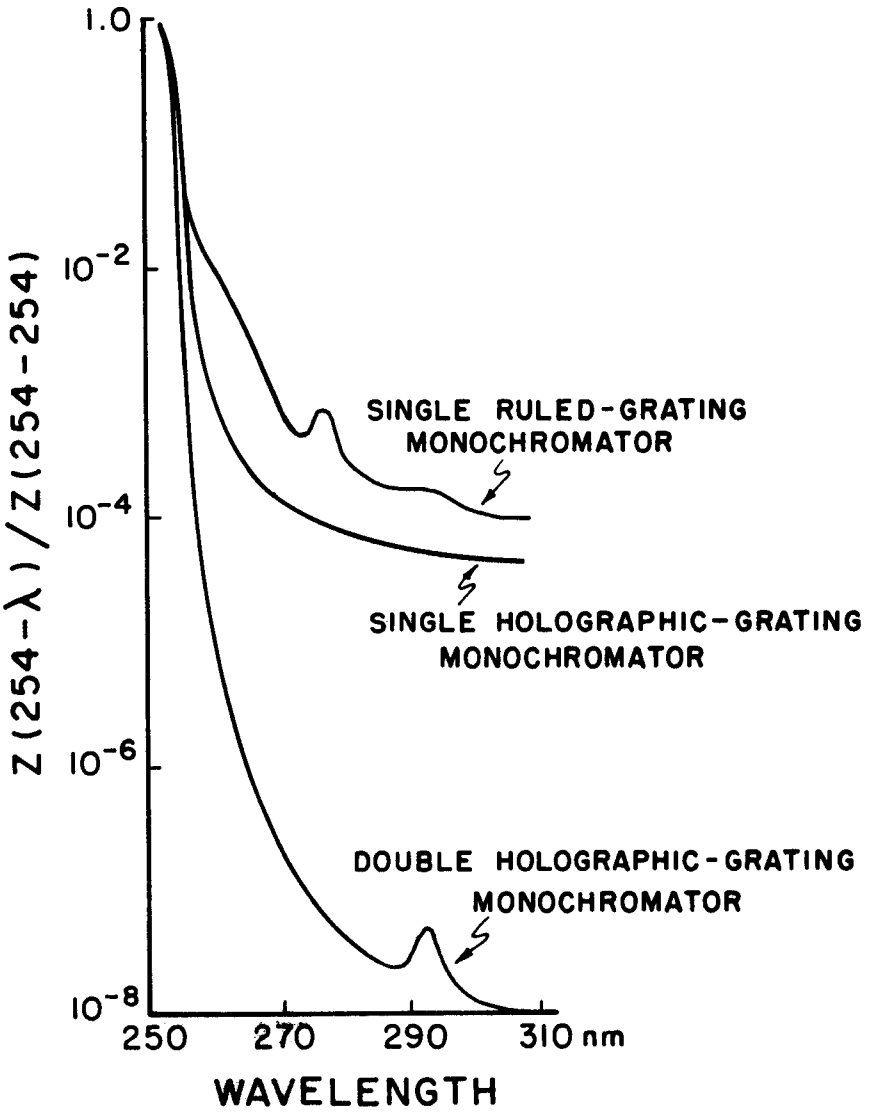


Figure 7.13. Long-wavelength side of experimentally determined slit-scattering functions for three different spectroradiometers, obtained by using the mercury line at 254 [nm] and scanning between 254 and 200 [nm].

for these calculations. In order that the calculations be independent of the spectral shape of the standard, we assume that $E^s_\lambda(\lambda)$ is constant over $\Delta\lambda$. Then

$$S^s(\Delta\lambda,\lambda_o) = K\cdot E^s_\lambda(\lambda_o)\cdot\int_{\Delta\lambda} r(\lambda_o,\lambda)\cdot d\lambda \ [S]. \tag{7.20}$$

Dividing eq. (7.10) by eq. (7.20) and multiplying by E^s_λ we obtain

$$\frac{S}{S^s}\cdot E^s_\lambda = \frac{\int_{\Delta\lambda} E_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda}{\int_{\Delta\lambda} r(\lambda_o,\lambda)d\lambda} \ [W\cdot m^{-2}\cdot nm^{-1}]. \tag{7.21}$$

But the left-hand side of eq. (7.21) is just the irradiance we would obtain [eq. (7.4)] from measurements of S and S^s without taking into account the convolution of $E_\lambda(\lambda)$ and $r(\lambda_o,\lambda)$. Designating this "observed" irradiance as $E^o_\lambda(\lambda_o)$, we have

$$E^o_\lambda(\lambda_o) = \frac{\int_{\Delta\lambda} E_\lambda(\lambda)\cdot r(\lambda_o,\lambda)\cdot d\lambda}{\int_{\Delta\lambda} r(\lambda_o,\lambda)\cdot d\lambda} \ [W\cdot m^{-2}\cdot nm^{-1}]. \tag{7.22}$$

Thus by normalizing the measurement equation with the integral of the responsivity function, the result is the irradiance we would obtain if we utilized the simple equation

$$E_\lambda(\lambda_o) = \frac{S}{S^s}\cdot E^s_\lambda(\lambda_o) \ [W\cdot m^{-2}\cdot nm^{-1}]. \tag{7.4}$$

Figure 7.14 illustrates the result of calculating E^o_λ from eq. (7.22) for the relative responsivity functions and spectral irradiance distributions shown in the figure. Figure 7.14a is the typical spectral scattering situation. A substantial value is obtained for E^o_λ for wavelengths less than λ_1, even though the actual spectral irradiance is zero or very small at these wavelengths. The integrand in the numerator of eq. (7.22) is zero for wavelengths up to λ_1 (see figure 7.14a) because $E_\lambda(\lambda)$ is zero in this range. But from λ_1 to longer wavelengths $E_\lambda(\lambda)$ becomes so large that, even though $r(\lambda_o,\lambda)$ is, for the most part, very small in most of this region, the cumulative contribution of their product to the integral over all wavelengths can be very large. Physically, $E^o_\lambda(\lambda_o)$ represents an accumulation over this large wavelength range, of the scattering, illustrated earlier in figure 7.4b for one wavelength, λ_5. An extreme example of spectral scattering of this type occurs when making terrestrial solar irradiance measurements below 320 [nm]. Absorption of solar radiation by atmospheric ozone produces an E_λ distribution similar to that shown in figure 7.14a. Spectral irradiance measurements at 300 [nm] when utilizing a single monochromator and at 290 [nm] with a double monochromator are typically too large by about a factor of two when the responsivity function is not taken into account; that is, when eq. (7.22) or (7.4) is used.

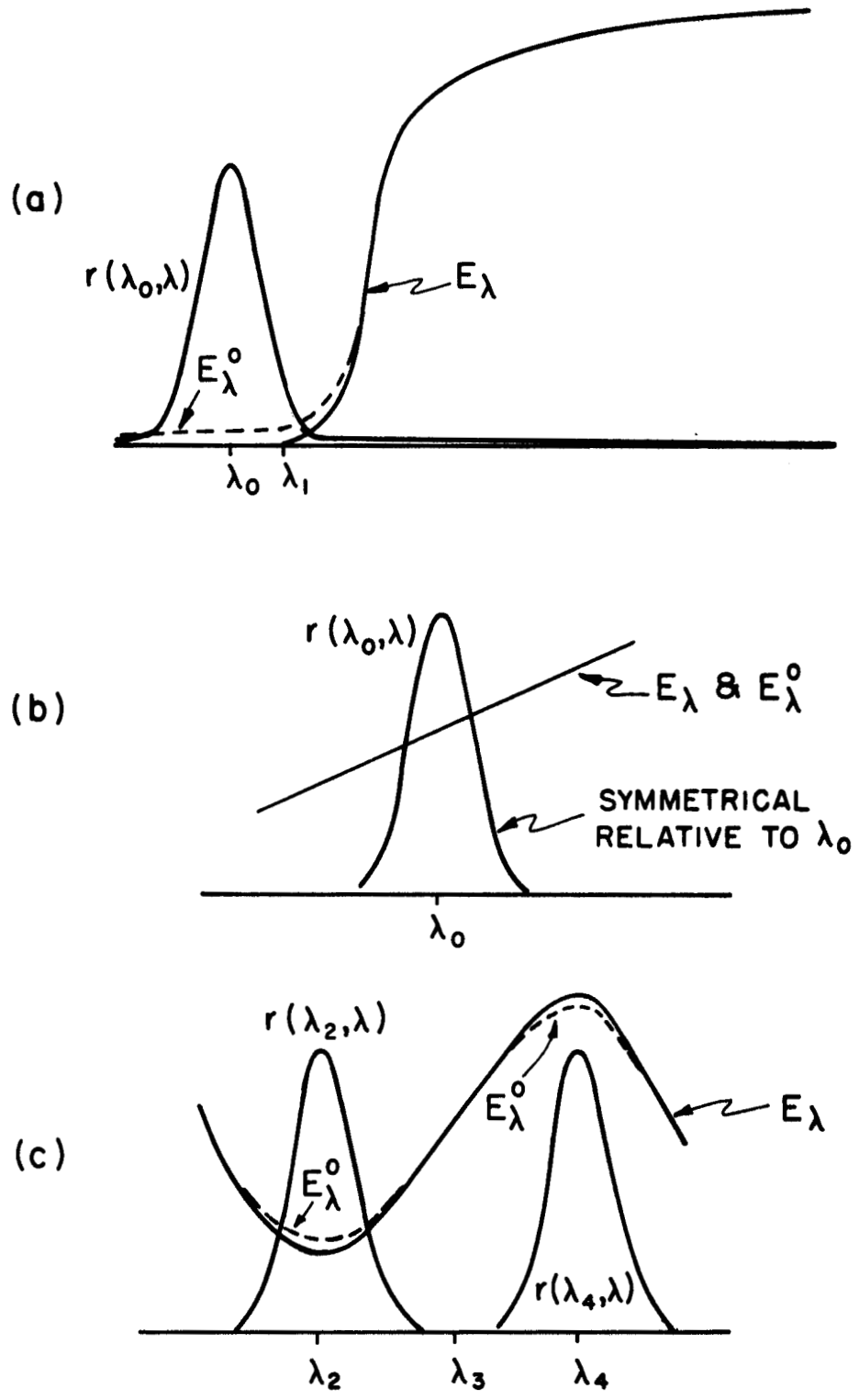


Figure 7.14. Examples of how the responsivity function distorts or modifies the spectral distribution being measured.

If the spectral distribution being measured is linear over the wavelength range $\Delta\lambda$ covered by the relative responsivity function $r(\lambda_o,\lambda)$ and this function is symmetrical with respect to λ_o, as pictured in figure 7.14b, no distortion results. The observed spectral irradiance is identical to that being measured. This is so because E^o_λ, as seen in eq. (7.22), is simply the weighted mean of E_λ over $\Delta\lambda$, weighted by the relative responsivity function $r(\lambda_o,\lambda)$. For the linearity and symmetry in figure 7.14b, the result at λ_o is the same as having an E_λ that is constant over $\Delta\lambda$. Put another way, the smaller contribution on the "low" E_λ side (where $\lambda < \lambda_o$) is exactly compensated by the larger contribution on the "high" E_λ side (where $\lambda > \lambda_o$).

For the spectral distribution $E_\lambda(\lambda)$ shown in figure 7.14c, the observed irradiance obtained from eq. (7.22) will be larger at λ_2 and smaller at λ_4 than the actual spectral irradiances at these wavelengths. There is no compensation as in figure 7.14b. The weighted mean of E_λ at λ_2 is higher and at λ_4 lower than the actual values at these wavelengths. At λ_3, $E_\lambda(\lambda)$ is linear and the situation is similar to that in figure 7.14b so that $E^o_\lambda(\lambda_3)$ is equal to $E_\lambda(\lambda_3)$. Note that the curvature at any point on the observed spectral distribution in figure 7.14c is either the same as or less than that of the actual or true spectral distribution.

Estimates for the magnitude of the distortion produced in any particular measurement can be calculated using eq. (7.22) if the relative responsivity function and an estimated spectral distribution for the quantity being measured are available. In fact such calculations form the basis of one deconvolution technique described in Chapter 8.

When either the RADIOMETRIC QUANTITY or the RESPONSIVITY FUNCTION or both VARY with POSITION and DIRECTION. The treatment of the responsivity and slit-scattering function in this chapter has assumed that both the radiometric quantity being measured and the responsivity function are uniform and isotropic throughout the beam of radiation incident on the monochromator. Ordinarily this is easily achieved when measuring spectral irradiance by incorporating an averaging sphere for the first optical element of the spectroradiometer. In this section, we examine the consequences when such an element is not used.

Let us address this situation by starting with the general measurement equation[1] {eq. (5.30) [7.1]}

$$S(\Delta\lambda,A,\omega,\lambda_o) = \int_{\Delta\lambda}\int_{A}\int_{\omega} L_\lambda(x,y,\theta,\phi,\lambda)\cdot R_\Phi(x,y,\theta,\phi,\lambda_o,\lambda)\cdot\cos\theta\cdot d\omega\cdot dA\cdot d\lambda \ [S]. \qquad (7.23)$$

When the responsivity is uniform and isotropic throughout A and ω, but the spectral radiance is not, then $R_\Phi(x,y,\theta,\phi,\lambda_o,\lambda) = R_\Phi(\lambda_o,\lambda)$ and eq. (7.23) may be written as

$$S(\Delta\lambda,A,\omega,\lambda_o) = \int_{\Delta\lambda} \bar{L}_\lambda(\lambda)\cdot R_L(\lambda_o,\lambda)\cdot d\lambda \ [S] \qquad (7.24)$$

---

[1]We are still assuming that polarization does not have an effect; that is, either the incident beam is unpolarized or the spectroradiometer is not polarization sensitive.

where

$$\bar{L}_\lambda(\lambda) = \frac{\int_A \int_\omega L_\lambda(x,y,\theta,\phi,\lambda)\cdot\cos\theta\cdot d\omega\cdot dA}{\int_A \int_\omega \cos\theta\cdot d\omega\cdot dA} \quad [W\cdot m^{-2}\cdot sr^{-1}\cdot nm^{-1}] \tag{7.25}$$

is the weighted average of the spectral radiance in A and ω, weighted with respect to $\cos\theta$, and $R_L(\lambda_o,\lambda) = R_\phi(\lambda_o,\lambda)\cdot\int_A \int_\omega \cos\theta\cdot d\omega\cdot dA$. Since eq. (7.24) is equivalent to eq. (7.1b), the responsivity and slit-scattering function material developed in this chapter for spectral irradiance also applies to the weighted average spectral radiance of eq. (7.25), as long as the responsivity is uniform and isotropic. When the responsivity is not uniform and isotropic but the spectral radiance is; that is, $L_\lambda(x,y,\theta,\phi,\lambda) = L_\lambda(\lambda)$, eq. (7.23) becomes

$$S(\Delta\lambda,A,\omega,\lambda_o) = \int_{\Delta\lambda} L_\lambda(\lambda)\cdot R_L(A,\omega,\lambda_o,\lambda)\cdot d\lambda \quad [S] \tag{7.26}$$

where

$$R_L(A,\omega,\lambda_o,\lambda) = \int_A \int_\omega R_\phi(x,y,\theta,\phi,\lambda_o,\lambda)\cdot\cos\theta\cdot d\omega\cdot dA \quad [S\cdot W^{-1}\cdot cm^2\cdot sr] \tag{7.27}$$

is the radiance responsivity. Equation (7.26) is also equivalent to eq. (7.1b), and therefore the material of this chapter also applies to the case of a non-uniform, non-isotropic flux responsivity provided the spectral radiance is uniform and isotropic throughout the beam being measured.

When neither L_λ nor R_ϕ is uniform and isotropic throughout the beam of interest, the beam to be measured must be reduced in size (A and ω) until at least one of the above quantities is sufficiently uniform and isotropic for the accuracy required. This amounts to determining the spectral radiance (and responsivity) as a function of direction and position and requires many more measurements and more difficult measurements than when uniformity and isotropy exist.

MEASUREMENT of SPECTRAL-LINE RADIATION. A chapter on the relative responsivity or slit-scattering function of a spectroradiometer would not be complete without at least an introduction to the measurement of spectral-line radiation. In radiometry it is usually not necessary to obtain the spectral distribution of a spectral line but only the radiance of the line radiation or the irradiance produced by the line radiation. These are defined[1] as

$$L(\lambda_{line}) = \int_{\Delta\lambda_{line}} L_\lambda(\lambda)\cdot d\lambda \quad [W\cdot m^{-2}\cdot sr^{-1}] \tag{7.28}$$

and

[1] In eqs. (7.28) and (7.29), it is assumed that there is no continuum radiation present.

$$E(\lambda_{line}) = \int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot d\lambda \ [W\cdot m^{-2}]. \tag{7.29}$$

In order to illustrate the manner in which the measurement equation is solved when determining the irradiance produced by narrow spectral-line radiation, we assume that the spectroradiometer is set on the wavelength λ_o, corresponding to the peak of the line, and that eq. (7.13) repeated below is valid,

$$S(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)\cdot d\lambda \ [S]. \tag{7.13}$$

A plot of the various functions used in determining $E(\lambda_{line})$ is shown in figure 7.15.

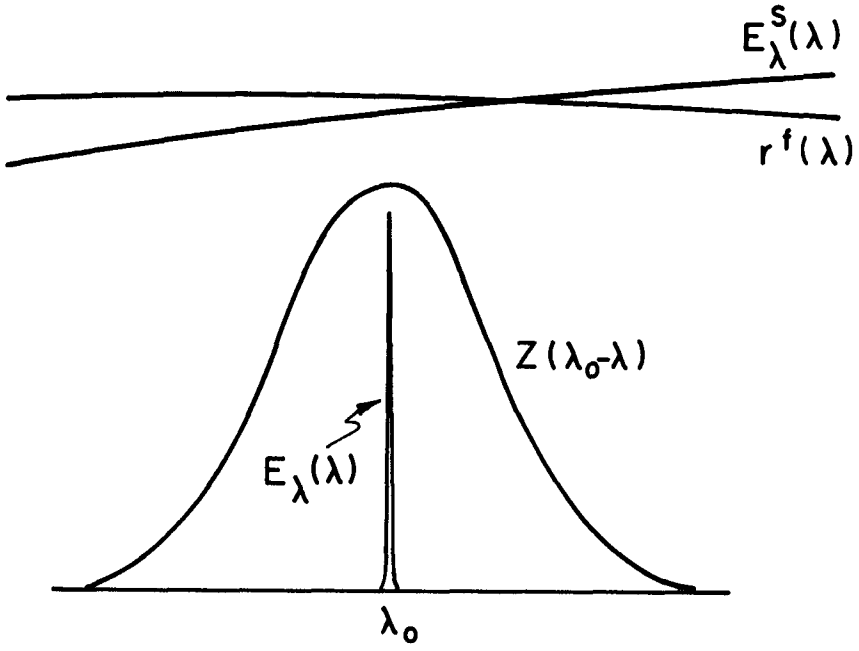


Figure 7.15. Representative functions when determining the irradiance produced by a narrow spectral-line source.

Using these functions, eq. (7.13) reduces to

$$S(\Delta\lambda,\lambda_o) \simeq K\cdot z(\lambda_o-\lambda_o)\cdot r^f(\lambda_o)\cdot\int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot d\lambda \ [S] \tag{7.30}$$

where $r^f(\lambda)$ and $z(\lambda_o-\lambda_o)$ $\{= z(0)\}$ have been removed from within the integral because they vary negligibly over the wavelength range of the spectral line $\Delta\lambda_{line}$. The responsivity factor evaluated at λ_o and the constant K can be eliminated by making a measurement of a standard of spectral irradiance at λ_o. In such a measurement

$$S^s(\Delta\lambda,\lambda_o) = K\cdot\int_{\Delta\lambda} E^s_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)\cdot d\lambda \ [S] \tag{7.13a}$$

and with the usual approximations

$$S^s(\Delta\lambda,\lambda_o) \simeq K\cdot r^f(\lambda_o)\cdot E^s_\lambda(\lambda_o)\cdot\int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda \ [S]. \tag{7.31}$$

Combining eqs. (7.30) and (7.31),

$$E(\lambda_{line}) \equiv \int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot d\lambda \simeq \frac{S(\lambda_o)}{S^s(\lambda_o)}\cdot E^s_\lambda(\lambda_o)\cdot\Delta\lambda_{eff} \ [W\cdot m^{-2}], \tag{7.32}$$

where

$$\Delta\lambda_{eff} = \frac{\int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda}{z(\lambda_o-\lambda_o)} \ [nm]. \tag{7.33}$$

All the quantities on the right-hand side of eqs. (7.32) and (7.33) are known or can be measured. If eq. (7.31) is not sufficiently accurate, S^s may be obtained for a large number of wavelengths throughout $\Delta\lambda$ in order to determine $K\cdot r^f$ from a deconvolution of eq. (7.13a). The $E(\lambda_{line})$ may then be determined from eqs. (7.29) and (7.30).

If the spectral line is too wide to be treated in the above manner or if there are two or more spectral lines within the central portion of the slit-scattering function, as shown in figures 7.16a or 7.16b, one can obtain an approximation for the irradiance produced by the line or lines by spectrally scanning and integrating the resulting output signal over the wavelength region for which the signal is significant. This is demonstrated mathematically in the following manner. Integrating eq. (7.13) over the wavelength range $\Delta\lambda_o$ for which the signal $S(\Delta\lambda,\lambda_o)$ is significant,

$$\int_{\Delta\lambda_o} S(\Delta\lambda,\lambda_o)\cdot d\lambda_o = K\cdot\int_{\Delta\lambda_o}\int_{\Delta\lambda} E_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot r^f(\lambda)\cdot d\lambda\cdot d\lambda_o \ [S\cdot nm]. \tag{7.34}$$

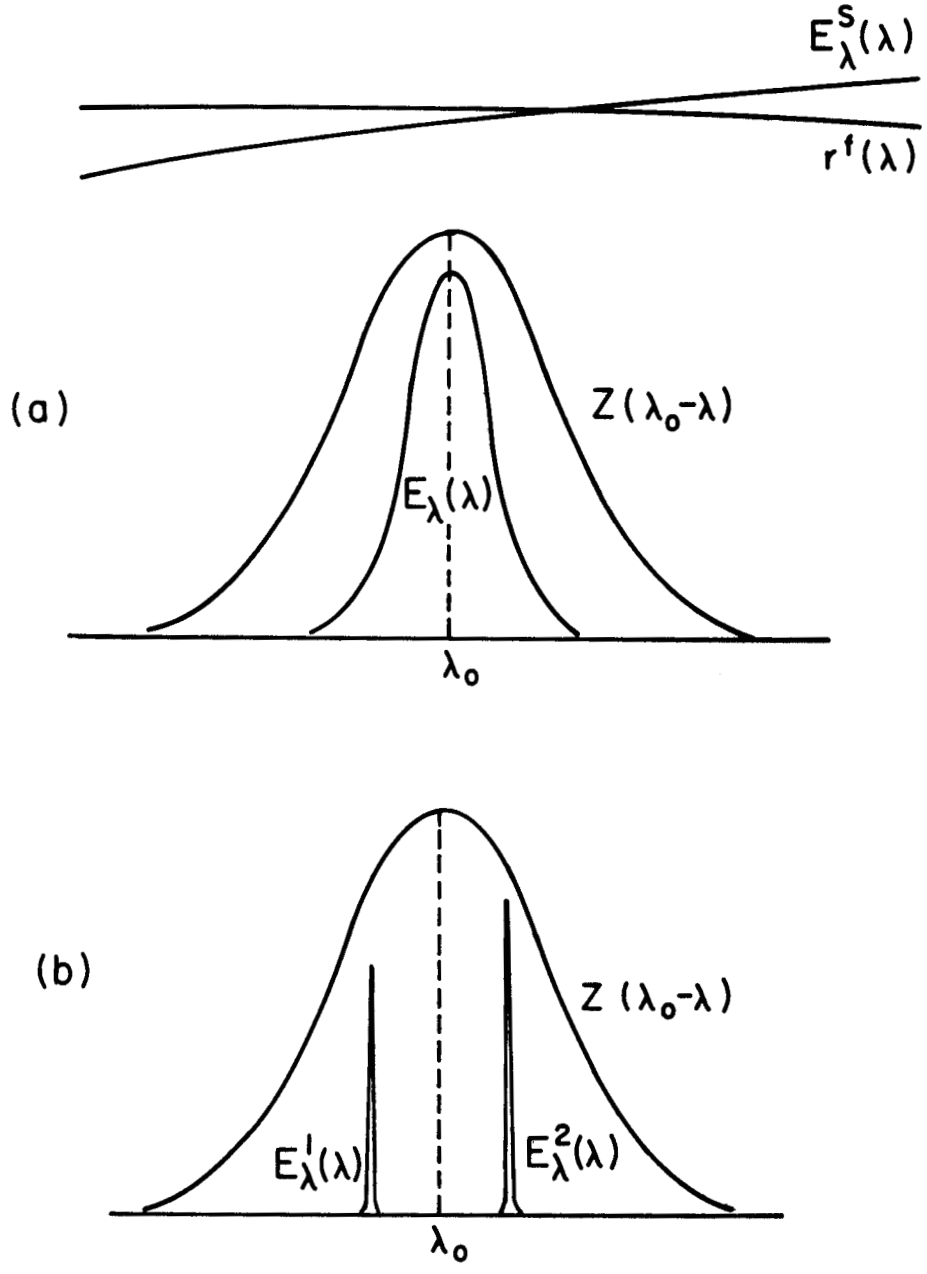


Figure 7.16. Representative functions when determining the irradiance produced by a source emitting a wide spectral line or two narrow spectral lines separated by a wavelength interval comparable to the spectral slit width.

Assuming that the responsivity factor does not change significantly over either integration interval, the responsivity factor can be removed from within both integrals and evaluated at the center of the $\Delta\lambda_o$ scan at wavelength λ_{mean},

$$\int_{\Delta\lambda_o} S(\Delta\lambda,\lambda_o)\cdot d\lambda_o \simeq K\cdot r^f(\lambda_{mean})\cdot\int_{\Delta\lambda_o}\int_{\Delta\lambda} E_\lambda(\lambda)\cdot z(\lambda_o-\lambda)\cdot d\lambda\cdot d\lambda_o \ [S\cdot nm]. \tag{7.35}$$

The only factor in the integrand of eq. (7.35) which depends upon λ_o is $z(\lambda_o-\lambda)$. The integral of this slit-scattering function over the entire range $\Delta\lambda_o$ of λ_o values for which there is any instrument response, yields a fixed number independent of λ, which can be removed from the integral over λ. Thus we obtain

$$\int_{\Delta\lambda_o} S(\Delta\lambda,\lambda_o)\cdot d\lambda_o \simeq K\cdot r^f(\lambda_{mean})\cdot\int_{\Delta\lambda_o} z(\lambda_o-\lambda)\cdot d\lambda_o \int_{\Delta\lambda} E_\lambda(\lambda)\cdot d\lambda \ [S\cdot nm]. \tag{7.36}$$

The responsivity factor at λ_{mean} and the constant K in eq. (7.36) may be obtained by a measurement of a spectral irradiance standard at λ_{mean}. This measurement is represented by a slight modification of eq. (7.31) or

$$S^s(\Delta\lambda,\lambda_{mean}) \simeq K\cdot r^f(\lambda_{mean})\cdot E^s_\lambda(\lambda_{mean})\cdot\int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda \ [S]. \tag{7.31a}$$

Note that the integration limits of the slit-scattering function integrals in eq. (7.31a) and (7.36) are different. However, since the magnitude of $\Delta\lambda$ was chosen sufficiently large so as to include any significant portion of the relative responsivity function {or $z(\lambda_o-\lambda)$}, integration of $z(\lambda_o-\lambda)$ over the larger wavelength range $\Delta\lambda_o$ is negligibly different from that over $\Delta\lambda$. That is, from eq. (7.33),

$$\int_{\Delta\lambda_o} z(\lambda_o-\lambda)\cdot d\lambda_o \simeq \int_{\Delta\lambda} z(\lambda_o-\lambda)\cdot d\lambda = z(0)\cdot\Delta\lambda_{eff} \ [nm]. \tag{7.37}$$

Combining eqs. (7.36), (7.31a) and (7.37) results in

$$\int_{\Delta\lambda} E_\lambda(\lambda)\cdot d\lambda = \frac{\int_{\Delta\lambda_o} S(\Delta\lambda,\lambda_o)\cdot d\lambda_o}{S^s(\Delta\lambda,\lambda_{mean})}\cdot E^s_\lambda(\lambda_{mean}) \ [W\cdot m^{-2}]. \tag{7.38}$$

Refinements and additional details on the determination of the irradiance produced by sources containing individual or multiple spectral lines and a mixture of lines and continua are planned for a future chapter of the Self-Study Manual.

SUMMARY of CHAPTER 7. For reliable spectral irradiance measurements, it is necessary to know the relative spectral responsivity function $r(\lambda_o,\lambda)$ in the spectral portion of the measurement equation {eq. (7.1b)}

$$S(\Delta\lambda,\lambda_o) = \int_{\Delta\lambda} E_\lambda(\lambda)\cdot K\cdot r(\lambda_o,\lambda)\cdot d\lambda \ [S], \tag{7.10}$$

where $K\cdot r(\lambda_o,\lambda) \equiv R_E(\lambda_o,\lambda)$ and K is a constant with respect to wavelength.

The direct method for obtaining $r(\lambda_o,\lambda)$ is by irradiating the spectroradiometer, with it set on wavelength λ_o, with a succession of monochromatic beams, each having the same irradiance but a different wavelength. These wavelengths should be distributed over a range large enough to cover the entire relative responsivity function. A plot of the resulting output signal of the spectroradiometer against wavelength is the relative responsivity function $r(\lambda_o,\lambda)$ of the instrument at λ_o. For a spectroradiometer containing a monochromator, when the width of the exit slit is equal to the width of the image of the entrance slit (in the neighborhood of the exit slit), the relative responsivity function, for a particular instrument setting λ_o, is a bell-shaped curve similar to that shown in figure 7.2. The width of the relative responsivity function is called the spectral slit width which is defined as either (1) the width at half-height or (2) the ratio $\int_{-\infty}^{\infty} r(\lambda_o,\lambda)d\lambda / r(\lambda_o,\lambda_o)$. The width of the central portion of the responsivity function is determined primarily by the linear dispersion and the width of the exit slit and the entrance-slit image. The value of the relative responsivity function far from the central region is primarily due to internal radiation scattering. Diffraction and optical aberrations also contribute to the shape of the responsivity function. If scattering, diffraction and aberration effects are negligible compared to the effects of the slits; and if the instrument dispersion, detector spectral responsivity and spectral propagance change negligibly over the central (wavelength) portion of the relative responsivity function, this function has approximately a triangular or trapezoidal shape depending on whether or not the slit widths mentioned above are equal. The width of the base $\Delta\lambda_B$ of such an ideal relative responsivity function is

$$\Delta\lambda_B = \left(\Delta x_{\substack{\text{image of}\\ \text{ent. slit}}} + \Delta x_{\substack{\text{exit}\\ \text{slit}}}\right) \Big/ \left(\frac{dx}{d\lambda}\right) \ [\text{nm}] \tag{7.8}$$

where $dx/d\lambda$ is the linear dispersion of the instrument and the Δx's are the slit widths referred to above. The width of the top portion $\Delta\lambda_T$ of the trapezoidal (unequal slit widths) responsivity function is

$$\Delta\lambda_T = \left|\Delta x_{\substack{\text{image of}\\ \text{ent. slit}}} - \Delta x_{\substack{\text{exit}\\ \text{slit}}}\right| \Big/ \left(\frac{dx}{d\lambda}\right) \ [\text{nm}]. \tag{7.9}$$

Since suitable monochromatic beams for which the wavelength is tunable have not been readily available, most measurements of the relative spectral responsivity function have been made by using an indirect approximate method rather than the above described direct method. In this indirect method, a fixed wavelength monochromatic beam is spectrally scanned by the spectroradiometer. The spectral plot of this output signal is the mirror image of the plot of a major factor of the relative responsivity function, a factor that we call the slit-scattering function. The remaining factor, which we call the responsivity factor, can be obtained by spectrally scanning a continuous source standard of spectral irradiance.

The indirect method for determining $r(\lambda_o,\lambda)$ is accurate only when the various factors affecting the shape of the slit-scattering function: dispersion, scattering, diffraction and optical aberrations, change negligibly over the entire wavelength range of the function. This condition is equivalent to the slit-scattering function being a function only of the difference between the wavelength setting of the monochromator λ_o and the wavelength of the flux. Then

$$r(\lambda_o,\lambda) = z(\lambda_o-\lambda)\cdot r^f(\lambda) \text{ [dimensionless]}, \tag{7.12}$$

where $z(\lambda_o-\lambda)$ is the slit-scattering function and $r^f(\lambda)$ is the responsivity factor referred to earlier. The $\lambda_o-\lambda$ dependence in eq. (7.12) is highly accurate only for small values of $|\lambda_o-\lambda|$, approximately equal to the spectral slit width or less. For much larger values of $|\lambda_o-\lambda|$, the relative responsivity function measured by the indirect method may be in error by 100% or more.

The slit-scattering function in the neighborhood of λ_o has a spectral shape similar to that of the responsivity function because, there, the responsivity factor generally changes much more slowly with wavelength than the slit-scattering function. The distant wings of the slit-scattering function typically have values relative to their peak of 10^{-4} and 10^{-8} for a single and double monochromator, respectively. Examples of such slit functions as well as a typical example of a responsivity factor are shown in figures 7.13 and 7.12.

The distortion produced by the relative responsivity function in spectroradiometric measurements is illustrated (figure 7.14) for a few characteristic spectral distributions. These include the familiar "out-of-band" spectral scattering so common in UV spectroradiometry and the reduction of spectral curvature of convex- and concave-shaped spectral distributions.

The responsivity and slit-scattering function material developed in this chapter for spectral irradiance measurements can also be applied to spectral radiance measurements provided either the spectral radiance or the responsivity function is uniform and isotropic throughout the measurement beam selected by the spectroradiometer. This can always be realized by making this beam sufficiently small, but then many more measurements are required in order to cover the original sized beam.

Since information about the relative responsivity function is required for radiometric measurements of spectral lines, an introduction to this subject is included. It is shown that the irradiance produced by an isolated spectral line of width small compared to the spectral slit width and without continuum radiation is given by

$$E(\lambda_{line}) \equiv \int_{\Delta\lambda_{line}} E_\lambda(\lambda)\cdot d\lambda \simeq \frac{S(\lambda_o)}{S^s(\lambda_o)} \cdot E^s_\lambda(\lambda_o)\cdot \frac{\int_{\Delta\lambda} z(\lambda_o-\lambda)d\lambda}{z(\lambda_o-\lambda_o)} \ [W\cdot m^{-2}] \qquad (7.32a)$$

where $S(\lambda_o)$ and $S^s(\lambda_o)$ are the output signals, respectively, of the spectroradiometer when measuring a spectral line source and a spectral standard continuum source at a wavelength setting λ_o corresponding to the wavelength of the spectral line. $E^s_\lambda(\lambda_o)$ is the spectral irradiance at λ_o of the standard, $z(\lambda_o-\lambda)$ is the slit-scattering function of the instrument at a wavelength setting λ_o, and $z(\lambda_o-\lambda_o)$ is the peak value of the slit-scattering function.

When the spectral width of the line is not small compared to the spectral slit width or when there are two or more lines separated by a wavelength interval roughly equal to the spectral slit width, the irradiance of the line (or lines) is given approximately by

$$E(_{\text{of all lines}}^{\text{Total}}) \simeq \frac{\int_{\Delta\lambda_o} S(\Delta\lambda,\lambda_o)\, d\lambda_o}{S^s(\Delta\lambda,\lambda_{mean})} \cdot E^s_\lambda(\lambda_{mean}) \ [W\cdot m^{-2}] \qquad (7.38a)$$

where the output signal $S(\Delta\lambda,\lambda_o)$ must now be integrated over a sufficently large wavelength range $\Delta\lambda_o$ so as to include all significant contributions to the signal from the spectral lines. The wavelength λ_{mean} at which the signal and spectral irradiance from the standard is obtained is at the center of the wavelength interval $\Delta\lambda_o$ and it is again assumed in eq. (7.38a) that no continuum radiation is present.

References

[7.1] Henry J. Kostkowski and Fred E. Nicodemus, "An Introduction to the Measurement Equation", Chapter 5 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech Note 910-2, 118 pages (Feb. 1978) 58-92.

[7.2] Louis Sica, "Use of Laser Radiation in the Calibration of Spectrometer Scanning Functions", J. Opt. Soc. Am. 57, No. 11 (Nov. 1967) 1366-1372.

[7.3] R. D. Saunders and J. B. Shumaker, "Optical Radiation Measurements: The 1973 NBS Scale of Spectral Irradiance", Nat. Bur. Stand. (U.S.), Tech Note 594-13, 36 pages (Apr. 1977).

Chapter 8. Deconvolution

by John B. Shumaker

In this CHAPTER. We discuss deconvolution, the numerical process of recovering an improved spectral distribution from spectroradiometric measurements which inevitably are smeared spectrally by the spectral-responsivity function of the radiometer. We first present a simple iterative technique which is extensively used and will be completely satisfactory in almost all radiometric situations. We also present a more sophisticated technique which should succeed in many of the instances where the simpler technique fails and which is founded on a sounder theoretical basis. Examples of computer programs for the two techniques are given in an appendix. We apply both techniques to a simple numerical example and briefly mention some of the sources of difficulty and the limitations of deconvolution. Finally, we touch on the question of errors and describe a couple of useful measures of accuracy.

INTRODUCTION. Deconvolution[1] refers to the problem of extracting spectral radiometric quantities such as spectral irradiance from measurements in which the spectral shape of the measured quantity has been distorted by the effects of a responsivity function. This is an example of a general problem which arises in any measurement situation where the resolution of the measuring instrument is less than perfect. Other examples in radiometry include measurements of the angular, temporal, or spatial (area) dependence of radiance using any real radiometer possessed of non-infinitesimal resolvabilities in these dimensions. In the case of spectral resolution we have {see eq. (5.32) [8.1]}[2]

$$S(\lambda_o) = \int R_E(\lambda_o,\lambda)\cdot E_\lambda(\lambda)\cdot d\lambda \quad [S], \tag{8.1}$$

where $E_\lambda(\lambda)$ $[W\cdot m^{-2}\cdot nm^{-1}]$ is the spectral irradiance whose value is desired. (Similar relationships, of course, exist for the other spectral radiometric quantities.) $R_E(\lambda_o,\lambda)$ $[S\cdot W^{-1}\cdot m^2]$ is the spectral irradiance responsivity {eq. (5.49) [8.1]} of the radiometer as a function of the wavelength λ of the radiation entering the instrument and the (mechanical) setting λ_o of the instrument itself. In principle, as we have noted in Chapters 5 [8.1] and 7, $R_E(\lambda_o,\lambda)$ can be measured directly with monochromatic beams.

---

[1]Strictly speaking, the terms "convolution" and "deconvolution" apply to integrals in which the kernel -- our responsivity function -- depends only upon the difference between its two arguments. However, there seems to be no good reason to use a different name to cover the usual radiometric case in which $R_E(\lambda_o,\lambda)$, in addition to depending mainly upon the difference $\lambda_o-\lambda$, may also depend on where on the wavelength scale this difference occurs, i.e., on λ or λ_o, and can be handled by a trivial generalization of the more restricted case.

[2]Figures in square brackets indicate literature references listed at the end of *this chapter*.

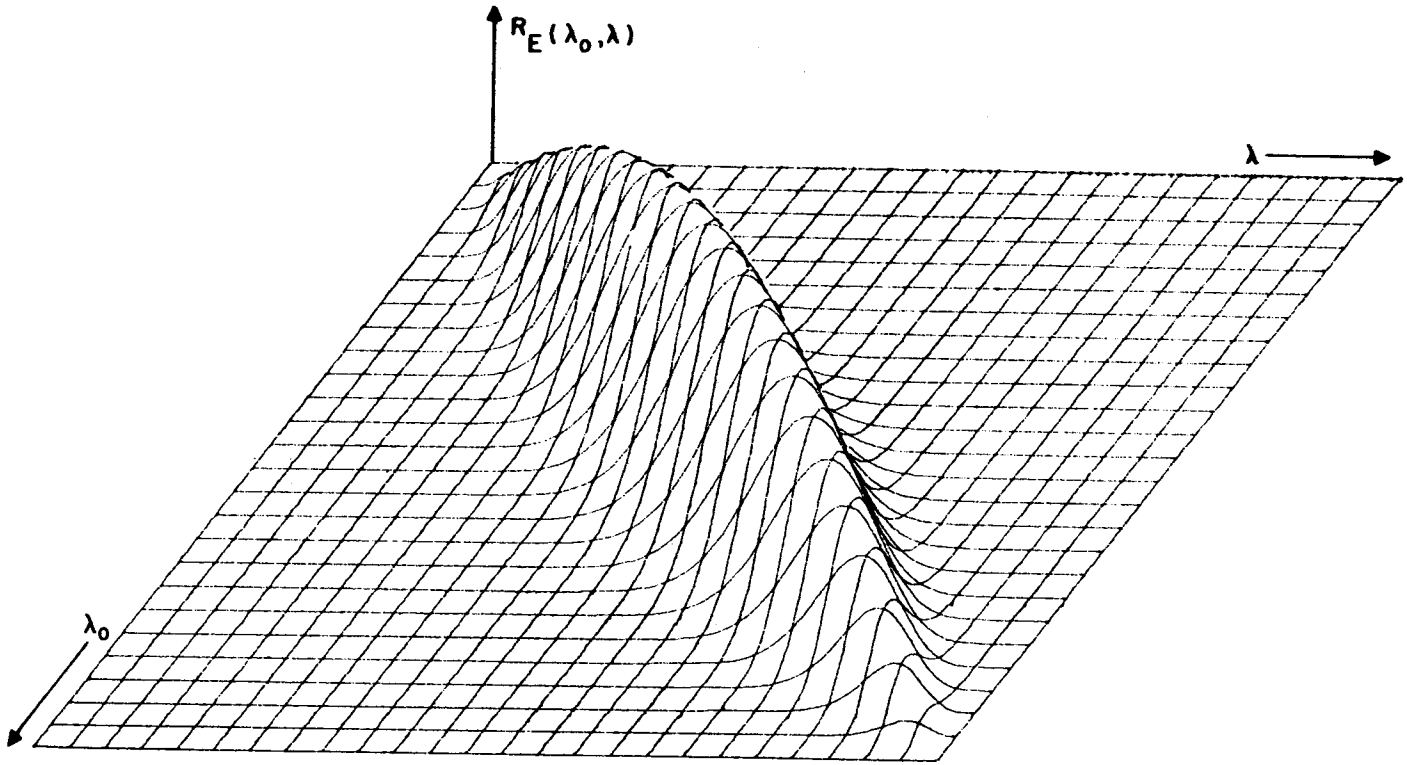


Figure 8.1. A three-dimensional representation of the spectral responsivity function.

Figure 8.1 is an attempt at a 3-dimensional sketch of the surface $R_E(\lambda_o,\lambda)$ as a function of λ_o and λ. $S(\lambda_o)$ is the output signal in volts or other appropriate units when the instrument is set at λ_o and views the source of spectral irradiance $E_\lambda(\lambda)$. The integration is carried out over all λ [nm] for which neither $R_E(\lambda_o,\lambda)$ nor $E_\lambda(\lambda)$ is zero. Thus, $S(\lambda_o)$ and $R_E(\lambda_o,\lambda)$ are the quantities which can be directly measured and the problem is to calculate the desired quantity $E_\lambda(\lambda)$ from the integral equation, eq. (8.1).

If the radiometer had perfect spectral resolution we could write[1]

$$R_E(\lambda_o,\lambda) = R_E^f(\lambda)\cdot\delta(\lambda_o-\lambda) \quad [S\cdot W^{-1}\cdot m^2] \tag{8.2}$$

and eq. (8.1) would become

$$S(\lambda_o) = \int R_E^f(\lambda)\cdot\delta(\lambda_o-\lambda)\cdot E_\lambda(\lambda)\cdot d\lambda = R_E^f(\lambda_o)\cdot E_\lambda(\lambda_o) \quad [S]. \tag{8.3}$$

In this case $E_\lambda(\lambda)$ can be computed directly from the measurements by division by a simple spectral irradiance responsivity factor $R_E^f(\lambda)$. The assumption of eq. (8.2) leading to eq. (8.3) is equivalent to ignoring the effects of the responsivity function. With less-than-perfect spectral resolution, $R_E(\lambda_o,\lambda)$ can no longer be represented by a delta function and both the measurement of the responsivity function $R_E(\lambda_o,\lambda)$ and the calculation of $E_\lambda(\lambda)$ become much more difficult.

Before going on, we should stop for a moment to note the distinction between this situation and the similar one of a measurement with a broad-band (broad-spectral-band) radiometer covered in the discussion following eq. (5.32) in Chapter 5 [8.1]. Equation (5.32) is quite similar to eq. (8.1) [or eq. (5.13)], the important difference being that in eq. (8.1) the responsivity is not only a function of wavelength λ, but also of a monochromator wavelength setting λ_o. For any one value of λ_o, i.e., for just one measurement at a single monochromator setting, the solution to eq. (8.1) or (5.13) is just as indeterminate as that to eq. (5.32) and everything said in the discussion of eq. (5.32) is equally applicable to the monochromator situation. However, with the monochromator, the incident radiation at each value of λ affects the output at a number of different values of λ_o. In other words, the outputs at several values of λ_o all contain information about the input at a given value of λ (along with information about inputs at other values of λ, as well). It is this additional overlapping information which enables us to say something about the dependence of $E_\lambda(\lambda)$ upon λ.

---

[1] $\delta(x)$ is the Dirac delta function with the properties:

$$\begin{aligned} &\delta(x) = 0, \quad \text{for} \quad x \neq 0, \quad \text{and} \\ &\int_{-a}^{a} \delta(x)\cdot dx = 1, \quad \text{for any} \quad a > 0. \end{aligned} \tag{8.4}$$

A SIMPLE ITERATIVE SOLUTION. Although many techniques[1] have been devised for the solution of the integral equation, eq. (8.1), the only one which has been extensively applied in optical spectrometry is based upon an iterative method for the solution of simultaneous linear equations [8.2 - 8.10]. If we assume that measurements $S(\lambda_o)$ are made at n equally spaced wavelength settings λ_k and if we approximate the integral in eq. (8.1) by a summation we can rewrite eq. (8.1) as

$$S_k = \Delta\lambda \cdot \sum_{i=1}^{n} R_{ki} \cdot E_i \quad [S], \qquad k = 1,2,\ldots,n. \tag{8.5}$$

In this equation $\Delta\lambda$ [nm] is the wavelength spacing, S_k [S] is the spectroradiometer output $S(\lambda_k)$ when set at the k^{th} wavelength, E_i [$W \cdot m^{-2} \cdot nm^{-1}$] is the desired spectral irradiance $E_\lambda(\lambda_i)$ (the subscript λ is omitted for simplicity) at the i^{th} wavelength and R_{ki} [$S \cdot W^{-1} \cdot m^2$] is the instrument spectral irradiance responsivity $R_E(\lambda_k,\lambda_i)$ (the spectral output per unit incident spectral irradiance) to monochromatic radiation at the i^{th} wavelength when the instrument is set to the k^{th} wavelength. In figure 8.1 the value of any R_{ki} is just the height of the surface, $R_E(\lambda_o,\lambda)$, at the point of intersection of the line drawn on the surface at constant $\lambda_o = \lambda_k$ with the line drawn on the surface at constant $\lambda = \lambda_i$.

Equations (8.5) consist of n equations in n unknowns where n in typical applications may be as small as 50 or as large as 10^5. Because of the size of n and because, as we shall see later, a mathematically exact solution of eqs. (8.5) is not needed, an iterative approximation technique is used for solving this set of equations. In this procedure[2] an assumed solution $E_i^{(0)}$ for the E_i together with the R_{ki} yields a computed set of $S_k^{(0)}$ from eqs. (8.5). Based upon a comparison of these computed values with the measured values of S_k the E_i are corrected and the process is repeated. The detailed steps are as follows: First we assume trial spectral irradiance values $E_i^{(0)}$ which we obtain from

$$E_i^{(0)} = S_i/(\Delta\lambda \cdot \sum_j R_{ij}) \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad i = 1,2,\ldots,n. \tag{8.6}$$

The value of $E_i^{(0)}$ is just the value which would have been obtained for E_i if deconvolution were not attempted [i.e., by solving eq. (7.2)]. Next, using eqs. (8.5), we calculate what effect the responsivity function would have on this spectrum:

---

[1]Hundreds of papers have been written on deconvolution and related subjects. The references cited in this chapter give only a sampling of this literature. Further references are given in [8.11,8.12,8.13]. Some interesting techniques which are not even hinted at in this chapter are described in [8.13-8.16].

[2]An illustrative computer program for this procedure is given in Appendix 8A.

$$S_k^{(0)} = \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i^{(0)} \quad [S], \qquad k = 1,2,\ldots,n. \tag{8.7}$$

These values of $S_k^{(0)}$ will in general not agree with the measured S_k. Therefore, we correct the $E_i^{(0)}$ using the difference between S_k and $S_k^{(0)}$ so as to improve our estimated values for E_i. Thus, we compute[1]

$$E_k^{(1)} = E_k^{(0)} + A_k \cdot [S_k - S_k^{(0)}] \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad k = 1,2,\ldots,n. \tag{8.8}$$

where A_k is a convergence factor whose value is not critical. Frequently a good value for A_k [as will be derived in eq. (8.81)] is

$$A_k = R_{kk} / (\Delta\lambda \cdot \sum_j R_{kj}^2) \quad [W \cdot m^{-2} \cdot nm^{-1} \cdot S^{-1}], \qquad k = 1,2,\ldots,n. \tag{8.9}$$

We shall return to a further brief discussion of A_k below. Now we continue in the same way using values of $E_i^{(1)}$ as a trial solution to obtain better estimates $E_i^{(2)}$, etc. In general[2]

$$E_k^{(\nu)} = E_k^{(\nu-1)} + A_k \cdot [S_k - S_k^{(\nu-1)}] \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad k = 1,2,\ldots,n, \tag{8.10}$$

with $$S_k^{(\nu-1)} = \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i^{(\nu-1)} \quad [S], \qquad k = 1,2,\ldots,n, \tag{8.11}$$

where the summation over i runs from $i = 1$ to $i = n$.

In many applications two or three iterations (i.e. $\nu = 2$ or 3) will suffice to reduce the correctable errors due to responsivity-function effects to less than the measurement noise. In fact, a criterion based upon consideration of the measurement noise should be used to terminate the iterations: when the computed $S_k^{(\nu)}$, given by eq. (8.11), agree with the measured S_k to within the experimental noise, for example, when [8.10]

---

[1]It would have been nice to be able invariably to associate the subscript k with the instrumental setting λ_o and the subscript i with the wavelength λ, but the varying requirements for these dummy indices in the different equations make any such consistent usage impossible to maintain.

[2]Iterative multiplicative corrections instead of additive corrections have been used [8.13, 8.17]:

$$E_k^{(\nu)} = E_k^{(\nu-1)} \cdot S_k / S_k^{(\nu-1)}. \tag{8.12}$$

Although rarely used it appears that this technique might be advantageous where $S(\lambda_o)$ spans several orders of magnitude of signal level with similar relative accuracy.

$$\sum_{k=1}^{n} \frac{(S_k^{(\nu)} - S_k)^2}{\sigma_k^2} \leq n \quad \text{[dimensionless]}, \tag{8.13}$$

with σ_k [S] being the estimated random experimental uncertainty in S_k, then the iterations should be stopped. Further iterations will begin to have a deleterious effect -- introducing spurious spectral distortions into the E_i which increase with ν and which are due entirely to noise in the S_k and to computational round-off error.

The convergence factors A_k are sometimes increased (or decreased) during the iteration process in an effort to achieve faster convergence. At convergence the factor $[S_k - S_k^{(\nu-1)}]$ in eq. (8.10) will have decreased to zero so that the value of $E_k^{(\nu)}$ reached no longer depends upon A_k. Consequently, the only role played by A_k is in determining the number of iterations required to achieve convergence by setting the magnitude of the correction to be applied in each iteration. Normally, all A_k's would be given the same value based upon a typical evaluation of eq. (8.9) and then, if after 3 or 4 iterations convergence is not close, this value would be modified upward (or downward) by trial-and-error to hasten convergence (or stem divergence). Somewhat similarly the ultimate values of $E_i^{(\nu)}$ achieved at convergence should not depend upon the starting values $E_i^{(0)}$. The initial values given by eq. (8.6) are an obvious starting point but starting, for example, with all $E_i^{(0)} = 0$ will often require only one or two more iterations. If a better set of initial values is available it should, of course, be used.

DECONVOLUTION in MATRIX NOTATION. Before turning to a simple numerical example to illustrate the application of eq. (8.10), let us digress for a moment to rewrite the equations as matrix equations. As we pointed out in Chapter 6 [8.18], no numerical advantage is gained by this; however, matrix equations are extremely useful in organizing calculations, particularly since some computer languages contain explicit matrix operations which can greatly reduce the chore of programming. We write the n values of the observed signal S_k as an n-dimensional vector

$$S = [S_k] = \begin{bmatrix} S_1 \\ S_2 \\ . \\ . \\ . \\ S_n \end{bmatrix} \quad [S] \tag{8.14}$$

and similarly, the n desired values E_k of spectral irradiance

$$E = [E_k] = \begin{bmatrix} E_1 \\ E_2 \\ \cdot \\ \cdot \\ \cdot \\ E_n \end{bmatrix} \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.15}$$

The responsivities R_{ki} form an n by n matrix

$$R = [R_{ki}] = \begin{bmatrix} R_{11} & R_{12} & \cdot & \cdot & \cdot & R_{1n} \\ R_{21} & R_{22} & \cdot & \cdot & \cdot & R_{2n} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ R_{n1} & R_{n2} & \cdot & \cdot & \cdot & R_{nn} \end{bmatrix} \quad [S \cdot W^{-1} \cdot m^2]. \tag{8.16}$$

where the matrix elements, R_{ki}, are just the heights of the intersection points on the $R_E(\lambda_o,\lambda)$ surface as sketched in figure 8.1. Now, using matrix "shorthand", the n simultaneous equations represented by eq. (8.5) are exactly what is meant by the one equation

$$S = \Delta\lambda \cdot R \cdot E \quad [S] \tag{8.17}$$

or

$$\begin{bmatrix} S_1 \\ S_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ S_k \\ \cdot \\ \cdot \\ S_n \end{bmatrix} = \Delta\lambda \cdot \begin{bmatrix} R_{11} & R_{12} & \cdot & \cdot & R_{1i} & \cdot & \cdot & \cdot & \cdot & R_{1n} \\ R_{21} & R_{22} & \cdot & \cdot & R_{2i} & \cdot & \cdot & \cdot & \cdot & R_{2n} \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ R_{k1} & R_{k2} & \cdot & \cdot & R_{ki} & \cdot & \cdot & \cdot & \cdot & R_{kn} \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ \cdot & \cdot & & & \cdot & & & & & \cdot \\ R_{n1} & R_{n2} & \cdot & \cdot & R_{ni} & \cdot & \cdot & \cdot & \cdot & R_{nn} \end{bmatrix} \cdot \begin{bmatrix} E_1 \\ E_2 \\ \cdot \\ \cdot \\ E_i \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ E_n \end{bmatrix} \quad [S]. \tag{8.18}$$

A typical member of this set of equations is[1]

$$S_k = \Delta\lambda \cdot (R_{k1} \cdot E_1 + R_{k2} \cdot E_2 + \ldots + R_{ki} \cdot E_i + \ldots + R_{kn} \cdot E_n). \tag{8.18a}$$

[1]This n-dimensional matrix multiplication is an obvious generalization of the 4-dimensional formulas given in Appendix 6 [8.19].

Thus the signal obtained at the k^{th} wavelength setting of the radiometer is associated with the k^{th} row of the spectral responsivity matrix R. If the radiometer is set at the k^{th} wavelength and an isolated spectral line source of known irradiance at wavelength λ_i is observed (i.e., all $E_\ell = 0$ except E_i) then the observed signal will be just $S_K = \Delta\lambda \cdot R_{ki} \cdot E_i$. If this line source is tunable so that it can be moved from λ_1 to λ_n a complete row of R_{ki} values can be determined. Moving the radiometer wavelength setting to other wavelengths and repeating these measurements will provide other rows of the spectral responsivity matrix. On the other hand, if the spectral line is fixed at the i^{th} wavelength while the radiometer setting is moved from λ_1 to λ_n a complete column $R_{1i}, R_{2i}, \ldots R_{ki}, \ldots R_{ni}$ of R values can be determined.

Going on now to eqs. (8.10) we see that in matrix notation they become

$$E^{(\nu)} = E^{(\nu-1)} + A \cdot \left[S - S^{(\nu-1)}\right] \tag{8.19}$$

with $$S^{(\nu-1)} = \Delta\lambda \cdot R \cdot E^{(\nu-1)} \tag{8.20}$$

and

$$A = \begin{bmatrix} A_1 & 0 & 0 & \cdots \\ 0 & A_2 & 0 & \cdots \\ 0 & 0 & A_3 & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots A_n \end{bmatrix} \quad [W \cdot m^{-2} \cdot nm^{-1} \cdot S^{-1}]. \tag{8.21}$$

In the slit-scattering function approximation we approximate the responsivity function by a product [eq. (7.12); K is defined following eq. (7.5)]

$$R(\lambda_o, \lambda) = K \cdot z(\lambda_o - \lambda) \cdot r^f(\lambda) \quad [S \cdot W^{-1} \cdot m^2]. \tag{8.22}$$

In terms of matrices we can say that the matrix R is proportional to the product of a slit-scattering function matrix

$$Z = \begin{bmatrix} z_0 & z_1 & z_2 & \cdots & z_{n-1} & z_n \\ z_{-1} & z_0 & z_1 & \cdots & z_{n-2} & z_{n-1} \\ z_{-2} & z_{-1} & z_0 & \cdots & z_{n-3} & z_{n-2} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ z_{1-n} & z_{2-n} & z_{3-n} & \cdots & z_0 & z_1 \\ z_{-n} & z_{1-n} & z_{2-n} & \cdots & z_{-1} & z_0 \end{bmatrix} \quad [\text{dimensionless}] \tag{8.23}$$

and a responsivity-factor matrix

$$R^f = \begin{bmatrix} r_1^f & 0 & \cdots & 0 \\ 0 & r_2^f & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & r_n^f \end{bmatrix} \quad [\text{dimensionless}]. \tag{8.24}$$

So eq. (8.5) becomes

$$S = \Delta\lambda \cdot K \cdot Z \cdot R^f \cdot E \quad [S]. \tag{8.25}$$

Usually it is convenient to perform the final multiplication and define a quantity

$$E' = R^f \cdot E = \begin{bmatrix} r_1^f \cdot E_1 \\ r_2^f \cdot E_2 \\ \cdot \\ \cdot \\ r_n^f \cdot E_n \end{bmatrix} \quad [W \cdot m^{-2} \cdot nm^{-1}] \tag{8.26}$$

so that, in terms of this quantity eq. (8.5) becomes

$$S = \Delta\lambda \cdot K \cdot Z \cdot E' \quad [S], \tag{8.27}$$

which is identical in form to eq. (8.17) and can be solved by the iterative technique we are discussing. If the responsivity factors r_i^f are known, this then permits the E_i to be calculated from the E_i'. If the r_i^f are not known it is necessary to repeat the measurements and the solution of eq. (8.27) for a standard of known spectral irradiance. Then the r_i^f can be calculated or they may be eliminated by using the ratios $E_i'/E_i'(s)$. That is,

$$E_i = \frac{E_i'}{E_i'(s)} \cdot E_i(s) \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.28}$$

[The argument (s), here is used to denote values associated with the standard of known spectral irradiance.] Forming the ratios $S_k/S_k(s)$ before deconvolution in order to avoid deconvolution of the standard source data will *not* give the same results unless the relative spectral distributions of both sources are identical. Returning for a moment to the slit-scattering function matrix of eq. (8.23), notice that a column of this matrix (associated with a radiometer wavelength scan of a source of a fixed wavelength spectral line) contains the same elements as a row (associated with viewing a spectral line of changing wavelength at a fixed radiometer setting) but in reverse order. This, of course, is simply another way of looking at the apparent slit-scattering function reversal with fixed-wavelength spectral

lines which has already been mentioned in Chapter 7. The great advantage of the slit-scattering function approximation when it is appropriate is that at most only $2n + 1$ values must be stored in the computer to specify the Z matrix; any element of the array is identical to one of these. For large n or small computers this is a vital consideration.

A NUMERICAL EXAMPLE. In this section we wish to illustrate the iterative deconvolution described by eqs. (8.10) with a simple numerical example. For our example we assume a "true" spectral-irradiance distribution and an instrument-responsivity function as shown in figures 8.2a and 8.2b. The responsivity function depends only upon the difference $\lambda_o - \lambda$. The product of the simple, piecewise, algebraic expressions which define $R(\lambda_o,\lambda)$ and $E_\lambda(\lambda)$ can be integrated to give exact values of the "observed" signal {eq. (8.1)}:

$$S(\lambda_o) = \int R(\lambda_o,\lambda)\cdot E_\lambda(\lambda)\cdot d\lambda \quad [S] \tag{8.29}$$

at any value of λ_o. Such values are shown as the second column of table 8-1 for wavelength intervals of 2 ($\Delta\lambda = 2$). Assuming that measurements with our radiometer had yielded these values of S_k we wish to obtain values of E_i using the technique of eqs. (8.10). If all goes well these should approximate values of $E_\lambda(\lambda)$ given in figure 8.2a. First we need a table of responsivity values at wavelength intervals of 2. These can be seen by inspection of figure 8.2b to be non-zero only for $\lambda = \lambda_o$ and $\lambda = \lambda_o \pm \Delta\lambda$. Thus we have $R_{k,k-1} = 2$, $R_{kk} = 4$, and $R_{k,k+1} = 2$ for any k and all other $R_{ki} = 0$. If we arrange the R_{ki} in a matrix it will look like

$$R = [R_{ki}] = \begin{bmatrix} 4 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & \cdots & 0 & 0 & 0 \\ & & \vdots & & & & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 2 & 4 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 & 4 \end{bmatrix} \quad [S\cdot W^{-1}\cdot m^2]. \tag{8.30}$$

We choose $A_k = 1/12$. This is the value given by eq. (8.9) for all k except $k = 0$ and $k = n$ of the R array of eq. (8.30). Since the values of A_k only affect the rate of convergence of the iterations and not the final results, it is usual to choose A_k to be constant, independent of k as we have done. The zeroth approximation to the values of E_i according to eq. (8.6) is simply the corresponding value of S_i multiplied by 1/16. Again we ignore possible errors in this factor at the ends of the wavelength range. These values of $E_i^{(0)}$ are given in column 3 of table 8-1. They are exactly the values of spectral irradiance which would be obtained if the responsivity-function correction were ignored. Now we proceed to the calculation of the first iterative correction according to eqs. (8.10). At each instrumental setting, k, we first calculate the sum $\sum_i R_{ki}\cdot E_i^{(0)}$ and then the

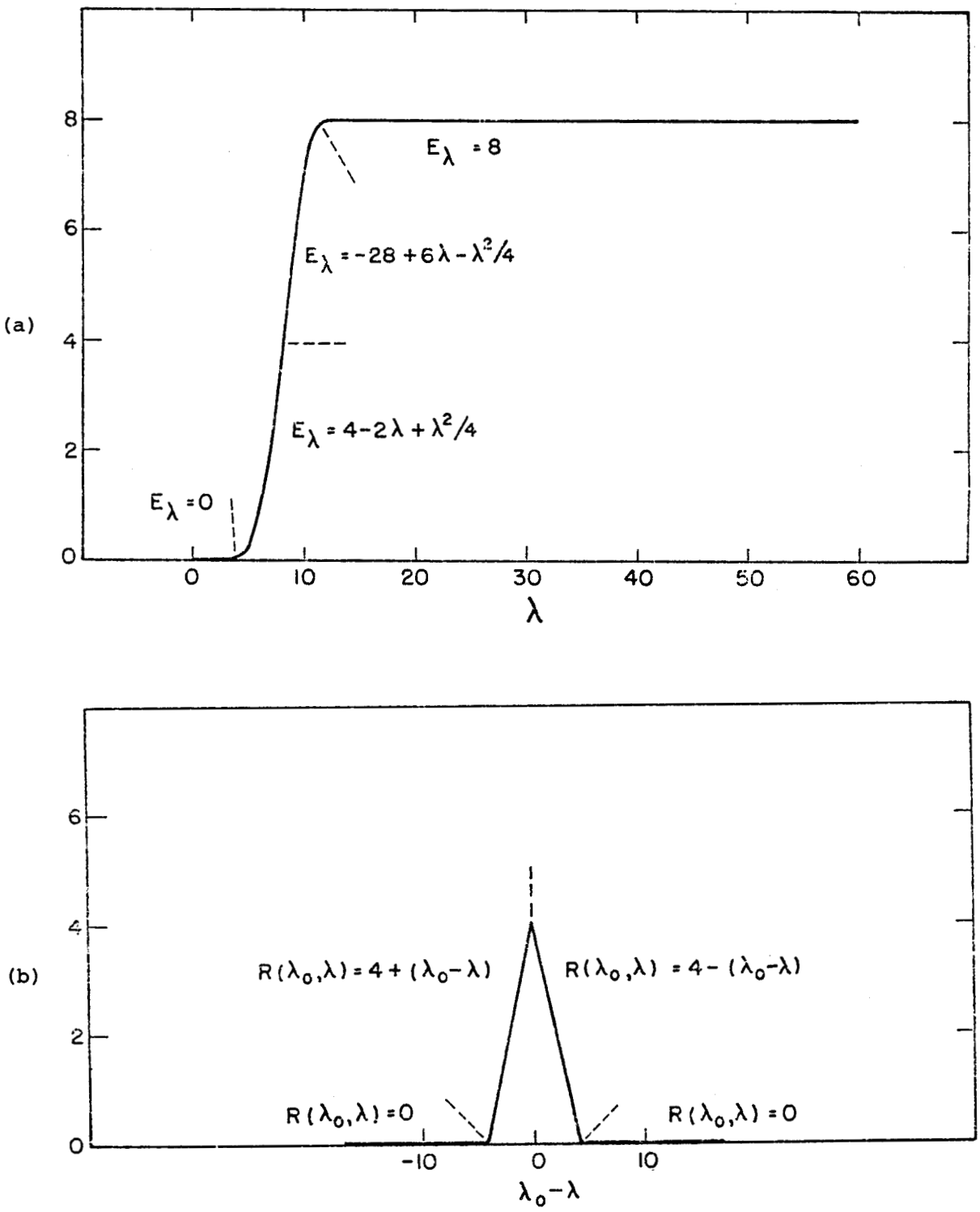


Figure 8.2. Spectral irradiance distribution (a), and spectral responsivity function (b), assumed for numerical examples in tables 8-1, 8-1a, 8-2, and 8-3.

Table 8-1. Simple iterative deconvolution

λ	S	$E^{(0)}$	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$	$E^{(7)}$	$E^{(8)}$	$E^{(9)}$	E	$S^{(9)}$
0	0.00	.00	-.01	.02	.01	.00	-.00	-.00	-.00	-.00	-.00	.00	-.00
2	0.33	.02	-.08	.00	-.00	.00	.00	.01	.00	.00	.00	.00	.34
4	5.33	.33	.01	.06	.06	.07	.07	.08	.08	.08	.08	.00	5.32
6	25.67	1.60	1.23	1.21	1.19	1.18	1.18	1.17	1.17	1.17	1.17	1.00	25.68
8	64.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	64.00
10	102.33	6.40	6.77	6.79	6.81	6.82	6.82	6.83	6.83	6.83	6.83	7.00	102.32
12	122.67	7.67	7.99	7.94	7.94	7.93	7.93	7.92	7.92	7.92	7.92	8.00	122.68
14	127.67	7.98	8.08	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	127.66
16	128.00	8.00	8.01	7.98	7.99	7.99	8.00	8.00	8.00	8.00	8.00	8.00	128.00
18	128.00	8.00	8.00	8.00	8.01	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.01
20	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
22	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
24	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
26	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
28	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
30	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
32	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
34	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
36	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
38	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
40	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
42	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
44	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
46	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.04
48	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.01	8.03	8.00	127.83
50	128.00	8.00	8.00	8.00	8.00	8.00	8.00	7.99	7.97	7.93	7.89	8.00	128.51
52	128.00	8.00	8.00	8.00	8.00	8.00	8.03	8.09	8.16	8.25	8.34	8.00	126.88
54	128.00	8.00	8.00	8.00	8.00	7.90	7.77	7.62	7.45	7.29	7.13	8.00	129.91
56	128.00	8.00	8.00	8.00	8.30	8.59	8.89	9.16	9.42	9.65	9.86	8.00	125.47
58	128.00	8.00	8.00	7.11	6.52	6.02	5.63	5.30	5.02	4.78	4.58	8.00	130.50
60	128.00	8.00	10.67	11.56	12.15	12.54	12.84	13.07	13.26	13.41	13.54	8.00	126.42
			2.13	.42	.12	.05	.03	.02	.01	.01	.01		

correction $A_k \cdot [S_k - \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i^{(0)}]$. For $k = 3$, for example, (at $\lambda = 4$) we have

$\sum_i R_{ki} \cdot E_i^{(0)} = \ldots + 0.0 + 2 \cdot 0.02 + 4 \cdot 0.33 + 2 \cdot 1.60 + 0.0 \ldots = 4.56$ and $S_k = 5.33$.
So the correction is $(5.33 - 2 \cdot 4.56)/12 = -0.32$. Thus, we obtain $E_3^{(1)} = 0.33 - 0.32 = 0.01$ in agreement with the third entry in the fourth column of table 8-1. In a similar fashion the other entries in the $E^{(1)}$ column are calculated. Then new corrections are calculated and the second approximation $E_i^{(2)}$ obtained. That is, still for $k = 3$,

$$\sum_i R_{ki} \cdot E_i^{(1)} = 2 \cdot (-0.08) + 4 \cdot 0.01 + 2 \cdot 1.23 = 2.34, \quad (5.33 - 2 \cdot 2.34)/12 = 0.05, \tag{8.31}$$

$E_3^{(2)} = 0.01 + 0.05 = 0.06$. Table 8-1 shows successive approximations $E_k^{(\nu)}$ through $\nu = 9$. The next-to-last column is the "true" spectral irradiance taken from the expression in figure 8.2a and should be compared to the various approximations, $E^{(\nu)}$. The final column shows how well the "observed" signal is reconstructed from the final iteration $E^{(9)}$ and the responsivity array R_{ki} using eqs. (8.11). The discrepancies between S_k and $S_k^{(9)}$ are quite small in the upper part of the table and since the correction to be applied to $E^{(\nu-1)}$ is proportional to the difference $S - S^{(\nu-1)}$ [eq. (8.10)] this means that convergence is essentially complete in this part of the table.

END-EFFECTS. In the lower part of table 8-1 it is clear that there is a problem which is becoming worse as the number of iterations increases. This is caused by our having arbitrarily terminated the table of S_k values at $\lambda_k = 60$ before S_k had returned to zero, and by the fact that the computational algorithm, in effect, proceeds as though all E_i for $\lambda_i > 60$ (and for $\lambda_i < 0$) are zero. If the data had, in fact, exhibited a step from $S = 128$ to $S = 0$ at $\lambda = 60$ the solution being generated would be correct. In table 8-1 the iterative solution is clearly valid up to within 2 or 3 spectral slit widths of the end of the data, although with significantly more iterations this will no longer be true. To understand why the iteration algorithm functions as though all values of E_i outside the measured range are zero consider the last equation implied by eq. (8.17) [see eq. (8.18a)]:

$$S_n = \Delta\lambda \cdot (\ldots + R_{nn} \cdot E_n). \tag{8.32}$$

If we had made one more measurement we would be writing this equation

$$S_n = \Delta\lambda \cdot (\ldots + R_{nn} \cdot E_n + R_{n,n+1} \cdot E_{n+1}). \tag{8.33}$$

[There would, of course, now be one more equation:

$$S_{n+1} = \Delta\lambda \cdot (\ldots + R_{n+1,n} \cdot E_n + R_{n+1,n+1} \cdot E_{n+1}) \tag{8.34}$$

but we aren't interested in it at the moment.] Since S_n and the R_{ki} and E_i are all fixed numbers, obviously both eqs. (8.32) and (8.33) can't be correct unless $E_{n+1} = 0$.

Thus, if we use eq. (8.30) for the matrix R we imply that all $E_i = 0$ for $i < 1$ and for $i > n$. The best way to avoid this problem is by extending the wavelength range of the measurements or by the use of spectral filters to assure that the instrument output will be negligible outside the measurement range. If this is impractical then it may be possible to extend the table of S_k values by measurement or estimation far enough that the end-effect after deconvolution will not affect the values of E_i in the wavelength region of interest. In the example of table 8-1 the wavelength span in which end-effects can be seen is given approximately by

$$\Delta\lambda_w \cdot \sqrt{\nu}, \tag{8.35}$$

where $\Delta\lambda_w$ is the spectral slit width (eq. 7.34). So, for ν iterations, measurements or estimates of S_k should be extended this far beyond the spectral region of interest on either end if the signal is not negligible there. In the present example we might assume that the wavelength region of interest is from $\lambda = 0$ to $\lambda = 46$ and that the S_k values from $\lambda = 48$ to $\lambda = 60$ were included only to avoid the end-effect. Since the data run smoothly to zero at $\lambda = 0$ there is no end-effect problem there.

The last row of table 8-1 gives the rms error in reconstructing S from the corresponding column of $E_i^{(\nu)}$ values, i.e., the quantity

$$\left\{\frac{1}{16}\sum_{i=1}^{16}(S_i - S_i^{(\nu)})^2\right\}^{\frac{1}{2}} \tag{8.36}$$

for each iteration. We have omitted from the sum the values in the lower half of the table ($\lambda > 30$) in order to exclude the end-effects region. If the measurements S are assumed to be accurate to within 0.1% of their maximum then three iterations ($E^{(3)}$) are evidently sufficient to achieve convergence to within the accuracy of the data and the iterations should be terminated there. Of course, the rms value of the error $E_i - E_i^{(\nu)}$ would be a more valuable statistic but this is never available in a real measurement situation since $E_\lambda(\lambda)$ would be unknown. In the next paragraph and near the end of this chapter we briefly discuss errors in the deconvoluted results, $E^{(\nu)}$.

ACCURACY. In spite of the fact that convergence is essentially complete after a few iterations in the upper part of table 8-1 there still remains a significant discrepancy between $E_k^{(9)}$ and the "true" E_k, especially at $\lambda = 4$, 6, 10, and 12. These errors are attributable primarily to the error in approximating the exact integral of eq. (8-1) by the summation of eq. (8.5). Equation (8.5) is exact if the integrand, $R(\lambda_o,\lambda)\cdot E(\lambda)$, is linear between data points. In figure 8.3 we show this product as a function of λ for $\lambda_o = 6$. Clearly, in any one segment, say from $\lambda = 4$ to $\lambda = 6$, the integrand is far from linear; however, the errors are positive in some parts of the range of integration and negative in others so that they tend to cancel. For this reason more elaborate integration formulas improve the accuracy very little. The certain way to improve the accuracy in the

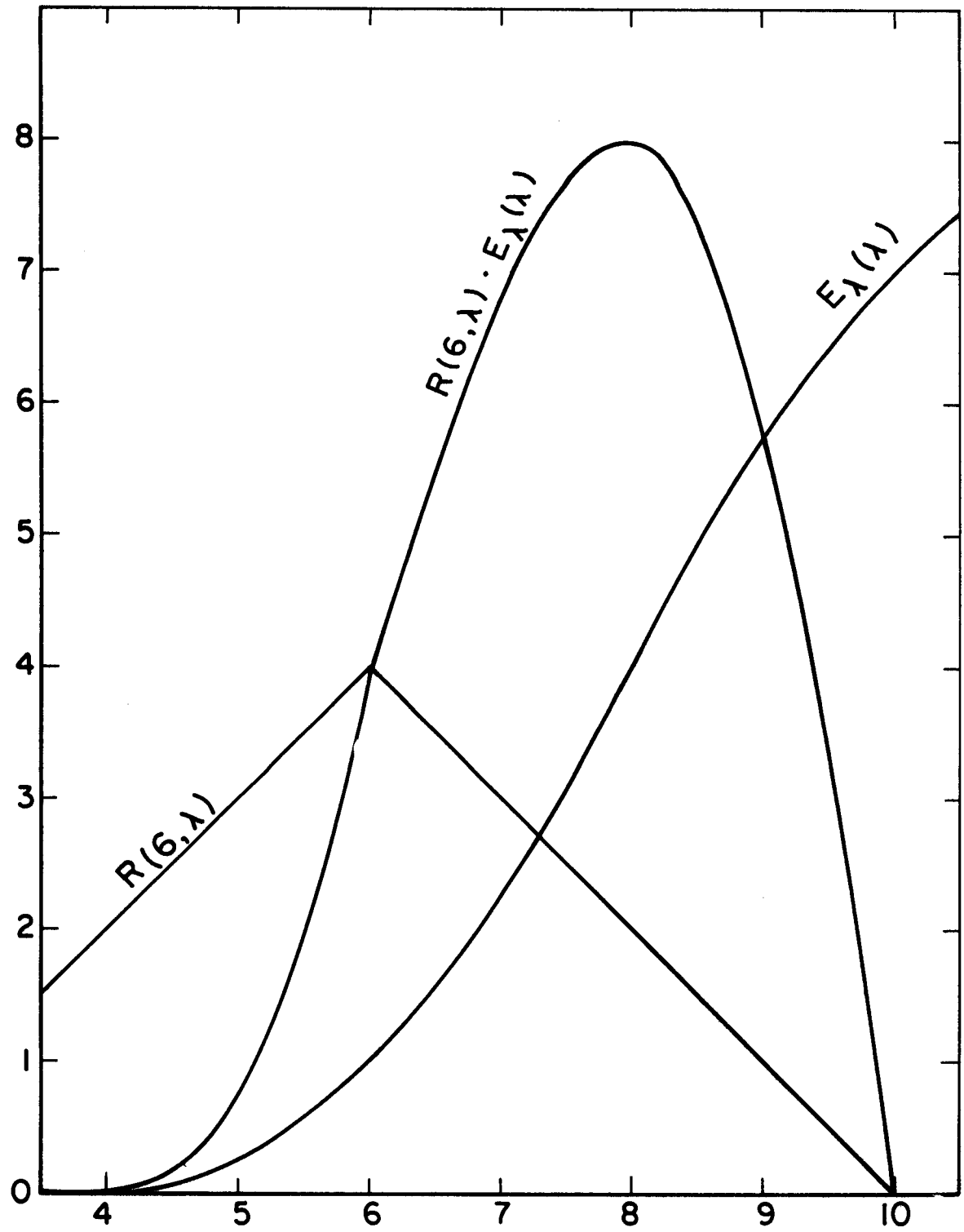


Figure 8.3. Responsivity function, spectral irradiance, and their product for $\lambda_o = 6$. From example of figure 8.2.

neighborhood of rapid changes in $S(\lambda_o)$ is to take data with a smaller $\Delta\lambda$, i.e., to take more data. Of course, even with a perfect mathematical solution there may still be errors if there exist measurement errors or if eq. (8.1) is not obeyed due, for example, to a nonlinear detector. The result of taking more data in our example is illustrated in table 8-1a in which $\Delta\lambda$ was taken as $\frac{1}{2}$. The accuracy of deconvolution after a few iterations is dramatically improved over the results given in table 8-1. The rate of convergence is essentially the same. Table 8-1a shows only the portion of the calculations corresponding to the upper portion of table 8-1. In the end-effects region we would see that the amplitude of the oscillations is much greater with the smaller sampling interval but the range of wavelengths over which the calculations are perturbed by the end-effect is the same.

The INVERSE SLIT-SCATTERING FUNCTION. When the slit-scattering function approximation, eq. (8.22), is appropriate a technique is available which saves considerable computational effort if two or more spectra are to be deconvoluted using the same slit-scattering function [8.20]. Let us write eqs. (8.5) in terms of the slit-scattering function

$$S_k = \Delta\lambda \cdot K \cdot \sum_i z_{k-i} \cdot E'_i \quad [S], \qquad k = 1,2,\ldots,n. \tag{8.37}$$

As in eq. (8.23) we have used $z_{k-i} \equiv z(\lambda_k - \lambda_i)$ [dimensionless] where λ_k [nm] is the instrumental setting and, as in eq. (8.27), we have combined E_i with the responsivity factor $r_i^f \equiv r^f(\lambda_i)$ [dimensionless] to give $E'_i \equiv r_i^f \cdot E_i$ [$W \cdot m^{-2} \cdot nm^{-1}$]. The summation is over all values of E'_i and we take the slit-scattering function z_{k-i} to be zero any time $k-i$ takes on a value outside the range for which the slit-scattering function has been tabulated. Let us also assume that n is an odd number and redefine i and k so that numbering begins at the mid-points of E and S. Thus if $n_2 = (n-1)/2$ we rewrite eq. (8.37)

$$S_k = \Delta\lambda \cdot K \cdot \sum_{i=-n_2}^{n_2} z_{k-i} \cdot E'_i \quad [S], \qquad k = -n_2, -n_2+1, \ldots, -1, 0, 1, \ldots, n_2. \tag{8.38}$$

Now we multiply both sides of this equation by a quantity $v_{\ell-k}$ (yet to be determined) and sum over k

$$\sum_{k=-n_2}^{n_2} v_{\ell-k} \cdot S_k = \Delta\lambda \cdot K \cdot \sum_{i=-n_2}^{n_2} E'_i \left(\sum_{k=-n_2}^{n_2} z_{k-i} \cdot v_{\ell-k} \right). \tag{8.39}$$

We see that if the values of $v_{\ell-k}$ could be chosen in such a way that[1]

---

[1] δ_{ik} in eq. (8.40) is the Kronecker delta function. It has the value 1 if $i = k$ and 0 if $i \neq k$. The Kronecker delta function has the same effect in summations as the Dirac delta function [eq. (8.4)] has in integrals.

Table 8-1a. Simple iterative deconvolution -- a significant portion of the computations of table 8-1 repeated with more data and smaller wavelength intervals.

λ	S	$E^{(0)}$	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$	$E^{(7)}$	$E^{(8)}$	$E^{(9)}$	E	$S^{(9)}$
0.0	0.00	.00	-.02	.05	.02	.01	.00	-.00	-.01	-.01	-.01	.00	.01
0.5	0.00	.00	-.03	.05	.03	.02	.01	.01	.00	.00	-.00	.00	.02
1.0	0.02	.00	-.06	.05	.03	.03	.02	.01	.01	.01	.01	.00	.04
1.5	0.11	.01	-.09	.04	.02	.02	.02	.02	.01	.01	.01	.00	.12
2.0	0.33	.02	-.13	.01	-.00	.01	.01	.01	.01	.01	.01	.00	.34
2.5	0.81	.05	-.17	-.02	-.03	-.01	-.00	-.00	.00	.00	.00	.00	.80
3.0	1.69	.11	-.19	-.04	-.05	-.03	-.02	-.02	-.01	-.01	-.00	.00	1.66
3.5	3.13	.20	-.18	-.05	-.06	-.04	-.03	-.03	-.02	-.02	-.01	.00	3.09
4.0	5.33	.33	-.12	-.02	-.04	-.02	-.02	-.01	-.01	-.00	-.00	.00	5.31
4.5	8.54	.53	.01	.09	.06	.07	.07	.07	.07	.07	.07	.06	8.53
5.0	12.94	.81	.25	.29	.26	.26	.26	.26	.26	.26	.25	.25	12.94
5.5	18.64	1.17	.61	.62	.58	.58	.57	.57	.57	.56	.56	.56	18.65
6.0	25.67	1.60	1.09	1.08	1.03	1.03	1.02	1.01	1.01	1.01	1.01	1.00	25.68
6.5	33.93	2.12	1.68	1.67	1.62	1.61	1.60	1.59	1.59	1.58	1.58	1.56	33.96
7.0	43.27	2.70	2.39	2.36	2.33	2.32	2.31	2.30	2.29	2.29	2.29	2.25	43.30
7.5	53.41	3.34	3.17	3.16	3.14	3.13	3.12	3.12	3.11	3.11	3.11	3.06	53.44
8.0	64.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	64.00
8.5	74.59	4.66	4.83	4.84	4.86	4.87	4.88	4.88	4.89	4.89	4.89	4.94	74.56
9.0	84.73	5.30	5.61	5.64	5.67	5.68	5.69	5.70	5.71	5.71	5.71	5.75	84.69
9.5	94.07	5.88	6.32	6.33	6.38	6.39	6.40	6.41	6.41	6.42	6.42	6.44	94.04
10.0	102.33	6.40	6.91	6.92	6.96	6.97	6.98	6.99	6.99	6.99	6.99	7.00	102.32
10.5	109.36	6.83	7.39	7.38	7.42	7.42	7.43	7.43	7.43	7.44	7.44	7.44	109.35
11.0	115.06	7.19	7.75	7.71	7.74	7.74	7.75	7.74	7.75	7.75	7.75	7.75	115.06
11.5	119.46	7.47	7.99	7.91	7.94	7.93	7.93	7.93	7.93	7.93	7.93	7.94	119.47
12.0	122.67	7.67	8.12	8.02	8.04	8.02	8.02	8.01	8.01	8.00	8.00	8.00	122.69
12.5	124.88	7.80	8.18	8.05	8.06	8.04	8.03	8.03	8.02	8.02	8.02	8.00	124.90
13.0	126.31	7.89	8.19	8.04	8.05	8.03	8.02	8.02	8.01	8.01	8.01	8.00	126.33
13.5	127.19	7.95	8.17	8.02	8.03	8.01	8.00	8.00	8.00	8.00	8.00	8.00	127.19
14.0	127.67	7.98	8.13	7.99	8.01	7.99	7.99	7.99	7.99	7.99	7.99	8.00	127.66
14.5	127.89	7.99	8.09	7.96	7.99	7.97	7.98	7.98	7.98	7.99	7.99	8.00	127.87
15.0	127.98	8.00	8.06	7.95	7.98	7.97	7.98	7.98	7.98	7.99	7.99	8.00	127.96
15.5	128.00	8.00	8.03	7.94	7.98	7.98	7.99	7.99	7.99	8.00	8.00	8.00	127.98
16.0	128.00	8.00	8.02	7.95	7.99	7.99	8.00	8.00	8.00	8.00	8.00	8.00	127.99
16.5	128.00	8.00	8.01	7.96	8.00	8.00	8.01	8.01	8.01	8.01	8.01	8.00	128.00
17.0	128.00	8.00	8.00	7.97	8.01	8.01	8.01	8.01	8.01	8.01	8.01	8.00	128.01
17.5	128.00	8.00	8.00	7.98	8.02	8.01	8.01	8.01	8.01	8.01	8.01	8.00	128.01
18.0	128.00	8.00	8.00	7.99	8.02	8.01	8.01	8.01	8.00	8.00	8.00	8.00	128.01
18.5	128.00	8.00	8.00	7.99	8.01	8.00	8.01	8.00	8.00	8.00	8.00	8.00	128.01
19.0	128.00	8.00	8.00	8.00	8.01	8.00	8.00	8.00	8.00	8.00	7.99	8.00	128.00
19.5	128.00	8.00	8.00	8.00	8.01	8.00	8.00	8.00	8.00	7.99	7.99	8.00	128.00
20.0	128.00	8.00	8.00	8.00	8.01	8.00	8.00	8.00	8.00	8.00	8.00	8.00	127.99
20.5	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	127.99
21.0	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	127.99
21.5	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	127.99
22.0	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
22.5	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00
23.0	128.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	128.00

$$K \cdot \sum_{k=-n_2}^{n_2} z_{k-i} \cdot v_{\ell-k} = \delta_{\ell i} \quad [\text{dimensionless}], \qquad \begin{matrix} i = -n_2,\ldots,n_2 \\ \ell = -n_2,\ldots,n_2 \end{matrix} \tag{8.40}$$

then eq. (8.39) would become

$$\sum_{k=-n_2}^{n_2} v_{\ell-k} \cdot S_k = \Delta\lambda \cdot \sum_{i=-n_2}^{n_2} E'_i \cdot \delta_{\ell i} = \Delta\lambda \cdot E'_\ell \quad [W \cdot m^{-2}] \tag{8.41}$$

or

$$E'_\ell = \frac{1}{\Delta\lambda} \cdot \sum_{k=-n_2}^{n_2} v_{\ell-k} \cdot S_k \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad \ell = -n_2,\ldots,n_2, \tag{8.42}$$

which is a simple, direct solution of eq. (8.38) for E'. If the z_i are known the v_i $[S^{-1} \cdot W \cdot m^{-2}]$ are uniquely determined by eq. (8.40) and the quantities E'_ℓ $[W \cdot m^{-2} \cdot nm^{-1}]$ can be evaluated for any number of spectra, S, using eq. (8.42) without recomputing the v_i as long as the slit function remains unchanged. Comparison of eqs. (8.38) and (8.42) shows that the vector of v_i $[S^{-1} \cdot W \cdot m^{-2}]$ values constitutes a kind of reciprocal responsivity function and it also possesses the appropriate unit dimensions for this interpretation. To obtain the v_i from eq. (8.40) we first multiply both sides by $\Delta\lambda$ and change the signs of i, k, and ℓ:

$$\Delta\lambda \cdot \delta_{\ell i} = \Delta\lambda \cdot K \cdot \sum_{k=n_2}^{-n_2} z_{i-k} \cdot v_{k-\ell} \quad [nm] \tag{8.43}$$

(by its definition, of course $\delta_{-\ell,-i} = \delta_{\ell i} = \delta_{i\ell}$). Then letting $\ell = 0$ and observing that as long as we sum over all values of k in the range from $-n_2$ to n_2 the limits on the summation can be interchanged, we obtain

$$\Delta\lambda \cdot \delta_{io} = \Delta\lambda \cdot K \cdot \sum_{k=-n_2}^{n_2} z_{i-k} \cdot v_k \quad [nm], \qquad i = -n_2,-n_2+1,\ldots,-1,0,1,\ldots n_2. \tag{8.44}$$

But eq. (8.44) has exactly the same form as eq. (8.38) so that any technique for solving eqs. (8.38) [or eqs. (8.5)] can be used to obtain the quantities v_i simply by starting with a "signal" which is $\Delta\lambda$ at the center and zero everywhere else.

So, in summary, to use this technique, we apply eqs. (8.10) to solve iteratively for values of $E_i^{(\nu)}$ when S_k is given numerically by

$$S_k = \Delta\lambda \cdot \delta_{kn_2}, \qquad k = 1,2,\ldots,n(\text{odd}) \tag{8.45}$$

where $n_2 = (n-1)/2$ and $R_{ki} = K \cdot z_{k-i}$. Now denoting this solution after ν iterations by $v_i^{(\nu)}$, i.e., numerically

$$v_i^{(\nu)} \equiv E_i^{(\nu)} \tag{8.46}$$

we can use these $v_i^{(\nu)}$ to compute a ν^{th} order deconvolution of any set of real measurements from

$$E'^{(\nu)}_{\ell} = \frac{1}{\Delta\lambda} \cdot \sum_{k=1}^{n} v^{(\nu)}_{\ell-k} \cdot S_k \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad \ell = 1,2,\ldots,n. \tag{8.47}$$

For want of a better name we will call the set of values $v_k^{(\nu)}$ the ν^{th} order <u>inverse slit-scattering function</u>. The idea of the ν^{th} order inverse slit-scattering function could presumably be generalized to apply to any arbitrary responsivity matrix by iteratively computing each column $V_{kj}^{(\nu)}$ of a complete ν^{th} order inverse responsivity function matrix using the appropriate $S_i = \Delta\lambda \cdot \delta_{ij}$ for $j = 1,2,\ldots,n$. However, as far as we know, this has never been done. In the slit-scattering function approximation a full matrix of the $V_{kj}^{(\nu)}$ values can, of course, be assembled looking exactly like the Z matrix of eq. (8.23) with z everywhere replaced by $v^{(\nu)}$: in this approximation the inverse slit-scattering function matrix $V^{(\nu)}$ has the same kind of symmetry[1] as that possessed by the slit-scattering function matrix. This matrix $V^{(\nu)}$ would satisfy the matrix equation

$$E'^{(\nu)} = \frac{1}{\Delta\lambda} \cdot V^{(\nu)} \cdot S \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.48}$$

In table 8-2 we show the results of inverse slit-scattering function calculations using the same slit-scattering function as in the example of table 8-1. In columns 3 to 12 we show successive approximations $v^{(\nu)}$ for the delta-function-like S given in column 2. These are computed using eqs. (8.10) exactly as the corresponding columns of table 8-1 were computed. In the last column we show the results of computing

$$E_k^{(9)} = \frac{1}{\Delta\lambda} \cdot \sum_{i=1}^{31} v^{(9)}_{k-i} \cdot S_i \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad k = 1,2,\ldots,31, \tag{8.49}$$

---

[1]The symmetry we are talking about in this chapter is not the symmetry that a mathematician has in mind when he speaks of a symmetric matrix. A symmetric matrix in that sense means that the matrix and its transpose are identical. Thus, a symmetric matrix is symmetric about its principle diagonal, which is the diagonal running from upper left to lower right. The special slit-scattering function matrix symmetry, or row-by-row invariance with which we are concerned is a kind of symmetry in which all elements of any diagonal parallel to the principle diagonal have the same value. Such a matrix will also exhibit the customary symmetry if the slit-scattering function is symmetric but we make no use of that matrix property in this chapter.

Table 8-2. Iterative calculation of inverse slit-scattering function

λ	$\Delta\lambda \cdot \delta_{k,16}$	$v^{(0)}$	$v^{(1)}$	$v^{(2)}$	$v^{(3)}$	$v^{(4)}$	$v^{(5)}$	$v^{(6)}$	$v^{(7)}$	$v^{(8)}$	$v^{(9)}$	$E'^{(9)}$
0	0	0	0	0	0	0	0	0	0	0	0	−.00
2	0	0	0	0	0	0	0	0	0	0	0	.00
4	0	0	0	0	0	0	0	0	0	0	0	.08
6	0	0	0	0	0	0	0	0	0	0	0	1.17
8	0	0	0	0	0	0	0	0	0	0	0	4.00
10	0	0	0	0	0	0	0	0	0	0	0	6.83
12	0	0	0	0	0	0	0	0	0	0	−.0000	7.92
14	0	0	0	0	0	0	0	0	0	.0000	.0001	8.00
16	0	0	0	0	0	0	0	0	−.0001	−.0002	−.0006	8.00
18	0	0	0	0	0	0	0	.0002	.0006	.0014	.0027	8.00
20	0	0	0	0	0	0	−.0005	−.0017	−.0037	−.0063	−.0097	8.00
22	0	0	0	0	0	.0015	.0046	.0091	.0147	.0212	.0285	8.00
24	0	0	0	0	−.0046	−.0123	−.0221	−.0333	−.0453	−.0580	−.0711	8.00
26	0	0	0	.0139	.0324	.0525	.0730	.0936	.1140	.1340	.1536	8.00
28	0	0	−.0417	−.0833	−.1204	−.1543	−.1857	−.2150	−.2426	−.2688	−.2937	8.00
30	2	1250	.2083	.2639	.3102	.3503	.3863	.4192	.4498	.4783	.5053	8.00
32	0	0	−.0417	−.0833	−.1204	−.1543	−.1857	−.2150	−.2426	−.2688	−.2937	8.00
34	0	0	0	.0139	.0324	.0525	.0730	.0936	.1140	.1340	.1536	8.00
36	0	0	0	0	−.0046	−.0123	−.0221	−.0333	−.0453	−.0580	−.0711	8.00
38	0	0	0	0	0	.0015	.0046	.0091	.0147	.0212	.0285	8.00
40	0	0	0	0	0	0	−.0005	−.0017	−.0037	−.0063	−.0097	8.00
42	0	0	0	0	0	0	0	.0002	.0006	.0014	.0027	8.00
44	0	0	0	0	0	0	0	0	−.0001	−.0002	−.0006	8.00
46	0	0	0	0	0	0	0	0	0	.0000	.0001	8.00
48	0	0	0	0	0	0	0	0	0	0	−.0000	8.00
50	0	0	0	0	0	0	0	0	0	0	0	8.03
52	0	0	0	0	0	0	0	0	0	0	0	7.86
54	0	0	0	0	0	0	0	0	0	0	0	8.48
56	0	0	0	0	0	0	0	0	0	0	0	6.65
58	0	0	0	0	0	0	0	0	0	0	0	11.20
60	0	0	0	0	0	0	0	0	0	0	0	1.37

where $v_j^{(9)} = 0$ if $j \leq 0$ or $j \geq 32$ and where S is now the "observed" signal given in column 2 of table 8-1. A comparison of this last column of table 8-2 with the $E^{(9)}$ column of table 8-1 shows no differences aside from the end-effects at the bottom of the table. The advantage of the calculation of the inverse slit-scattering function $v^{(\nu)}$ of table 8-2 is that any set of data taken using this slit-scattering function may now be deconvoluted using eq. (8.42), whereas without this calculation the entire iterative process must be repeated for each new set of data. There are two minor disadvantages, however: (1) to see convergence one must convolute the inverse slit-scattering functions with data; one usually cannot detect convergence in the sequence of inverse slit-scattering functions of successive orders. (2) The inverse slit-scattering function approach, as we have outlined it, is valid only when the responsivity function can be approximated by the product of a slit-scattering function and a responsivity factor.

UNIQUENESS and NOISE in DECONVOLUTION. If $R(\lambda_o,\lambda)$ is accurately known and $S(\lambda_o)$ is measured it would seem, in principle, that enough information should be available to calculate $E(\lambda)$ with about the same degree of accuracy as $S(\lambda_o)$ possesses. Unfortunately, we cannot extract this much information from eq. (8.1). First of all, the solution of eq. (8.1) is not mathematically unique. Consider, for example, the simple rectangularly shaped responsivity function

$$\begin{aligned} R_E(\lambda_o,\lambda) &= 1 \quad \text{for} \quad |\lambda_o-\lambda| < 1/2 \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{8.50}$$

Then eq. (8.1) becomes

$$S(\lambda_o) = \int_{\lambda_o-1/2}^{\lambda_o+1/2} E_\lambda(\lambda)\cdot d\lambda \quad [S]. \tag{8.51}$$

Now suppose we have a solution $E_\lambda(\lambda)$ which satisfies eq. (8.51) for some set of observations $S(\lambda_o)$. Clearly

$$E_\lambda(\lambda) + \sum_{n=1}^{\infty} (A_n\cdot\cos 2\pi n\lambda + B_n\cdot\sin 2\pi n\lambda) \tag{8.52}$$

will also satisfy eq. (8.51) for any values of the A_n and B_n because

$$\int_{\lambda_o-1/2}^{\lambda_o+1/2} \cos 2\pi n\lambda\cdot d\lambda = \int_{\lambda_o-1/2}^{\lambda_o+1/2} \sin 2\pi n\lambda\cdot d\lambda = 0. \tag{8.53}$$

Thus there are an infinity of solutions to eq. (8.51) represented by eq. (8.52) and the one "true" solution cannot be determined from eq. (8.51) alone. Appeal must be made to other desired properties of the solution such as smoothness or non-negativity to find an acceptable solution. In the iterative deconvolution technique, for example, we obtain a solution

which has the property that it satisfies eqs. (8.5) within the measurement uncertainty and exhibits, in some sense, minimal departures from the spectral distribution of the measured signal. Later we will discuss a deconvolution technique which picks out the solution which, within limits, optimizes a measure of the smoothness of $E_\lambda(\lambda)$. Such solutions may still not precisely describe the true spectral irradiance but they will usually be acceptable solutions in terms of one's intuitive notions about the nature of the spectral irradiance distribution.

In principle there is another similar cause of non-uniqueness in a real deconvolution due to the wavelength interval of sampling, $\Delta\lambda$. When $S(\lambda_o)$ is known only at a number of discrete points, the curve of $S(\lambda_o)$ may have any conceivable shape in the intervals $\Delta\lambda$ between measured points and each of these possible $S(\lambda_o)$ curves will give rise to a different solution $E_\lambda(\lambda)$; even the values of $E_\lambda(\lambda_i)$ at the measured points, λ_i, will depend upon how $S(\lambda_o)$ is interpolated between the measured points. Fortunately, in practice one intuitively takes the wavelength interval $\Delta\lambda$ small enough so that usually no significant excursions of $S(\lambda_o)$ remain unsampled. In any case the choice of the wavelength interval is controlled by the experimenter so that non-uniqueness from this source can be minimized easily whereas that discussed above, which is due to the breadth and shape of the responsivity function, can be reduced only by modifying the responsivity function -- by narrowing the slits, for example. Presumably, however, if the responsivity function could have been narrowed, the experimenter would already have done this, balancing the improved spectral resolution against the increased noise accompanying the lower signal levels to arrive at over-all optimum experimental conditions for the total measurement time available.

Another related aspect of the problem is that when the spectroradiometer physically converts $E_\lambda(\lambda)$ into $S(\lambda_o)$ it performs a smoothing operation, flattening out the hills and filling in the valleys of $E_\lambda(\lambda)$ (see, e.g., figure 7.14c). In fact, the sharper the detail in $E_\lambda(\lambda)$ the more it is smoothed out by the instrument. Therefore any mathematical scheme which reconstructs $E_\lambda(\lambda)$ from $S(\lambda_o)$ must amplify any small bumps and dips of $S(\lambda_o)$ to rebuild the hills and valleys of $E_\lambda(\lambda)$. And the finer the detail, the greater must be the amplification. The result of this is that random experimental noise in the $S(\lambda_o)$ values will receive great amplification so that in the presence of noise any exact solution of eq. (8.1) is likely to produce a distorted spectral irradiance distribution containing a wealth of spurious spectral detail. In many respects deconvolution may be thought of as equivalent to carrying out the same experiment in the same length of time but with narrowed spectroradiometer slits and a smaller wavelength interval: the total number of photons collected during the experiment is the same but it is now distributed among more samples resulting in a large increase in the relative noise in the high resolution samples compared to that in the original, actually measured samples. Any attempt to achieve perfect spectral resolution must result in infinite relative noise. A further consequence of the physical smoothing performed by the spectroradiometer is that any spectral detail in $E_\lambda(\lambda)$ which is smoothed to the point where it is lost in the measurement noise [as the sinusoidal

terms in the example of eq. (8.52) would be] will be irretrievably lost: it is not recoverable by any conceivable mathematical manipulation.

We can easily see how the iterative procedure selectively amplifies sharp spectral features. Suppose the measurements S_k contain a sharp spectral feature at λ_ℓ. Then, of course, $E_\ell^{(0)}$ will contain this same structure [see eq. (8.6)]. However, the calculated values of $S_k^{(0)}$ in the neighborhood of λ_ℓ will contain a much broader version of this feature because the calculation $S_k^{(\nu)} = \Delta\lambda \cdot \sum_i R_{ik} \cdot E_i^{(\nu)}$ simply performs mathematically the same smoothing operation on the $E_i^{(\nu)}$ as the radiometer does on the physical spectrum. Therefore the correction term $S_k - S_k^{(0)}$ will still contain the feature at λ_ℓ in a slightly attenuated version. When such corrections are added again and again in the course of generating $E^{(\nu)}$ the magnitude of the excursion at λ_ℓ will have increased by a factor of nearly $A_\ell \cdot \nu$. A very broad spectral feature, on the other hand, will be reproduced quite faithfully in computing $S^{(0)}$ so that the correction term $S - S^{(\nu)}$ will quickly vanish as ν increases.

A GENERALIZED LEAST-SQUARES DECONVOLUTION. We have described the iterative deconvolution method in detail because of its intuitive appeal, its frequent appearance in the literature, and because it will work in almost all radiometric applications. The number of iterations required to reach a particular degree of precision depends mainly upon the amount of distortion that must be removed (and, of course, upon the degree of precision required). Assuming that the ultimate spectral resolution sought is comparable to the wavelength interval $\Delta\lambda$, if $\Delta\lambda$ is wider than or comparable to the spectral slit width then a few iterations will suffice. However, if $\Delta\lambda$ is much smaller than the spectral slit width then dozens or even hundreds of iterations may be required to achieve a satisfactory deconvolution. Although successful iterative deconvolutions can be effected in these cases [8.2,8.20] -- usually by making use of techniques for accelerating convergence and for smoothing to minimize the amplification of experimental noise which accompanies the iterations -- this approach to deconvolution is not generally recommended [8.11,8.21] if more than 5 or 10 iterations prove to be required. Occasionally the method fails by divergence of the iterations before the condition of eq. (8.13) is satisfied. The method is most likely to fail [8.21] with markedly asymmetric responsivity functions, with noisy data, and when spectral detail narrower in width than the spectral slit width is sought. In the next sections we present a deconvolution technique which is somewhat less likely to fail. This is not to say that the solution will always be satisfactory: attempts to extract fine spectral details from noisy data taken with coarse slits will always lead to solutions with a substantial credibility gap. The technique to be described is a generalized least-squares or constrained least-squares deconvolution method and has been described in various forms by many authors [8.13,8.21,8.22,8.23]. We will not attempt a derivation of the method here but will simply outline the approach.

Consider the following expression evaluated for any arbitrary estimate of the spectral irradiance function $E_\lambda(\lambda)$

$$\int \left\{ S(\lambda_o) - \int R(\lambda_o,\lambda) \cdot E_\lambda(\lambda) \cdot d\lambda \right\}^2 \cdot d\lambda_o + \gamma \cdot \int \left| \frac{d^2 E_\lambda(\lambda)}{d\lambda^2} \right|^2 \cdot d\lambda \quad [S^2 \cdot nm]. \tag{8.54}$$

The integrand of the first integral is the square of the discrepancy between the measurement $S(\lambda_o)$ at λ_o and the instrument output calculated from this $E_\lambda(\lambda)$. The integral over λ_o is then a measure of the total error throughout the spectrum. The last integral is a measure of the roughness of the function $E_\lambda(\lambda)$: if this $E_\lambda(\lambda)$ has sharp excursions then the second derivative will take on large values and this integral will be large. The complete expression is thus simultaneously a measure of the magnitude of the error and the roughness or irregularity of the function $E_\lambda(\lambda)$. The parameter γ $[S^2 \cdot W^{-2} \cdot m^4 \cdot nm^4]$ determines how much relative weight is to be given to these two attributes of $E_\lambda(\lambda)$ (it also absorbs some dimensional constants since the dimensions of the two components of expression (8.54) are otherwise obviously different). The minimization of expression (8.54) with respect to the parameters which describe $E_\lambda(\lambda)$ then yields a generalized least-squares solution which simultaneously minimizes errors and roughness with a relative weighting determined by γ. In the case of discrete measurements, expression (8.54) takes the form

$$\sum_k \left(S_k - \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i \right)^2 + \gamma \cdot \sum_i \left(E_{i-1} - 2E_i + E_{i+1} \right)^2 / (\Delta\lambda)^2 \quad [S^2] \tag{8.55}$$

and the minimization leads to the solution

$$\mathsf{E} = \frac{1}{\Delta\lambda} \cdot \left(\mathsf{R}^T \cdot \mathsf{R} + \gamma \cdot \mathsf{H} \right)^{-1} \cdot \mathsf{R}^T \cdot \mathsf{S} \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.56}$$

This is a matrix equation in which the quantities E and S, the spectral irradiance and instrument output vectors, and R the responsivity matrix, have already been introduced. The superscript T on a matrix symbol means the matrix transpose of that matrix -- obtained by interchanging rows and columns: i.e., $R^T_{r,c} = R_{c,r}$. The superscript -1 on a matrix symbol means the matrix inverse of that matrix. The product of a matrix by its inverse results in a matrix known as the identity or unit matrix[1]:

[1]The notation for the matrix inverse and for the unit matrix brings out the analogy between these quantities and the corresponding algebraic quantities, reciprocal and unity. Unfortunately the matrix inverse is difficult to calculate especially for large matrices and the analogy is further marred by some of the other idiosyncrasies of matrix expressions such as the importance of the order of the factors in multiplication. Since we are not attempting to derive the equations in this section, familiarity with matrix algebra should not be necessary. For interested readers, brief introductions to matrix algebra are given in [8.13] and [8.24]. Complete treatments can be found in textbooks on matrix algebra or linear algebra.

$$A^{-1}\cdot A = I = \begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & . & . & . \\ 0 & 0 & 1 & & & \\ & . & & 1 & 0 & \\ & . & & 0 & 1 & \end{bmatrix} \quad [\text{dimensionless}] \qquad (8.57).$$

The matrix H is given by

$$H = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & & \\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & & \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & . . . & \\ 0 & 0 & 1 & -4 & 6 & -4 & 1 & & \\ & & & . & & & & & \\ & & & . & & & & & \\ & & & . & & & & & \end{bmatrix} \quad [nm^{-4}]. \qquad (8.58)$$

The matrix is symmetrical about both diagonals and except for the first two rows and the last two rows, all rows are similar.

The matrix H given in eq. (8.58) tends to force $E_\lambda(\lambda)$ in the direction of linearity. Other choices for the matrix H are possible corresponding to different choices for the smoothing term in expression (8.54). For example, the matrix

$$H = \begin{bmatrix} 1 & -3 & 3 & -1 & 0 & 0 & 0 & 0 & \\ -3 & 10 & -12 & 6 & -1 & 0 & 0 & 0 & \\ 3 & -12 & 19 & -15 & 6 & -1 & 0 & 0 & . . \\ -1 & 6 & -15 & 20 & -15 & 6 & -1 & 0 & \\ 0 & -1 & 6 & -15 & 20 & -15 & 6 & -1 & \\ & & & & . & & & & \\ & & & & . & & & & \end{bmatrix} \quad [nm^{-5}] \qquad (8.59)$$

corresponding to a term $\gamma\cdot\int \left\{\frac{d^3E_\lambda(\lambda)}{d\lambda^3}\right\}^2\cdot d\lambda \ [S^2\cdot nm]$ in expression (8.54) tends to force $E_\lambda(\lambda)$ toward a quadratic. Usually there is little difference in the final results. Other examples are given in [8.21,8.22,8.25], and especially [8.13].

When $\gamma = 0$ in eq. (8.56) the solution reduces to

$$E = \frac{1}{\Delta\lambda}\cdot\left(R^T\cdot R\right)^{-1}\cdot R^T\cdot S \quad [W\cdot m^{-2}\cdot nm^{-1}], \qquad (8.60)$$

which is the least-squares solution for a set of equations like eq. (8.17). When $\gamma = 0$ and the number of equations is exactly equal to the number of unknowns, as it is in our case, eq. (8.60) further reduces to

$$E = \frac{1}{\Delta\lambda} \cdot R^{-1} \cdot S \quad [W \cdot m^{-2} \cdot nm^{-1}], \tag{8.61}$$

which is just the formal, exact solution of eq. (8.17). As we have noted this solution is likely to be unsatisfactory in deconvolution because of the problem of noise amplification. As γ increases eq. (8.56) gives solutions of increasing smoothness. Eventually, for large γ, the calculation of the matrix inverse $\left(R^T \cdot R + \gamma \cdot H\right)^{-1}$ will generally fail. This appears to be a numerical problem associated with the fact that the determinant of H is zero.

The practical difference between eq. (8.56) and eq. (8.61) as solutions of the deconvolution problem is that because of the smoothing represented by the term $\gamma \cdot H$ $[S^2 \cdot W^{-2} \cdot m^4]$ eq. (8.56) may be solved by exact methods. If explicit matrix operations are available in the computer language employed and if n is not too large the matrix operations can be carried out directly. In languages without explicit matrix operations the calculation is more trouble, but still straightforward: the matrix inversion is the most difficult step and algorithms for this are available in most computer centers.

So, in summary, in the generalized least-squares technique we choose a value of γ $[S^2 \cdot W^{-2} \cdot m^4 \cdot nm^4]$ and compute the values of $E_i(\gamma)$ $[W \cdot m^{-2} \cdot nm^{-1}]$ using eq. (8.56). One might start, for example, with $\gamma \approx .01R^{*2}$ where R^* is the largest element in the responsivity matrix R $[S \cdot W^{-1} \cdot m^2]$. We then evaluate some measure of the agreement between the reconstructed signal and the measured signal, such as the standard deviation:

$$\sigma(\gamma) = \left\{ \frac{1}{n-1} \cdot \sum_k \left[S_k - \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i(\gamma) \right]^2 \right\}^{\frac{1}{2}} \quad [S]. \tag{8.62}$$

Now, depending upon whether $\sigma(\gamma)$ is larger or smaller than the estimated random measurement uncertainties, a smaller or larger value of γ is chosen -- say by a factor of 10 -- and the calculation is repeated. The goal is to introduce as much smoothing as possible -- as large a value of γ as possible -- consistent with the measured signal, allowing for its uncertainty. The solution is not very sensitive to γ and a less-than-optimum choice by an order-of-magnitude will not seriously affect the results.

In table 8-3 we show the results of solving the deconvolution problem of table 8-1 using this generalized least-squares technique. The last four columns of the table are solutions obtained using different values of γ. Notice that we still have an end-effect at the bottom of the table. This is because the technique attempts to recreate the sudden jump in S at $\lambda = 60$ and, as we have said, the results would be perfectly correct if the measured S_k values actually dropped abruptly to zero at this point. The similarity between this family of solutions and the successive iterates of table 8-1 is apparent. The last row

Table 8-3. Generalized least-squares deconvolution

λ	S	γ=.01	γ=.1	γ=1	γ=10
0	.00	.00	.00	.04	-.10
2	.33	-.00	.00	-.06	-.16
4	5.33	.08	.07	.06	.20
6	25.67	1.17	1.19	1.29	1.59
8	64.00	4.00	4.00	4.00	4.00
10	102.33	6.83	6.81	6.71	6.41
12	122.67	7.91	7.93	7.95	7.81
14	127.67	8.01	8.00	8.04	8.16
16	128.00	7.99	8.00	7.98	8.07
18	128.00	8.01	8.00	8.00	7.99
20	128.00	7.99	8.00	8.00	7.98
22	128.00	8.00	8.00	8.00	8.00
24	128.00	8.00	8.00	8.00	8.00
26	128.00	7.99	8.00	8.00	8.00
28	128.00	8.00	8.00	8.00	8.00
30	128.00	8.00	8.00	8.00	8.00
32	128.00	7.98	8.00	8.00	8.00
34	128.00	8.06	8.00	8.00	8.00
36	128.00	7.89	8.00	8.00	8.00
38	128.00	8.18	7.99	8.00	8.00
40	128.00	7.73	8.02	8.00	8.00
42	128.00	8.37	7.97	8.00	8.00
44	128.00	7.54	8.01	7.99	8.00
46	128.00	8.51	8.04	8.00	8.00
48	128.00	7.53	7.84	8.02	8.01
50	128.00	8.26	8.34	7.95	8.03
52	128.00	8.19	7.52	7.99	8.01
54	128.00	7.02	8.39	8.29	7.83
56	128.00	10.17	8.30	7.47	7.70
58	128.00	4.20	6.05	7.62	8.49
60	128.00	13.80	12.66	11.43	10.34
		.02	.05	.34	1.66

of table 8-3 is again the rms deviation of the computed $\Delta\lambda \cdot \sum_i R_{ik} \cdot E_i$ from the measured S_k for the first sixteen values of the four columns of E_i values. For measurement uncertainties in S_k of 0.1% something between $\gamma = .1$ and $\gamma = 1$ would apparently be appropriate. The matrices R and H used in this solution have been given in eqs. (8.30) and (8.58). As an aid to readers unaccustomed to working with matrices we will briefly indicate the appearance of the rest of the matrices involved in this example. Since R is symmetric about the diagonal running from upper left to lower right R and R^T are the same. The product $R^T \cdot R$ is

$$R^T \cdot R = \begin{bmatrix} 20 & 16 & 4 & 0 & 0 & 0 & 0 & & \\ 16 & 24 & 16 & 4 & 0 & 0 & 0 & & \\ 4 & 16 & 24 & 16 & 4 & 0 & 0 & . \; . \; . & \\ 0 & 4 & 16 & 24 & 16 & 4 & 0 & & \\ 0 & 0 & 4 & 16 & 24 & 16 & 4 & & \\ & & & . & & & & & \\ & & & . & & & & & \end{bmatrix} [S^2 \cdot W^{-2} \cdot m^4]. \qquad (8.63)$$

For $\gamma = 1$ the matrix $R^T \cdot R + \gamma \cdot H$ becomes

$$R^T \cdot R + H = \begin{bmatrix} 21 & 14 & 5 & 0 & 0 & 0 & 0 & & \\ 14 & 29 & 12 & 5 & 0 & 0 & 0 & & \\ 5 & 12 & 30 & 12 & 5 & 0 & 0 & . \; . \; . & \\ 0 & 5 & 12 & 30 & 12 & 5 & 0 & & \\ 0 & 0 & 5 & 12 & 30 & 12 & 5 & & \\ & & & . & & & & & \\ & & & . & & & & & \end{bmatrix} [S^2 \cdot W^{-2} \cdot m^4]. \qquad (8.64)$$

A portion of the middle row (or column) of the inverse of the matrix of eq. (8.64) is

$$10^{-6}(. \; . \; . \; -73 \; 395 \; -425 \; -1125 \; 4375 \; -3125 \; -15625 \; 46875 \; -15625 \; -3125 \; . \; . \; .). \qquad (8.65)$$

Neighboring rows (or columns) are similar but displaced just as in eq. (8.64); however, the departures from this typical row at the edges of the array are more pronounced than in eq. (8.64). Finally, the product of $\left(R^T \cdot R + H\right)^{-1}$ by R^T for $\gamma = 1$ gives a matrix whose middle row (or column) is

$$(. \; . \; . \; .0004 \; .0006 \; -.0032 \; .0034 \; .009 \; -.035 \; .025 \; .125 \; .025 \; -.035 \; . \; . \; .). \qquad (8.66)$$

This matrix is functionally the equivalent of the inverse slit-scattering function matrix V described earlier.

There is no requirement in this procedure that the number of E_i values be equal to the number of S_k values or that the wavelength spacings be uniform. In our numerical example, for instance, if an additional measurement were made at $\lambda_o = 5$, it would only be necessary to insert an appropriate additional row into R

$$R = \begin{bmatrix} 4 & 2 & 0 & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & 0 & 0 \\ 0 & 1 & 3 & 3 & 1 & 0 \\ 0 & 0 & 2 & 4 & 2 & 0 \\ & & \vdots & & & \end{bmatrix} \longleftarrow \qquad (8.67)$$

expressing how $S(\lambda_o = 5)$ is to be calculated from the $E(\lambda)$ $(\lambda = 0,2,4,6,8,...)$. The matrix R now will have 32 rows and 31 columns. Matrix multiplication requires that the number of columns of the factor on the left be equal to the number of rows of the factor on the right and then the product matrix will contain as many rows as the first matrix and as many columns as the second. The following diagram illustrates this rule schematically

$$\begin{bmatrix} * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \end{bmatrix} \cdot \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \qquad (8.68)$$

If the matrix R has m $(> n)$ rows and n columns then R^T will have n rows and m columns and matrix $R^T \cdot R$ will be an n by n square matrix. The matrix H then must be chosen as an n by n matrix and we must invert an n by n matrix to solve the deconvolution problem. Note that the size of the matrix to be inverted is given by the number of equally spaced E_i values desired and $\Delta\lambda$ is the spacing of these values.

An ITERATIVE SOLUTION for LARGE NUMBERS of POINTS. If the number of measurement points, n, is large -- greater than something like 100 -- limitations of computer memory capacity and increasing inaccuracies of matrix inversion routines make the direct evaluation of E by eq. (8.56) extremely difficult if not impossible. In this case an iterative solution similar to that of eqs. (8.10) and (8.11) can be carried out. Let us write

$$T = R^T \cdot R + \gamma \cdot H \quad [s^2 \cdot w^{-2} \cdot m^4] \qquad (8.69)$$

and

$$Q = R^T \cdot S \quad [s^2 \cdot w^{-1} \cdot m^2]; \tag{8.70}$$

then eq. (8.56) can be written

$$E = \frac{1}{\Delta\lambda} \cdot T^{-1} \cdot Q \quad [W \cdot m^{-2} \cdot nm^{-1}] \tag{8.71}$$

or $Q = \Delta\lambda \cdot T \cdot E$ which is the same form as eq. (8.17). Thus the iterative solution can be written [by analogy with eq. (8.19)]

$$E^{(\nu)} = E^{(\nu-1)} + A \cdot \left[Q - \Delta\lambda \cdot T \cdot E^{(\nu-1)} \right] \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.72}$$

$E^{(o)}$ is computed from S by eqs. (8.6) exactly as in the simple iterative deconvolution we discussed earlier. The execution of the iteration scheme of eq. (8.72) normally can be carried further toward convergence than can the simple iterative deconvolution scheme of eq. (8.10) because the term $\gamma \cdot H$ prevents noise build-up. Notice that if the matrix A were taken as the exact inverse of $\Delta\lambda \cdot T$, that is, if $A = \frac{1}{\Delta\lambda} \cdot T^{-1}$ we would have

$$E^{(\nu)} = E^{(\nu-1)} + \frac{1}{\Delta\lambda} \cdot T^{-1} \cdot Q - I \cdot E^{(\nu-1)} = \frac{1}{\Delta\lambda} \cdot T^{-1} \cdot Q \quad [W \cdot m^{-2} \cdot nm^{-1}], \tag{8.73}$$

which is just eq. (8.71) or (8.56). In this case $E^{(\nu)} \left(=E\right)$ is independent of the trial spectral irradiance vector $E^{(\nu-1)}$ and consequently convergence is immediate with no iterations required. Therefore, if A can be chosen as a good approximation to $\frac{1}{\Delta\lambda} \cdot T^{-1}$ the iterations of eq. (8.72) can be expected to converge very quickly. The following procedure may be used to compute a good approximation to $\frac{1}{\Delta\lambda} \cdot T^{-1}$ $[s^{-2} \cdot w^2 \cdot m^{-4} \cdot nm^{-1}]$.

We are looking for a matrix A which satisfies

$$A \cdot T \approx \frac{1}{\Delta\lambda} \cdot I \quad [nm^{-1}], \tag{8.74}$$

where I is the unit matrix. In obtaining A we make use of the fact that the elements of the matrix T are everywhere nearly zero except for a narrow band along the principle diagonal

$$T = \begin{bmatrix} & & & & & & \\ & 0 & 0 & 0 & & & \\ & & 0 & 0 & . & . & . \\ 0 & & & 0 & & & \\ 0 & 0 & & & & & \\ 0 & 0 & 0 & & & & \\ & . & & & & & \\ & . & & & & & \\ & . & & & & & \end{bmatrix} \quad [s^2 \cdot w^{-2} \cdot m^4] \tag{8.75}$$

and that each row is identical or nearly identical to its neighbor except for a one-step horizontal displacement. Then we determine the elements of a matrix A $[s^{-2} \cdot w^{2} \cdot m^{-4} \cdot nm^{-1}]$, with a similar narrow diagonal band of identical non-zero rows which approximately satisfies eq. (8.74). Thus we have a matrix equation which looks something like this

$$\begin{bmatrix} 0 & 0 & & \\ & 0 & \ddots & \\ 0 & & & \\ 0 & 0 & & \\ & \ddots & & \end{bmatrix}_{A} \cdot \begin{bmatrix} 0 & 0 & & \\ & 0 & \ddots & \\ 0 & & & \\ 0 & 0 & & \\ & \ddots & & \end{bmatrix}_{T} \approx \frac{1}{\Delta\lambda} \cdot \begin{bmatrix} 1 & 0 & 0 & & & 0 \\ 0 & 1 & 0 & \ddots & & 0 \\ 0 & 0 & \ddots & & & 0 \\ & \ddots & & \ddots & & \ddots \\ & \ddots & & & 1 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 1 \end{bmatrix}_{I} [nm^{-1}], \qquad (8.76)$$

where the diagonal bands, because of their assumed row-by-row invariance look like

$$\begin{array}{ccccccccc} & & & \vdots & & & & & \\ & & & \vdots & & & & & \\ \cdots & A_{-2} & A_{-1} & A_{0} & A_{1} & A_{2} & \cdots & & \\ & \cdots & A_{-2} & A_{-1} & A_{0} & A_{1} & A_{2} & \cdots & \\ & & \cdots & A_{-2} & A_{-1} & A_{0} & A_{1} & A_{2} & \cdots \\ & & & \cdots & A_{-2} & A_{-1} & A_{0} & A_{1} & A_{2} \cdots \\ & & & & & & \vdots & & \\ & & & & & & \vdots & & \end{array} \qquad (8.76a)$$

and

$$\begin{array}{ccccccccc} & & & \vdots & & & & & \\ & & & \vdots & & & & & \\ \cdots & T_{-2} & T_{-1} & T_{0} & T_{1} & T_{2} & \cdots & & \\ & \cdots & T_{-2} & T_{-1} & T_{0} & T_{1} & T_{2} & \cdots & \\ & & \cdots & T_{-2} & T_{-1} & T_{0} & T_{1} & T_{2} & \cdots \\ & & & \cdots & T_{-2} & T_{-1} & T_{0} & T_{1} & T_{2} \cdots \\ & & & & & & \vdots & & \\ & & & & & & \vdots & & \end{array} \qquad (8.76b)$$

Consider now the equations which generate a single column near the middle of the unit matrix:

$$\begin{array}{lllll}
T_2 \cdot A_1 & & & & = 0 \\
T_1 \cdot A_1 & + & T_2 \cdot A_0 & & = 0 \\
T_0 \cdot A_1 & + & T_1 \cdot A_0 & + & T_2 \cdot A_{-1} = 0 \\
T_{-1} \cdot A_1 & + & T_0 \cdot A_0 & + & T_1 \cdot A_{-1} = 1/\Delta\lambda \\
T_{-2} \cdot A_1 & + & T_{-1} \cdot A_0 & + & T_0 \cdot A_{-1} = 0 \\
 & & T_{-2} \cdot A_0 & + & T_{-1} \cdot A_{-1} = 0 \\
 & & & & T_{-2} \cdot A_{-1} = 0
\end{array} \qquad (8.77)$$

In eqs. (8.77) we have assumed that only the T elements from T_{-2} to T_2 are significantly different from zero and that we desire a matrix A with only 3 non-zero diagonals. In general, with the number n_T of such non-zero diagonals in the T matrix and the number n_A of A elements desired we will have $n_T + n_A - 1$ equations in the n_A unknown A matrix elements. These equations can be solved in a least-squares sense to give a set of A elements which will define a good band-matrix approximation to $\frac{1}{\Delta\lambda} \cdot T^{-1}$. The least-squares solution is given by [compare eq. (8.60)]

$$X = \left(T'^{T} \cdot T'\right)^{-1} \cdot T'^{T} \cdot \begin{bmatrix} \cdot \\ \cdot \\ 0 \\ 1/\Delta\lambda \\ 0 \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \quad [s^{-2} \cdot w^2 \cdot m^{-4} \cdot nm^{-1}], \qquad (8.78)$$

where the elements of vector X, from top-to-bottom are the values A_i, $i = \ldots 2,1,0,-1,-2,\ldots$ and T' is the abbreviated $n_T + n_A - 1$ by n_A matrix

$$\mathbf{T}' = \overbrace{\begin{bmatrix} \cdot & \cdot & 0 & 0 & \cdot \\ & \cdot & \cdot & 0 & \\ & T_2 & \cdot & \cdot & \\ & T_1 & T_2 & \cdot & \\ & T_0 & T_1 & T_2 & \\ \cdot & T_{-1} & T_0 & T_1 & \cdot \\ & T_{-2} & T_{-1} & T_0 & \\ & \cdot & T_{-2} & T_{-1} & \\ & \cdot & \cdot & T_{-2} & \\ & 0 & \cdot & \cdot & \\ \cdot & 0 & 0 & \cdot & \cdot \end{bmatrix}}^{(n_A \text{ columns})} \Bigg\} (n_T + n_A - 1 \text{ rows}) \qquad (8.79)$$

The matrix $\mathbf{T}'^{T} \cdot \mathbf{T}'$ in eq. (8.78) which now must be inverted is only an n_A by n_A matrix. Thus, by using this approximation and accepting the complication of iteration we are able to reduce the dimensions of the matrix which must be inverted from n by n (n is the total number of measured points) to n_A by n_A where n_A is a small number usually taken equal to the number of significantly non-zero elements in a row of the responsivity matrix $\mathbf{R}$. Thus, n_A will normally be equal to 2 or 3 times the nominal spectral slit width in units of $\Delta\lambda$, so that the calculation of $\left(\mathbf{T}'^{T} \cdot \mathbf{T}'\right)^{-1}$ should pose no problem. Of course, n_A may always be chosen smaller, if necessary, at the expense of additional iterations.

If $n_A = 1$ eqs. (8.77) reduce to

$$\begin{aligned} &\cdot \\ &\cdot \\ T_2 \cdot A_0 &= 0 \\ T_1 \cdot A_0 &= 0 \\ T_0 \cdot A_0 &= 1/\Delta\lambda \\ T_{-1} \cdot A_0 &= 0 \\ T_{-2} \cdot A_0 &= 0 \\ &\cdot \\ &\cdot \end{aligned} \qquad (8.80)$$

with the least-squares solution

$$A_0 = T_0/[\Delta\lambda \cdot (\ \ldots T_2^2 + T_1^2 + T_0^2 + T_{-1}^2 + T_{-2}^2 \ldots)] \quad [S^{-2} \cdot W^2 \cdot m^{-4} \cdot nm^{-1}]. \qquad (8.81))$$

This obviously does not solve eqs. (8.80) very well but it represents the best available approximation to $\frac{1}{\Delta\lambda}\cdot\mathsf{T}^{-1}$ in a least-squares sense which consists of a single diagonal.

$$\mathsf{A} = \begin{bmatrix} A_0 & 0 & 0 & 0 & \cdot & \cdot \\ 0 & A_0 & 0 & 0 & & \\ 0 & 0 & A_0 & 0 & & \\ 0 & 0 & 0 & A_0 & & \\ & & & & \cdot & \\ & & & & & \cdot \end{bmatrix} \approx (1/\Delta\lambda)\cdot\mathsf{T}^{-1} \quad [S^{-2}\cdot W^2\cdot m^{-4}\cdot nm^{-1}]. \tag{8.82}$$

This solution, for $n_A = 1$, when applied to the responsivity matrix of the simple iterative deconvolution discussed earlier leads to eq. (8.9). More elaborate A matrices corresponding to $n_A > 1$ can be used in the simple iterative deconvolution just as in the iterations of the generalized least-squares method and will usually result in more rapid convergence. However, one of the most appealing features of the iterative deconvolution technique is its simplicity, both computational and conceptual, and, in our opinion, the effort required to provide an improved A matrix would be better devoted to applying the generalized least-squares technique or a Fourier technique [8.6,8.7,8.26,8.27,8.28] to one's data.

As in the simple iterative deconvolution the value of the matrix A should only affect the rate of convergence, not the final values of spectral irradiance obtained at convergence. If the rows of matrix T are not invariant [as they are in eq. (8.79)], which means that R cannot be approximated by a slit-scattering function, then A will depend upon which particular group of n_A columns of T are chosen for the submatrix T' [eq. (8.79)]. Thus any particular T' may lead to a matrix A which is inappropriate for other spectral regions of the calculation in the sense that convergence may be slow in those regions. In extreme cases it may be necessary to calculate different matrices A for use in different spectral regions; in other words, to break the problem into smaller problems in each of which a single $\mathsf{A} \sim \frac{1}{\Delta\lambda}\cdot\mathsf{T}^{-1}$ provides adequate convergence.

For brevity we have presented the generalized least-squares method, for the most part, in matrix notation. We wish to stress that the implementation of the technique on a computer need not use matrix operations except for the single matrix inversion, which can usually be handled by a call to a suitable subroutine available in most computer-center program libraries. Moreover, to the extent that the matrices are band matrices with invariant rows, the matrix multiplications can be replaced by convolutions. Convolution calculations reduce computer-memory requirements by storing only the non-zero elements of one row of such a matrix and obtaining other matrix elements by juggling the indices of this row. For example, instead of

$$W_r = \left(T \cdot E\right)_r = \sum_{k=1}^{n} T_{rk} \cdot E_k \qquad r = 1,2,\ldots,n$$

and (8.83)

$$P_{rc} = \left(R^T \cdot R\right)_{rc} = \sum_{k=1}^{n} R^T_{rk} \cdot R_{kc} = \sum_{k=1}^{n} R_{kr} \cdot R_{kc} \qquad r,c = 1,2,\ldots,n$$

we can write

$$W(i) = \sum_{k=-n_2}^{n_2} T(k) \cdot E(i-k) \qquad i = -n_2,\ldots,-1,0,1,\ldots,n_2$$

and (8.84)

$$P(i) = \sum_{k=-n_2}^{n_2} R^T(k) \cdot R(i-k) = \sum_{k=-n_2}^{n_2} R(-k) \cdot R(i-k) \qquad i = -n_2,\ldots,-1,0,1,\ldots,n_2.$$

In these expressions (T_{rc}), (R_{rc}) and (P_{rc}) are n by n matrices with n assumed to be an odd number and $n_2 = \frac{1}{2}(n-1)$. The quantities T(k), R(k) and P(k) are the entries along the invariant rows of these matrices as in eqs. (8.76) and (8.76a). That is, for any r:

$$T(k) = T_{r,r+k}, \quad R(k) = R_{r,r+k}, \quad P(k) = P_{r,r+k} = \left(R^T \cdot R\right)_{r,r+k}$$

and (8.85)

$$E(k) = E_{n_2+k+1}, \quad W(k) = W_{n_2+k+1}.$$

Examples of computer programs for both the simple iterative deconvolution technique and the iterative generalized least-squares technique are given in Appendix 8A. Both these programs use convolution instead of matrix multiplication and hence, are applicable only when the slit-scattering function approximation is adequate. The only matrix operations used are those associated with evaluating eq. (8.78).

CONFIDENCE BANDS and the RESIDUAL SPECTRAL RESPONSIVITY FUNCTION. If we can obtain an estimate of the matrix which converts the measured signal vector S into the deconvoluted spectral irradiance vector E then we can calculate how the random uncertainties in the values of E_i are related to the uncertainties in the measured S_i [8.21]. Calling this matrix L we need to know L in the equation

$$\mathsf{E} = \mathsf{L} \cdot \mathsf{S} \quad [W \cdot m^{-2} \cdot nm^{-1}] \tag{8.86}$$

where E, here, is the calculated set of spectral irradiance values which result from deconvolution. Luckily, we already have expressions for L. For the simple iterative deconvolution we have [from eq. (8.48)]

$$L = \frac{1}{\Delta\lambda}\cdot V^{(\nu)} \quad [s^{-1}\cdot W\cdot m^{-2}\cdot nm^{-1}]. \tag{8.87}$$

For the generalized least-squares solution [from eqs. (8.70) and (8.71)]

$$L = \frac{1}{\Delta\lambda}\cdot T^{-1}\cdot R^{T} \quad [s^{-1}\cdot W\cdot m^{-2}\cdot nm^{-1}] \tag{8.88}$$

or, if the iterative routine is used

$$L \approx A\cdot R^{T} \quad [s^{-1}\cdot W\cdot m^{-2}\cdot nm^{-1}]. \tag{8.89}$$

Now if E and S are related by eq. (8.86) then the uncertainties in E are related to the uncertainties in S by

$$\mathrm{cov}(E) = L\cdot\mathrm{cov}(S)\cdot L^{T} \quad [W^2\cdot m^{-4}\cdot nm^{-2}], \tag{8.90}$$

where $\mathrm{cov}(S)$ means the convariance matrix of S, etc. For our purposes we shall make the usual assumption that the random errors in S are uncorrelated, that is, that the magnitude and direction of the error in S_ℓ, say, doesn't depend upon the error that was made in measuring any other value, S_k. In such a case $\mathrm{cov}(S)$ is just the diagonal matrix

$$\mathrm{cov}(S) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & & \\ 0 & \sigma_2^2 & 0 & . & . \\ 0 & 0 & \sigma_3^2 & & \\ & . & & & \\ & . & & & \sigma_n^2 \end{bmatrix} \quad [s^2], \tag{8.91}$$

where σ_k, as in eq. (8.13), is the standard deviation of the measurement of S_k. The diagonal elements of the matrix $\mathrm{cov}(E)$ computed from eq. (8.90) are similarly the squares of the standard deviations, ξ_k, of the values of E_k. The off-diagonal elements of $\mathrm{cov}(E)$ describe the correlation between errors at different wavelengths but we will not discuss these quantities here. In the simplest case where all S_i are subject to the same random error σ, $\mathrm{cov}(S)$ reduces to $\sigma^2\cdot I$ $[s^2]$ where I is the unit matrix. Then eq. (8.90) reduces to

$$\mathrm{cov}(E) = \sigma^2\cdot L\cdot L^{T} \quad [W^2\cdot m^{-4}\cdot nm^{-2}] \tag{8.92}$$

whence

$$\xi_i^2 = \sigma^2\cdot\sum_k L_{ik}\cdot L_{ki}^{T} = \sigma^2\cdot\sum_k L_{ik}^2 \quad [W^2\cdot m^{-4}\cdot nm^{-2}]. \tag{8.93}$$

Applying this formula to the example of table 8-2 using eq. (8.87) we find that, since all rows of the matrix V are identical we need only sum the squares of the column corresponding to the appropriate number of iterations. For $\nu = 3$ iterations this leads to

$$\xi_i^2 = \sigma^2 \cdot \frac{1}{\Delta\lambda^2} \cdot \sum_k \left(V_{k,16}^{(3)}\right)^2 \quad [W^2 \cdot m^{-4} \cdot nm^{-2}] \tag{8.94}$$

or $\xi_i = 0.18\ \sigma$ for any i. Assuming an uncertainty in the S measurements of 0.1% of the maximum or 0.13 we obtain $\xi_i = 0.023$ or about 0.3% of the maximum of E. Once the standard deviations, ξ_i, of the E_i are known, confidence bands can be drawn around the calculated E_i values. For example, the probability is 0.997 that the "true" values (deconvoluted random-error-free data) are within three standard deviations of the calculated values -- between $E_i - 3\xi_i$ and $E_i + 3\xi_i$. (The corresponding probability for two standard deviations is 0.954 and for one standard deviation, 0.683.)

The uncertainties ξ_i are due entirely to the random error in S and do not include the difference between the deconvoluted spectrum and the true spectral irradiance caused, for example, by premature termination of the iteration or smoothing. These additional errors are described by a function we shall call the residual spectral responsivity function. This is the effective responsivity function of the overall process including measurement and deconvolution. If E is the true spectral irradiance vector of a source and R is the responsivity matrix of our spectroradiometer, we will observe a signal S which, even assuming linearity, no error from replacing integration by summation, etc., is not given exactly by

$$S = \Delta\lambda \cdot R \cdot E \quad [S] \tag{8.95}$$

because of the introduction of random experimental noise both in the beam reaching the radiometer and in the measuring process. In following this measurement by a deconvolution we endeavor to reverse the instrumental smearing described by R. As we have noted, even where it is mathematically and numerically possible to solve eq. (8.95) exactly for the vector E, it is rarely desirable to do so because the S we observe is not the quantity which satisfies eq. (8.95) and the discrepancies -- the noise -- will be enormously amplified. Consequently, in deconvolution we actually calculate a quantity[1] $\hat{E}$, which is some approximation to E, by a process

$$\hat{E} = L \cdot S \quad [W \cdot m^{-2} \cdot nm^{-1}] \tag{8.96}$$

which is some approximation to the exact inverse

$$E = \frac{1}{\Delta\lambda} \cdot R^{-1} \cdot S \quad [W \cdot m^{-2} \cdot nm^{-1}]. \tag{8.97}$$

---

[1] In this section we distinguish between a true spectral irradiance E and a calculated approximation to it which we call $\hat{E}$. We have not had to make this distinction previously because it is usually obvious from the context whether E or $\hat{E}$ is intended.

We may think of the matrix $\mathbf{L}$ in eq. (8.96) as an explicit deconvolution matrix as given in eqs. (8.87), (8.88), or (8.89) or we may think of it merely as signifying the process of deconvolution of $\mathbf{S}$ by whatever means -- i.e., as an operator. The degree to which the process $\mathbf{L}$ approximates the exact $\frac{1}{\Delta\lambda}\cdot\mathbf{R}^{-1}$ and consequently the degree to which $\hat{\mathbf{E}}$ approximates $\mathbf{E}$ is dependent indirectly upon the estimated noise level in the data $\mathbf{S}$ because the noise level limits the number of iterations ν or governs the amount of smoothing γ used in the deconvolution process $\mathbf{L}$. We define the residual spectral responsivity function $\mathbf{F}$ as the function which describes the relationship between $\hat{\mathbf{E}}$ and $\mathbf{E}$. That is

$$\hat{\mathbf{E}} = \Delta\lambda\cdot\mathbf{F}\cdot\mathbf{E} \quad [\mathrm{W\cdot m^{-2}\cdot nm^{-1}}]. \tag{8.98}$$

By substituting eq. (8.95) in eq. (8.96) and comparing with eq. (8.98) we see that the matrix $\mathbf{F}$ is given by

$$\mathbf{F} = \mathbf{L}\cdot\mathbf{R} \quad [\mathrm{nm^{-1}}], \tag{8.99}$$

Although $\mathbf{F}$ and $\mathbf{R}$ in this equation are matrices they can be treated column-by-column as vectors. In other words eq. (8.99) can be written as n matrix equations, one for each column of $\mathbf{R}$ and corresponding column of $\mathbf{F}$:

$$\begin{bmatrix} F_{1i} \\ F_{2i} \\ \cdot \\ \cdot \\ \cdot \\ F_{ni} \end{bmatrix} = \mathbf{L}\cdot\begin{bmatrix} R_{1i} \\ R_{2i} \\ \cdot \\ \cdot \\ \cdot \\ R_{ni} \end{bmatrix} [\mathrm{nm^{-1}}], \qquad i=1,2,\ldots,n \tag{8.100}$$

Since these equations are exactly the form of eq. (8.96) -- the deconvolution process -- we see that the columns of the residual responsivity function can be obtained by deconvoluting the corresponding columns of the spectral responsivity matrix as though the columns of the responsivity matrix were observed signals, $\mathbf{S}$. The deconvolution process, $\mathbf{L}$, in eq. (8.100) must be exactly the same matrix or represent the same sequence of calculations which was used in deconvoluting the measurements $\mathbf{S}$ to give the computed spectral irradiance distribution $\hat{\mathbf{E}}$; it must use the same values of any parameters such as γ or A_k and the same number of iterations, if any. The final result, the complete function $\mathbf{F}$, depending upon both λ and λ_o, is the responsivity function of an imaginary spectroradiometer which would yield directly, without deconvolution, a spectral irradiance output identical to the computed results of the real system consisting of radiometer plus deconvolution.

The significance of the residual spectral responsivity function is that it describes the the spectral averaging which remains in the deconvoluted results and which is normally not removable by further deconvolution: it describes the ultimate[1] resolution which can be extracted from the data because of the limits imposed by noise and imprecision in the data itself. Consequently, if it is desired to compare deconvoluted experimental spectral irradiance measurements with, for example, a theory, one should first convolute the theoretical spectrum with the residual responsivity function -- in other words, calculate, using eq. (8.98), what the theoretical spectrum would look like if measured with the imaginary radiometer exhibiting a responsivity function given by the residual responsivity function -- and then compare these results with an appropriate confidence band of the deconvoluted experimental results.

Normally the residual responsivity function is used merely to indicate the improvement in spectral resolution achieved by the deconvolution process. For this purpose usually the calculation of a single representative column of F is adequate. In the limit of a perfect deconvolution where $\mathsf{L} = \frac{1}{\Delta\lambda}\cdot\mathsf{R}^{-1}$ the matrix F becomes

$$\mathsf{F} = \mathsf{L}\cdot\mathsf{R} = \frac{1}{\Delta\lambda}\cdot\mathsf{R}^{-1}\cdot\mathsf{R} = \frac{1}{\Delta\lambda}\cdot\mathsf{I} \quad [\mathrm{nm}^{-1}] \tag{8.101}$$

where I is the identity matrix. Each column of F is then the vector equivalent of a delta function $\delta(\lambda_o-\lambda)$. At the opposite extreme, where no deconvolution at all is carried out, we may take $\hat{\mathsf{E}}$ proportional to S, which is equivalent to setting[2] $\mathsf{L} \propto \mathsf{I}$ [(eq. (8.96)] and we obtain

$$\mathsf{F} = \mathsf{L}\cdot\mathsf{R} \propto \mathsf{I}\cdot\mathsf{R} = \mathsf{R}. \tag{8.102}$$

In this case each column of F is simply proportional to a column of the responsivity matrix R which, in turn, is proportional to the observed response of the spectroradiometer when scanned over a sharp, fixed-wavelength spectral line. Any real deconvolution, then, will lead to a function $F(\lambda_o,\lambda)$ whose dependence upon λ_o at any fixed λ_i lies somewhere between a delta function and the raw instrument distortion of a spectral line. Examples of the calculation of this function are shown in figure 8.4 for the example of table 8-1a. Since $R(\lambda_o,\lambda)$ normally depends strongly upon $\lambda_o-\lambda$ and only weakly (not at all in our example) upon λ or λ_o separately, it is usual to assume a similar behavior for $F(\lambda_o,\lambda)$. Then the abscissa can be labeled $\lambda_o-\lambda$, even though in our example it was calculated using the specific column of the responsivity matrix for $\lambda = 30$. The steady improvement in the overall spectral resolvability of the process as the number of iterations increases is

---

[1]We assume here that the iterations were terminated or the degree of smoothing was established by reference to the noise content of the data. If not, then, presumably, some further resolution enhancement is still possible.

[2]The symbol $\propto$, here, is the symbol of proportionality, and is read, "is proportional to".

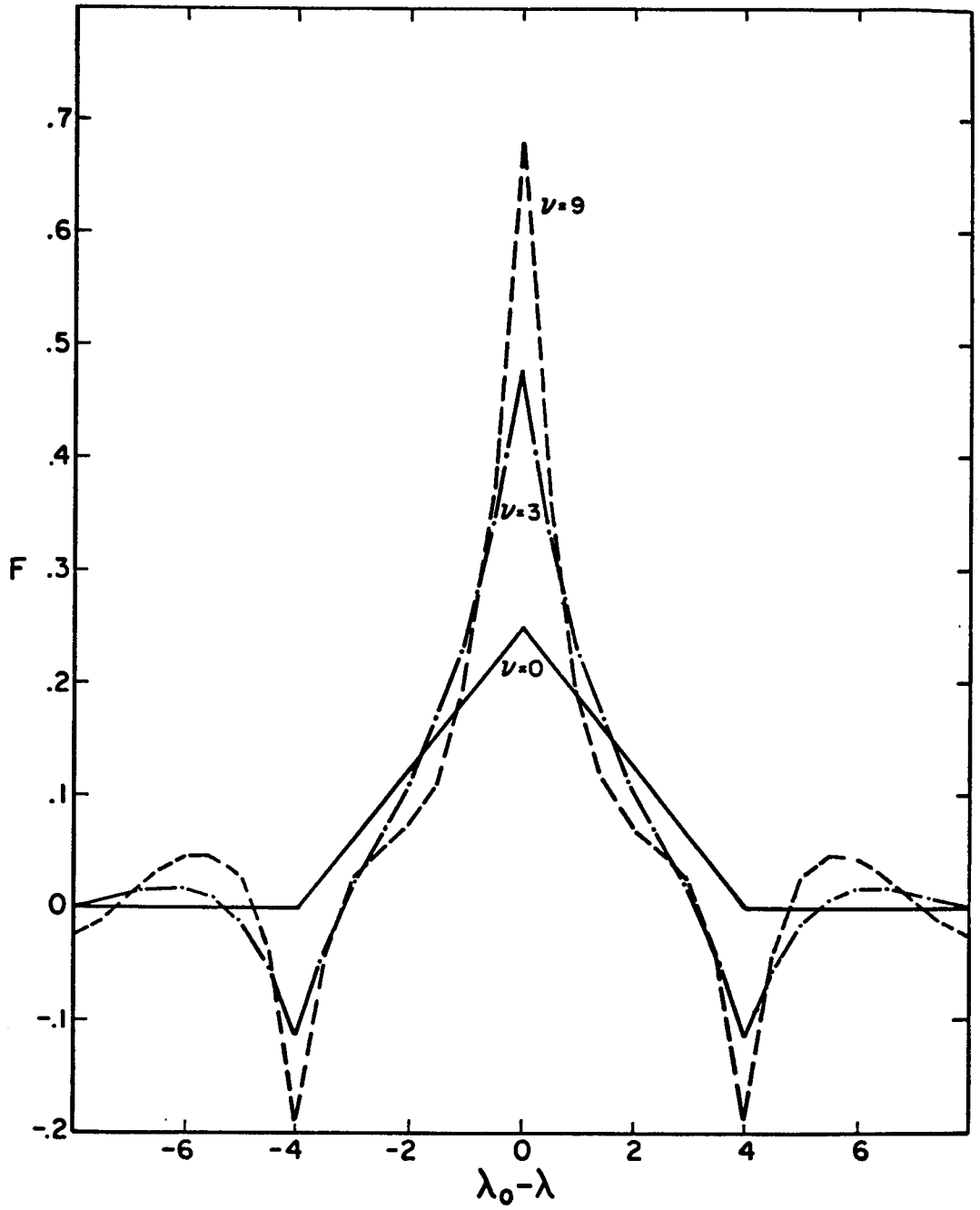


Figure 8.4. Residual responsivity functions for example of table 8-1a. Simple iterative deconvolution using 0, 3, and 9 iterations.

clearly shown in the figure by the sharpening of the central peak with increasing ν. Usually the most important feature of these curves from the point of view of characterizing the resolution improvement due to deconvolution is the full width at half maximum of the curve. In figure 8.4 this is seen to change from 4 at $\nu = 0$ (no deconvolution) to 1.1 at $\nu = 9$ which is roughly equivalent to carrying out the measurements with an instrument whose slit width was nearly a quarter that of the original instrument. Just as the measurement $S(\lambda_o)$ of a source is of limited usefulness unless information about the instrument spectral bandpass is available, so the deconvoluted spectral irradiance $\hat{E}(\lambda)$ is not really complete unless information about the overall spectral resolution is available. Thus, any careful work involving deconvolution should include a computation of the residual spectral responsivity function.

SUMMARY of CHAPTER 8. By the process of deconvolution we mean the recovery of a spectral distribution such as spectral irradiance, $E_\lambda(\lambda)$, from spectroradiometric measurements $S(\lambda_o)$ when the two functions are related by

$$S(\lambda_o) = \int R_E(\lambda_o,\lambda)\cdot E_\lambda(\lambda)\cdot d\lambda \quad [S]. \tag{8.1}$$

In this equation $R_E(\lambda_o,\lambda)$ is the spectral irradiance responsivity function of the measuring spectroradiometer and the variable λ_o is the wavelength setting of the radiometer. The integral is understood to extend over all wavelengths, λ, for which neither $R_E(\lambda_o,\lambda)$ nor $E_\lambda(\lambda)$ vanish. The solution of this integral equation for $E_\lambda(\lambda)$ is complicated by non-uniqueness and, in the case of real measurements, by the amplification of measurement noise.

In any application of eq. (8.1) to n real discrete measurements it is necessary to interpret the integral as a summation, so the starting point of a practical treatment of deconvolution is the set of equations

$$S_k = \sum_{i=1}^{n} R_{ki}\cdot E_i\cdot\Delta\lambda \quad [S], \qquad k = 1,2,\ldots,n, \tag{8.5}$$

where $S_k = S(\lambda_o = \lambda_k)$ [S]

$E_i = E_\lambda(\lambda = \lambda_i)$ [$W\cdot m^{-2}\cdot nm^{-1}$]

and $\Delta\lambda = \lambda_i - \lambda_{i-1}$ [nm] is the spacing between wavelengths, which we shall assume is constant. For the quantities R_{ki} [$S\cdot W^{-1}\cdot m^2$] we use values of $R(\lambda_o,\lambda)$ at $\lambda_o = \lambda_k$ and $\lambda = \lambda_i$, thus passing from eq. (8.1) to eq. (8.5) by the trapezoidal integration approximation. More elaborate integration formulas can still be expressed as eq. (8.5) provided the interpretation of the R_{ki} is slightly modified. As written, eq. (8.5) is a set of n equations in n unknowns and can be solved exactly. Because of the non-uniqueness and noise amplification problems of deconvolution, however, this solution will often prove to be erratic and wholly unacceptable physically. What is needed is an approximate solution of eqs. (8.5) which also meets additional criteria of acceptability such as smoothness.

Probably the most commonly used deconvolution technique in spectroscopy and spectroradiometry is based on requiring the spectral distribution $E_\lambda(\lambda)$ to resemble the distribution $S(\lambda_o)$. This is a simple iterative technique which starts by assuming a trial set of E_i values

$$E_i^{(0)} = S_i/(\Delta\lambda \cdot \sum_j R_{ij}) \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad i = 1,2,\ldots,n, \tag{8.6}$$

which is just the set which would be accepted if no corrections for instrumental spectral broadening were to be made. This set is substituted into eqs. (8.5) and the computed values of S_k are compared with the measured ones to calculate a correction to the trial E_i values. In general, starting with the ν-1th set of E_i values an improved ν^{th} set is computed from

$$E_k^{(\nu)} = E_k^{(\nu-1)} + A_k \cdot \left\{S_k - S_k^{(\nu-1)}\right\} \quad [W \cdot m^{-2} \cdot nm^{-1}], \qquad k = 1,2,\ldots,n, \tag{8.10}$$

where

$$S_k^{(\nu-1)} = \Delta\lambda \cdot \sum_{i=1}^{n} R_{ki} \cdot E_i^{(\nu-1)} \quad [S], \qquad k = 1,2,\ldots,n. \tag{8.11}$$

The A_k, here, are convergence factors whose values only affect the rate of convergence toward the final solution. Frequently these factors are modified by trial-and-error during the course of the iterations. A reasonable starting value is

$$A_k = R_{kk}/(\Delta\lambda \cdot \sum_j R_{kj}^2) \quad [W \cdot m^{-2} \cdot nm^{-1} \cdot S^{-1}] \qquad k = 1,2,\ldots,n. \tag{8.9}$$

The iterations should be stopped when the reconstructed values, $S_k^{(\nu)}$ agree with the measured S_k values within the estimated experimental error. Further iterations tend to make the solution more nearly mathematically exact and less physically acceptable. This simple iterative deconvolution is usually not advocated if more than five or ten iterations are required because of the problem of noise amplification; however, digital smoothing can be added to each iteration stage to permit large numbers of iterations if necessary [8.3, 8.4,8.20].

A useful technique, when the spectral responsivity function can be represented by the slit-scattering function approximation

$$R(\lambda_o,\lambda) = K \cdot z(\lambda_o - \lambda) \cdot r^f(\lambda) \quad [S \cdot W^{-1} \cdot m^2]. \tag{8.22}$$

is to solve eqs. (8.5) for numerical values $S_k = \Delta\lambda \cdot \delta_{k\ell}$, $k = 1,2,\ldots,n$ where ℓ is $n/2$ or $(n-1)/2$, whichever is an integer. The numerical solution which results after ν iterations

$$v_i^{(\nu)} \equiv E_i^{(\nu)} \tag{8.46}$$

is effectively an inverse slit-scattering function with the property that, for any set of real measurements S_k, a ν^{th}-order solution can be calculated from

$$E'^{(\nu)}_{\ell} = \frac{1}{\Delta\lambda}\cdot\sum_{k=1}^{n} v^{(\nu)}_{\ell-k}\cdot S_k \quad [W\cdot m^{-2}\cdot nm^{-1}] \qquad \ell = 1,2,\ldots,n \tag{8.47}$$

and

$$E^{(\nu)}_{\ell} = E'^{(\nu)}_{\ell}/r^f_{\ell}$$

where $v^{(\nu)}_{\ell-k}$ $[S^{-1}\cdot W\cdot m^{-2}]$ is taken to be zero whenever $\ell-k$ is outside the range from 1 to n. There is thus no further need for iterations as long as the slit-scattering function doesn't change.

A different deconvolution technique in which the additional criterion of acceptability is more explicitly related to the smoothness of the result is known by such names as generalized least-squares or constrained least-squares deconvolution. In this technique the E_i are chosen in such a manner as to minimize a sum of two terms, one of which is a measure of the error in reconstructing the measured S_k, and the other of which is some measure of the non-smoothness of the E. For example, in the expression

$$\sum_k (S_k - \Delta\lambda\cdot\sum_i R_{ki}\cdot E_i)^2 + \gamma\cdot\sum_i (E_{i-1}-2E_i+E_{i+1})^2/(\Delta\lambda)^2 \quad [S^2] \tag{8.55}$$

the first term is the sum of the squares of the errors in reconstructing S while the second term is the sum of squares of an approximate measure of the departure from linearity of E in the neighborhood of λ_i. The parameter γ $[S^2\cdot W^{-2}\cdot m^4\cdot nm^4]$ simply sets the relative importance to be attached to these two measures contained in expression (8.55). The minimization of such an expression with respect to the values E_i leads to a solution for the E_i which is most compactly expressed by the matrix equation

$$E = \frac{1}{\Delta\lambda}\cdot\left(R^T\cdot R + \gamma\cdot H\right)^{-1}\cdot R^T\cdot S \quad [W\cdot m^{-2}\cdot nm^{-1}], \tag{8.56}$$

in which R^T $[S\cdot W^{-1}\cdot m^2]$ denotes the matrix transpose of the matrix R and $\left(R^T\cdot R + \gamma\cdot H\right)^{-1}$ denotes the matrix inverse of the matrix $\left(R^T\cdot R + \gamma\cdot H\right)$. The matrix H $[nm^{-4}]$ in eq. (8.56) depends upon the form chosen for the term describing non-smoothness in expression (8.55). The commonly used term shown above leads to

$$\mathbf{H} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & & \\ -2 & 5 & -4 & 1 & 0 & 0 & & \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdot\ \cdot\ \cdot & \\ 0 & 1 & -4 & 6 & -4 & 1 & & \\ & & & \cdot & & & & \\ & & & \cdot & & & & \\ & & & \cdot & & & & \end{bmatrix} \quad [\mathrm{nm}^{-4}], \tag{8.58}$$

but other choices for $\mathbf{H}$ are often used [8.13, 8.21, 8.22, 8.25]. Usually there is little difference in the final solution for $\mathbf{E}$.

The complete numerical procedure for generalized least-squares deconvolution, then, is to evaluate $\mathbf{E}$ from eq. (8.56) for a range of values of γ -- say for $\gamma \approx 0.001R^{*2}$, $0.01R^{*2}$, and $0.1R^{*2}$ where R^* is the largest element of the spectral responsivity matrix $\mathbf{R}$. For each solution the standard deviation of the error in reconstructing the measured signal

$$\sigma(\gamma) = \left\{ \frac{1}{n-1} \cdot \sum_k \left[S_k - \Delta\lambda \cdot \sum_i R_{ki} \cdot E_i(\gamma) \right]^2 \right\}^{1/2} \quad [S] \tag{8.62}$$

should be calculated. The proper value of γ to use is then the largest one (the most smoothing) for which $\sigma(\gamma)$ is not larger than the estimated measurement uncertainty. Generally, additional values of γ outside the range indicated above will have to be tried; however, the solution is not critically dependent upon γ and values within an order-of-magnitude of the optimum one will yield essentially the same set of spectral irradiance values, E_i.

If the number of data points is so large that the matrix operations of eq. (8.56) are not feasible, an iterative technique

$$\mathbf{T} = \mathbf{R}^T \cdot \mathbf{R} + \gamma \cdot \mathbf{H} \quad [S^2 \cdot W^{-2} \cdot m^4], \tag{8.69}$$

$$\mathbf{Q} = \mathbf{R}^T \cdot \mathbf{S} \quad [S^2 \cdot W^{-1} \cdot m^2], \tag{8.70}$$

$$\mathbf{E}^{(\nu)} = \mathbf{E}^{(\nu-1)} + \mathbf{A} \cdot \left\{ \mathbf{Q} - \Delta\lambda \cdot \mathbf{T} \cdot \mathbf{E}^{(\nu-1)} \right\} \quad [W \cdot m^{-2} \cdot nm^{-1}], \tag{8.72}$$

can be applied in which $\mathbf{A}$ $[S^{-2} \cdot W^2 \cdot m^{-4} \cdot nm^{-1}]$ is only an approximate inverse of $\Delta\lambda \cdot \mathbf{T}$. This procedure can be expressed in much the same way as eq. (8.10) and requires little more computer memory than the simple iterative deconvolution.

The dependence of errors in the computed values of E_i upon measurement errors in S_i can be estimated provided the deconvolution calculations which produce $\mathbf{E}$ from $\mathbf{S}$ permit the explicit evaluation of the matrix elements L_{ik} in $\mathbf{L}$ where $\mathbf{L}$ is defined by

$$\mathbf{E} = \mathbf{L} \cdot \mathbf{S} \quad [\mathrm{W \cdot m^{-2} \cdot nm^{-1}}]. \tag{8.86}$$

The matrix $\mathbf{L}$ is an approximation to an inverse responsivity matrix $(1/\Delta\lambda) \cdot \mathbf{R}^{-1}$. Its elements usually can be easily evaluated during deconvolution {see eqs. (8.87), (8.88), and (8.89)} and then

$$\mathrm{cov}(\mathbf{E}) = \mathbf{L} \cdot \mathrm{cov}(\mathbf{S}) \cdot \mathbf{L}^{T} \quad [\mathrm{W^2 \cdot m^{-4} \cdot nm^{-2}}]. \tag{8.90}$$

In this equation $\mathrm{cov}(\mathbf{S})$ is the covariance matrix of the measured S_i -- usually just the diagonal matrix

$$\mathrm{cov}(\mathbf{S}) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & & & \\ 0 & \sigma_2^2 & 0 & . & . & . \\ 0 & 0 & \sigma_3^2 & & & \\ & . & & . & & \\ & . & & & . & \\ & . & & & & \sigma_n^2 \end{bmatrix} \quad [S^2], \tag{8.91}$$

where σ_i is the measured or estimated standard deviation of S_i. The diagonal elements $\left[\mathrm{cov}(\mathbf{E})\right]_{ii}$ of the covariance matrix $\mathrm{cov}(\mathbf{E})$ are likewise the squares of the standard deviations of the computed E_i. The off-diagonal elements of $\mathrm{cov}(\mathbf{E})$ describe how errors at one wavelength are correlated with errors at another wavelength and are generally ignored.

Even if there were no measurement errors, there would still remain, after deconvolution, some spectral smearing of the spectral irradiance distribution. The residual spectral responsivity function is a measure of this residual spectral distortion. This is the computed response, to a monochromatic input beam, of the overall system including measurement and deconvolution. It is obtained by deconvoluting a "signal" numerically equal to the responsivity function, using exactly the same computational steps as the real deconvolution used, including the same values, if any, of A_k and γ, and the same number of iterations. The computed spectral distribution which results is the effective spectral responsivity of the total measurement-and-deconvolution process and normally cannot be improved by further computations; the residual responsivity function represents the best spectral definition which can be achieved in view of the noise content of the measurement data.

References

[8.1] Henry J. Kostkowski and Fred E. Nicodemus, "An Introduction to the Measurement Equation", Chapter 5 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech Note 910-2, 118 pages (Feb. 1978) 58-92.

[8.2] A. L. Khidir and J. C. Decius, "Numerical Methods for the Correction of Apparent Band Shapes due to Finite Slit Width", Spectrochim. Acta 18, (1962) 1629-1639.

[8.3] P. A. Jansson, "Method for Determining the Response Function of a High-Resolution Infrared Spectrometer", J. Opt. Soc. Am. 60, No. 2 (Feb. 1970) 184-191.

[8.4] Peter A. Jansson, Robert H. Hunt, and Earle K. Plyler, "Resolution Enhancement of Spectra", J. Opt. Soc. Am. 60, No. 5 (May 1970) 596-599.

[8.5] S. H. Sheen and J. G. Skofronick, "Deconvolution Methods in Elastic Total Cross Section Measurements", J. Chem. Phys. 61, No. 4 (15 Aug. 1974) 1430-1434.

[8.6] J. S. Rollet and L. A. Higgs, "Correction of Spectroscopic Line Profiles for Instrumental Broadening by a Fourier Analysis Method", Proc. Phys. Soc. 79, (1962) 87-93.

[8.7] G. K. Wertheim, "Deconvolution and Smoothing: Applications in ESCA", J. Electron Spectrosc. Relat. Phenom. 6, (1975) 239-251.

[8.8] Sabri Ergun, "Direct Method for Unfolding Convolution Products -- Its Application to X-ray Scattering Intensities", J. Appl. Cryst. 1, (1968) 19-23.

[8.9] Fukuo Zenitani and Shigeo Minami, "An Analysis of the Iterative Method for Deconvolving Spectroscopic Data Containing a Random Noise", Jap. J. Appl. Phys. 12, No. 3 (Mar. 1973) 379-387.

[8.10] L. D. Skarsgard, H. E. Johns, and L. E. S. Green, "Iterative Response Correction for a Scintillation Spectrometer", Radiation Research 14, (1961) 261-280.

[8.11] A. F. Jones and D. L. Misell, "The Problem of Error in Deconvolution", J. Phys. A 3, (1970) 462-472.

[8.12] P. S. Schoenfeld and J. R. DeVoe, "Statistical and Mathematical Methods in Analytical Chemistry", Anal. Chem. 48, No. 5 (Apr. 1976) 403R-411R.

[8.13] S. Twomey, "Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements" (Elsevier Scientific Pub. Co., New York, 1977).

[8.14] G. Backus and F. Gilbert, "Uniqueness in the Inversion of Inaccurate Gross Earth Data", Phil. Trans. Roy. Soc. Lond., A 266, (Mar. 1970) 123-192.

[8.15] Arend den Harder and Leo de Galan, "Evaluation of a Method for Real-Time Deconvolution", Anal. Chem. 46, No. 11 (Sept. 1974) 1464-1470.

[8.16] B. V. Bronk and W. B. Whitten, "Partial Deconvolution as a Means of Studying the Limitations on Resolution of Spectra Imposed by Noise", Nucl. Instr. and Meth. 106, (1973) 319-331.

[8.17] R. N. Jones, R. Venkataraghavan, and J. W. Hopkins, "The Control of Errors in Infrared Spectrophotometry -- I The Reduction of Finite Spectral Slit Distortion by the Method of 'Pseudo-Deconvolution'", Spectrochim. Acta 23A, (1967) 925-939.

[8.18] John B. Shumaker, "Distribution of Optical Radiation with Respect to Polarization", Chapter 6 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.) Tech Note 910-3, 62 pages (June 1977) 1-46.

[8.19] John B. Shumaker, "Matrix Multiplication", Appendix 6 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.) Tech Note 910-3, 62 pages (June 1977) 47-48.

[8.20] G. Halsey and W. E. Blass, "Deconvolution of IR Spectra in Real Time", Appl. Optics 16, No. 2 (Feb. 1977) 286-288.

[8.21] P. Paatero, S. Manninen, and T. Paakkari, "Deconvolution in Compton Profile Measurements", Philos. Mag. 30, (1974) 1281-1294.

[8.22] S. Twomey, "The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements", J. Franklin Inst. 279, No. 2 (Feb. 1965) 95-109; "On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature", J. Assoc. Comp. Mach. 10, (1963) 97-101.

[8.23] D. Louër, D. Weigel, and R. Louboutin, "Méthode Directe de Correction des Profils de Raies de Diffraction des Rayons X", Acta Cryst. A25, (1969) 335-350.

[8.24] H. Margenau and G. M. Murphy, "The Mathematics of Physics and Chemistry", Vol. 1, 2nd ed. (D. Van Nostrand Co., Inc., New York, 1956).

[8.25] Joel N. Franklin, "Well-Posed Stochastic Extensions of Ill-Posed Linear Problems", J. Math. Anal. and Appl. 31, (1970) 682-716.

[8.26] L. Moore, "Deconvolution of Physical Data", Brit. J. Appl. Phys. (J. Phys. D) (1968) Ser. 2 1, 237-245.

[8.27] H. Schrijver, "Unfolding of Spectra by Fourier Transformation and the Application of a Numerical Filter", Physica 49, (1970) 135-140.

[8.28] Gary Horlick, "Resolution Enhancement of Line Emission Spectra by Deconvolution", Appl. Spectrosc. 26, No. 3 (1972) 395-399.

Appendix 8A. Deconvolution Programs

by John B. Shumaker

In this appendix we present two computer programs for performing deconvolution. The first is the simple iterative deconvolution method of eq. (8.10). The second is the iterative generalized least-squares method of eq. (8.72) with the non-zero elements of a row of the band matrix $\mathbf{A}$ computed by eq. (8.78). They are written in the BASIC language and as far as possible the variable names agree with the notation of Chapter 8. The programs are intended primarily to provide an alternative description of the deconvolution algorithms and a working starting point for programming them. Each program is followed by a cross-reference table of variable names and line numbers. The inefficient use of memory by not writing over program variables when they are no longer needed and especially the data input and output are illustrative rather than practical.

Both programs employ the slit-scattering function approximation and denote by R and E the quantities z and E' of Chapter 8. Thus, if the responsivity function depends upon the instrumental setting λ_o and wavelength λ only through their difference, $\lambda_o-\lambda$, then the program variable R may be taken as the spectral responsivity and the program variable E as the spectral irradiance. The programs also use D, G, and N for the quantities $\Delta\lambda$ (wavelength interval), γ (smoothing parameter), and ν (number of iterations) of Chapter 8.

The number of points in R, E, S and the other vectors is always taken as an odd integer. The slit-scattering function data are assumed to be in the order from top-to-bottom of a column of the responsivity matrix -- as if obtained by observing a fixed wavelength monochromatic source while tuning the radiometer through the spectrum in the same direction as the measured signal data were taken.

All matrix multiplications are performed by the same convolution routine based on eqs. (8.83), (8.84), and (8.85): the convolution of X with Y to form Z assumes that the lengths of all three vectors are odd integers and the indexing is arranged so that the summation which produces the central entry in Z includes the product of the central entries of X and Y. Algebraically, if the lengths of vectors X, Y, and Z are $2x_2+1$, $2y_2+1$, and $2z_2+1$, we have

$$Z_j = \sum_{i=i_1}^{i_2} X_i \cdot Y_{j-i} \qquad j = -z_2,-z_2+1,\ldots,-1,0,1,\ldots,z_2 \qquad (8A.1)$$

where $i_1 = j-y_2$ for $j > y_2-x_2$

otherwise $i_1 = -x_2$

and $i_2 = j + y_2$ for $j < x_2 - y_2$

otherwise $i_2 = x_2$.

In BASIC this translates to

```
DIM X(2*X2+1),Y(2*Y2+1),Z(2*Z2+1)
FOR J=-Z2 TO Z2
 LET Z(J+Z2+1)=0
 LET I1=-X2
 LET I2=X2
 IF J>Y2-X2 LET I1=J-Y2
 IF J>X2-Y2 LET I2=J+Y2
 FOR I=I1 TO I2
   LET Z(J+Z2+1)=Z(J+Z2+1)+X(I+X2+1)*Y(J+Y2+1-I)
 NEXT I
NEXT J
```

In iterative calculations, using convolution instead of matrix multiplication, there is always a question of how to avoid increasing the vector lengths with every iteration. The increases come about because the number of significant elements of Z which can be calculated by the convolution of X with Y is approximately equal to the sum of the numbers of significant elements in X and Y. In these programs the following scheme is used: in calculating the quantity[1] W = S - D*R*E (or W = Q - D*T*E in the second program) the lengths of W and S (or Q) are made large enough to accommodate all significant elements in the convolution of R (or T) with E. In computing the improved E by convolution of A with W, only the desired elements are retained.

The programs print out the complete set of E values after each iteration followed by the rms value of the error in reconstructing the S values from this set of E values. If end-effects (see Chapter 8) are important the rms-error calculation should be restricted to the central wavelength region which is free of this problem. In the simple iterative deconvolution the iterations should be terminated when the rms error reaches a minimum or becomes comparable with the experimental error. In the generalized least-squares program the iterations should be continued well beyond the point where the improvement per iteration is comparable with experimental accuracy and then the rms error can be used to judge which value of the smoothing constant G is most suitable.

---

[1]The asterisk stands for the convolution operation. That is, $R*E \equiv \sum_i R_{j-i} \cdot E_i$, or the equivalent integral. This notation is commonly used in the literature, probably because it suggests matrix multiplication, which is what convolution becomes when the slit-scattering function approximation is inappropriate and the full matrices must be used.

```
0010 REM A PROGRAM FOR SIMPLE ITERATIVE DECONVOLUTION
0020 PRINT "WAVELENGTH SPACING";
0030 INPUT D
0040 PRINT
0050 REM READ SLIT FUNCTION DATA
0060 REM FIRST ENTRY IN DATA BLOCK IS # OF DATA POINTS TO FOLLOW.
0070 REM ORDER OF SLIT-FUNCTION DATA IS AS IF A SHARP LINE
0080 REM WERE SCANNED EXACTLY AS NORMAL MEASURED SIGNAL IS SCANNED.
0090 DATA 3
0100 DATA 2,4,2
0110 PRINT
0120 READ R9
0130 LET R2=(R9-1)/2
0140 DIM R(R9)
0150 FOR I=1 TO R9
0160  READ R(I)
0170 NEXT I
0180 REM READ MEASURED SIGNAL DATA
0190 DATA 31
0200 DATA 0,.333333,5.33333,25.6667,64,102.333,122.667
0210 DATA 127.667,128,128,128,128,128,128,128
0220 DATA 128,128,128,128,128,128,128,128,128,128
0230 DATA 128,128,128,128,128,128
0240 READ E9
0250 LET S9=E9+R9-1
0260 DIM S(S9),E(E9),W(S9)
0270 FOR I=1 TO S9
0280  LET S(I)=0
0290 NEXT I
0300 FOR I=1 TO E9
0310  READ S(I+R2)
0320 NEXT I
0330 LET R0=0
0340 LET R1=0
0350 FOR I=1 TO R9
0360  LET R0=R0+R(I)
0370  LET R1=R1+R(I)^2
0380 NEXT I
0390 LET A=R(R2+1)/R1/D
0400 LET S2=(S9-1)/2
0410 LET E2=(E9-1)/2
0420 REM CALC ZEROTH ESTIMATE OF E
0430 FOR J=-E2 TO E2
0440  LET E(J+E2+1)=S(J+S2+1)/R0/D
0450 NEXT J
0460 REM CONVOLUTE R WITH E AND CALC W=S-D*R*E
0470 LET X=0
0480 FOR J=-S2 TO S2
0490  LET W(J+S2+1)=S(J+S2+1)
0500  LET I1=-R2
0510  LET N=0
0520  LET I2=R2
0530  IF J>E2-R2 LET I1=J-E2
0540  IF J<R2-E2 LET I2=J+E2
0550  FOR I=I1 TO I2
0560    LET W(J+S2+1)=W(J+S2+1)-R(I+R2+1)*E(J+E2-I+1)*D
0570  NEXT I
0580  LET X=X+W(J+S2+1)^2
0590 NEXT J
0600 REM CALC E=E+A*W
```

```
0610 FOR J=-E2 TO E2
0620  LET E(J+E2+1)=E(J+E2+1)+A*W(J+S2+1)
0630  PRINT E(J+E2+1),
0640 NEXT J
0650 PRINT
0660 PRINT SQR(X/E9)
0670 LET N=N+1
0680 IF N<11 GOTO 0460
0690 END
```

SID 2/ 6/79

A	0390	0620													
D	0030	0390	0440	0560											
E	0260	0440	0560	0620	0630										
E2	0410	0430	0440	0530	0540	0560	0610	0620	0630						
E9	0240	0250	0260	0300	0410	0660									
I	0150	0160	0170	0270	0280	0290	0300	0310	0320	0350	0360	0370	0380	0550	0560
	0570														
I1	0500	0530	0550												
I2	0520	0540	0550												
J	0430	0440	0450	0480	0490	0530	0540	0560	0580	0590	0610	0620	0630	0640	
N	0510	0670	0680												
R	0140	0160	0360	0370	0390	0560									
R0	0330	0360	0440												
R1	0340	0370	0390												
R2	0130	0310	0390	0500	0520	0530	0540	0560							
R9	0120	0130	0140	0150	0250	0350									
S	0260	0280	0310	0440	0490										
S2	0400	0440	0480	0490	0560	0580	0620								
S9	0250	0260	0270	0400											
W	0260	0490	0560	0580	0620										
X	0470	0580	0660												
0460	0680														

```
0010 REM A PROGRAM FOR ITERATIVE GENERALIZED LEAST-SQUARES DECONVOLUTION
0020 PRINT "WAVELENGTH INTERVAL";
0030 INPUT D
0040 PRINT
0050 PRINT "GAMMA";
0060 INPUT G
0070 PRINT
0080 REM FOR MORE ACCURATE A, INCREASE A1
0090 LET A1=0
0100 REM READ SLIT FUNCTION
0110 REM FIRST ENTRY IN DATA BLOCK IS # OF DATA POINTS TO FOLLOW
0120 REM ORDER OF SLIT-FUNCTION DATA IS AS IF A SHARP LINE
0130 REM WERE SCANNED EXACTLY AS NORMAL MEASURED SIGNAL IS SCANNED.
0140 DATA 3
0150 DATA 2,4,2
0160 READ R9
0170 LET T2=R9-1
0180 LET R2=(R9-1)/2
0190 DIM R(R9),T(2*T2+1)
0200 LET R0=0
0210 FOR I=1 TO R9
0220  READ R(I)
0230  LET R0=R0+R(I)
0240 NEXT I
0250 REM CONVOLUTE R WITH R-TRANSPOSE TO GIVE T
0260 FOR J=1-R9 TO R9-1
0270  LET T(J+R9)=0
0280  LET I1=-R2
0290  LET I2=R2
0300  IF J>0 LET I1=J-R2
0310  IF J<0 LET I2=J+R2
0320  FOR I=I1 TO I2
0330    LET T(J+R9)=T(J+R9)+R(R2-I+1)*R(J+R2-I+1)
0340  NEXT I
0350 NEXT J
0360 REM ADD IN SMOOTHING TERM  G*H
0370 LET T(R9)=T(R9)+6*G
0380 LET T(R9+1)=T(R9+1)-4*G
0390 LET T(R9-1)=T(R9-1)-4*G
0400 LET T(R9+2)=T(R9+2)+G
0410 LET T(R9-2)=T(R9-2)+G
0420 REM COMPUTE APPROX INVERSE BY LEAST-SQUARES
0430 LET A9=R9+2*INT(A1)
0440 DIM B(A9,A9),A(A9,1),Z(A9,1)
0450 LET A2=(A9-1)/2
0460 REM CONVOLUTE T-TRANSPOSE WITH T AND FILL OUT MATRIX
0470 MAT B=ZER
0480 FOR J=-A9 TO A9
0490  IF ABS(J)>A2 GOTO 0520
0500  LET Z(J+A2+1,1)=0
0510  IF J+R9>0 IF J+R9<=2*T2+1 LET Z(J+A2+1,1)=T(J+R9)/D
0520  LET B0=0
0530  LET I1=-T2
0540  LET I2=T2
0550  IF J>0 LET I1=J-T2
0560  IF J<0 LET I2=J+T2
0570  FOR I=I1 TO I2
0580    LET B0=B0+T(T2-I+1)*T(J+T2-I+1)
0590  NEXT I
0600  FOR L=-A2 TO A2
```

```
0610     LET B1=J+L+A2+1
0620     IF B1>0 IF B1<=A9 LET B(L+A2+1,B1)=B0
0630   NEXT L
0640  NEXT J
0650  MAT B=INV(B)
0660  MAT A=B*Z
C670  REM READ SIGNAL DATA
0680  DATA 31
0690  DATA 0,.333333,5.33333,25.6667,64,102.333,122.667
0700  DATA 127.667,128,128,128,128,128,128,128
0710  DATA 128,128,128,128,128,128,128,128,128,128
0720  DATA 128,128,128,128,128,128
C730  READ S9
0740  DIM S(S9),E(S9)
0750  FOR I=1 TO S9
0760   READ S(I)
0770  NEXT I
0780  LET Q9=2*R9+S9-2
0790  REM CONVOLUTE R-TRANSPOSE WITH S TO GIVE Q
0800  DIM Q(Q9),W(Q9)
0810  LET Q2=(Q9-1)/2
0820  LET S2=(S9-1)/2
0830  FOR J=-Q2 TO Q2
0840   LET Q(J+Q2+1)=0
0850   LET I1=-R2
0860   LET I2=R2
0870   IF J>S2-R2 LET I1=J-S2
0880   IF J<R2-S2 LET I2=J+S2
0890   FOR I=I1 TO I2
0900     LET Q(J+Q2+1)=Q(J+Q2+1)+R(R2-I+1)*S(J+S2-I+1)
0910   NEXT I
0920  NEXT J
0930  REM CALC ZEROTH ESTIMATE OF E
0940  FOR I=1 TO S9
0950   LET E(I)=S(I)/R0/D
0960  NEXT I
0970  LET N=0
0980  REM CONVOLUTE T WITH E AND CALC W=Q-D*T*E
0990  FOR J=-Q2 TO Q2
1000   LET K=J+Q2+1
1010   LET W(K)=Q(K)
1020   LET I1=-T2
1030   LET I2=T2
1040   IF J>S2-T2 LET I1=J-S2
1050   IF J<T2-S2 LET I2=J+S2
1060   FOR I=I1 TO I2
1070     LET W(K)=W(K)-T(I+T2+1)*E(J+S2-I+1)*D
1080   NEXT I
1090  NEXT J
1100  REM CONVOLUTE A WITH W AND CALC E=E+A*W
1110  FOR J=-S2 TO S2
1120   LET I1=-A2
1130   LET I2=A2
1140   IF J>Q2-A2 LET I1=J-Q2
1150   IF J<A2-Q2 LET I2=J+Q2
1160   FOR I=I1 TO I2
1170     LET E(J+S2+1)=E(J+S2+1)+A(A2-I+1,1)*W(J+Q2-I+1)
1180   NEXT I
1190   PRINT E(J+S2+1),
1200  NEXT J
1210  PRINT
1220  REM CONVOLUTE R WITH E AND CALC ERRORS:  S-D*R*E
```

```
1230 LET X=0
1240 FOR J=-S2 TO S2
1250  LET X0=S(J+S2+1)
1260  LET I1=-R2
1270  LET I2=R2
1280  IF J>S2-R2 LET I1=J-S2
1290  IF J<R2-S2 LET I2=J+S2
1300  FOR I=I1 TO I2
1310    LET X0=X0-R(I+R2+1)*E(J+S2-I+1)*D
1320  NEXT I
1330  LET X=X+X0^2
1340 NEXT J
1350 PRINT SQR(X/S9)
1360 LET N=N+1
1370 IF N<10 GOTO 0980
1380 END
```

IGLSD 2/ 6/79

A	0440	0660	1170												
A1	0090	0430													
A2	0450	0490	0500	0510	0600	0610	0620	1120	1130	1140	1150	1170			
A9	0430	0440	0450	0480	0620										
B	0440	0470	0620	0650	0660										
B0	0520	0580	0620												
B1	0610	0620													
D	0030	0510	0950	1070	1310										
E	0740	0950	1070	1170	1190	1310									
G	0060	0370	0380	0390	0400	0410									
I	0210	0220	0230	0240	0320	0330	0340	0570	0580	0590	0750	0760	0770	0890	0900
	0910	0940	0950	0960	1060	1070	1080	1160	1170	1180	1300	1310	1320		
I1	0280	0300	0320	0530	0550	0570	0850	0870	0890	1020	1040	1060	1120	1140	1160
	1260	1280	1300												
I2	0290	0310	0320	0540	0560	0570	0860	0880	0890	1030	1050	1060	1130	1150	1160
	1270	1290	1300												
J	0260	0270	0300	0310	0330	0350	0480	0490	0500	0510	0550	0560	0580	0610	0640
	0830	0840	0870	0880	0900	0920	0990	1000	1040	1050	1070	1090	1110	1140	1150
	1170	1190	1200	1240	1250	1280	1290	1310	1340						
K	1000	1010	1070												
L	0600	0610	0620	0630											
N	0970	1360	1370												
Q	0800	0840	0900	1010											
Q2	0810	0830	0840	0900	0990	1000	1140	1150	1170						
Q9	0780	0800	0810												
R	0190	0220	0230	0330	0900	1310									
R0	0200	0230	0950												
R2	0180	0280	0290	0300	0310	0330	0850	0860	0870	0880	0900	1260	1270	1280	1290
	1310														
R9	0160	0170	0180	0190	0210	0260	0270	0330	0370	0380	0390	0400	0410	0430	0510
	0780														
S	0740	0760	0900	0950	1250										
S2	0820	0870	0880	0900	1040	1050	1070	1110	1170	1190	1240	1250	1280	1290	1310
S9	0730	0740	0750	0780	0820	0940	1350								
T	0190	0270	0330	0370	0380	0390	0400	0410	0510	0580	1070				
T2	0170	0190	0510	0530	0540	0550	0560	0580	1020	1030	1040	1050	1070		
W	0800	1010	1070	1170											
X	1230	1330	1350												
X0	1250	1310	1330												
Z	0440	0500	0510	0660											
0520	0490														
0980	1370														

Chapter 9. Physically Defining Measurement-Beam Geometry by Using Opaque Barriers

by Fred E. Nicodemus

In this CHAPTER. In previous discussions of the distribution of optical radiation with respect to position and direction (Chapter 2 [9.1][1] and Chapter 4 [9.2]), we merely assumed that opaque barriers, such as the opaque screens in figure 2.2 [9.1], could be used to define beams by blocking all rays that are not a part of the desired beam. Now we want to examine more closely the way in which such opaque barriers are used to physically define or limit the extent of radiation beams in making radiometric measurements. We'll do this mostly in terms of throughput, the integrated product of projected area and solid angle (or area and projected solid angle), and the limits of the throughput integral(s), because they are the quantitative measures of beam geometry. Even though exact numerical evaluation of these quantities may not be needed very often, it is only in these terms that an adequate understanding of their significance can be achieved, including the ability to confidently decide, in all cases, whether or not the numerical evaluation is required.

The geometrical-optics or ray-optics limits provided by opaque barriers are the starting point for assigning the limits of integration for the spatial parameters of area and solid angle (position and direction) in the measurement equation {e.g., eq. (5.5) [9.3]}. When accuracy requirements are not too stringent, these limits may be adequate without refinement. However, in most instances, the effects of scattering, diffraction, or optical-system aberrations (imperfect imaging) may extend the responsivity of a radiometer substantially beyond the nominal limits of the ray-optics beam boundaries. Hence, for optimum accuracy, the integrals of the measurement equation must likewise be extended beyond those nominal limits, to the extent determined by experimental checks of instrument responsivity as a function of position and direction.[2] The opaque-obstacle-shadow boundaries that we discuss in this chapter, then, while they are very important for design purposes and as a starting point for analysis, are no more than that -- a starting point or first approximation -- and must never be considered as absolute limits that cannot be exceeded. With that caveat, however, we'll devote the rest of this chapter, except for an occasional reminder about the real world, to the simpler domain of ideal geometrical optics with sharp ray-defined shadows and perfect images in clear non-scattering media.

The chapter starts with a discussion of the simplest physical definition of a radiation beam -- by two apertures in opaque screens in a homogeneous (iso-refractive-index) medium where all rays are straight lines. The concept of vignetting is also illustrated for this simple case.

---

[1]Figures in square brackets indicate literature references listed at the end of *this chapter*.

[2]Another approach, and in some ways a sounder approach, is to always extend the integration limits as far as possible -- e.g., $0 \leq \theta \leq \pi/2$ [rad] and $0 \leq \phi \leq 2\pi$ [rad] -- even though the integrand may vanish over much of each interval.

There follows a discussion of the beam defined by two apertures with a thin lens between them. This brings out the way in which aperture and/or aperture-image pairs can be equivalent to the earlier simple case of two apertures defining a beam of straight-line rays in a homogeneous medium. The concepts of object space and image space are developed, bringing out that both members of a beam-defining aperture/aperture-image pair exist in the same "space".

The geometrical-optics concepts of aperture stop and field stop, and their images in object space, the entrance pupil and entrance window, respectively, are developed. This is followed by a discussion of the somewhat "loose" way in which these terms are used in radiometry. The exit pupil and exit window, the images of the respective stops in image space are also briefly defined. In addition, the special case of an optical system focused on infinity (entrance window at infinity) is treated.

Some practical problems relating to the role of entrance pupils as receiving apertures in radiometric measurements are discussed, followed by a discussion of diffusers and their use in connection with the measurement of irradiance. Attention is given to the problem of the exact position of the entrance pupil (receiving aperture) and the effects of diffusers empirically shaped to achieve a "cosine response" characteristic.

Finally, there are brief discussions of the role of baffles, of some more aspects of vignetting, and of the effects of beam geometry on overall instrument radiance and irradiance responsivities.

BEAM GEOMETRY in RADIOMETRY. In Chapter 5 of this Manual, we found [9.3 (p.90)] that the measurement equation has basically the same form for all radiometric measurements {eq. (5.30b)}

$$S(A,\omega,\Delta\lambda) = \int_{\Delta\lambda}\int_{A}\int_{\omega} R_{\Phi}(x,y,\theta,\phi,\lambda)\cdot L_{\lambda}(x,y,\theta,\phi,\lambda)\cdot\cos\theta\cdot d\omega\cdot dA\cdot d\lambda \quad [S]. \tag{9.1}$$

Here $S(A,\omega,\Delta\lambda)$ [S] is the output signal of an instrument with spectral-ray flux responsivity $R_{\Phi}(x,y,\theta,\phi,\lambda)$ $[S\cdot W^{-1}]$ for the incident flux element

$$d\Phi(x,y,\theta,\phi,\lambda) \equiv L_{\lambda}(x,y,\theta,\phi,\lambda)\cdot\cos\theta\cdot d\omega\cdot dA\cdot d\omega \quad [W]$$

associated with the ray, of spectral radiance $L_{\lambda}(x,y,\theta,\phi,\lambda)$ $[W\cdot m^{-2}\cdot sr^{-1}\cdot nm^{-1}]$, incident on the receiving-aperture-area element dA $[m^2]$ at the point x,y, within the solid-angle element $d\omega$ [sr] from the direction θ,ϕ, and within the spectral-wavelength element $d\lambda$ [nm] at the wavelength λ. The effects of the radiation parameters of time or frequency-f (of fluctuation or modulation; $f << \nu$), and of polarization (see Chapter 6 [9.4]) have been ignored here and, for simplicity, we will continue to ignore them in this chapter while we examine the way in which measurement-beam geometry is physically defined. We'll look at the physical considerations that determine the values of such quantities as the spatial limits of integration, the area A and the solid angle ω [actually, $\omega(x,y)$], in eq. (9.1) that define the beam and its throughput ($\Theta \equiv \iint\cos\theta\cdot d\omega\cdot dA$, see next section).

We'll discuss the choice of reference surface to facilitate the beam (throughput) definition and will emphasize the purely ray-optics or geometrical-optics considerations.

Even when the physical limits that define the throughput are clearly and unambiguously identified, accurate measurements to establish the values of the area(s) and solid angle(s) involved can be quite difficult. However, we won't go into such precision-measurement problems in this chapter but will concern ourselves primarily with the ways in which the physical limits are established.

Just because these measurements are so difficult, we try to avoid the need for accurate evaluation of the throughput wherever possible in making radiometric measurements, particularly when the best accuracy is required. That's why, in all three illustrative problems in Chapter 5 [9.3], calibration measurements against a standard source are made under conditions that are enough like those of the actual measurement so that all of the geometrical factors cancel and do not appear in the final expression for the measured quantity. However, a thorough understanding of the physical limits is needed: (1) to be sure that the beam geometry is, in fact, not significantly changed between the calibration measurement of a standard and the corresponding measurement of an unknown radiometric quantity, (2) to insure adequate specification of the measurement beam in any complete report of a radiometric measurement, (3) to evaluate the adequacy of throughput estimates used in planning a measurement, particularly in establishing the required detector capability, and (4) as an essential factor in an error analysis to estimate the uncertainty of the results of a radiometric measurement.

SIMPLE BEAM-DEFINING APERTURES. In the "thought experiment" to develop the concept of radiance in Chapter 2 (figure 2.2 [9.1]) we introduced one of the simplest beam-defining aperture configurations--a beam of all rays through a pair of apertures, one in each of two opaque screens. With no other obstructions from source to receiver, the selected beam is clearly and unambiguously defined *at either aperture* as consisting of *all rays through every point of that aperture within the solid angle subtended at each point by the other aperture*. If we need to evaluate the throughput Θ of the aperture pair filled by the beam {see eq. (2.27) [9.1]}, we can choose either aperture (plane) as the reference surface (see figure 9.1) and evaluate

$$\Theta = \int_A \int_{\omega(x,y)} \cos\theta \cdot d\omega \cdot dA \ [m^2 \cdot sr], \tag{9.2}$$

where

A $[m^2]$ is the reference-surface aperture area (includes the intersection of every ray of the beam with the reference surface),

$\omega(x,y)$ [sr] is the solid angle subtended by the other aperture A' at the point x,y within the aperture area A (includes every ray of the beam that intersects that point),

dA $[m^2]$ is an element of reference-surface area at the point x,y [$dA = dx \cdot dy$],

θ [rad] is the polar angle from the normal to dA at x,y that, together with the azimuth angle ϕ [rad] from the X-axis in the reference-surface plane, constitute spherical co-ordinates about x,y, and

$d\omega$ [sr] is an element of solid angle in the direction θ,ϕ from x,y (within ω); [$d\omega = \sin\theta \cdot d\theta \cdot d\phi$].

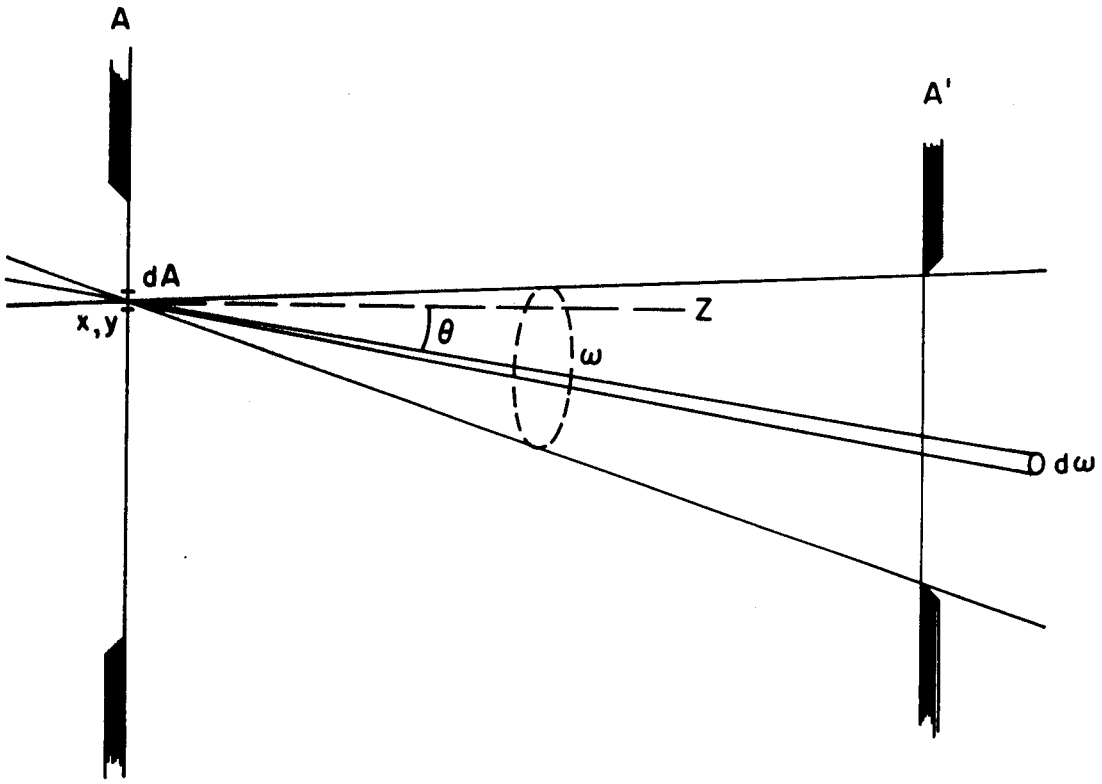


Figure 9.1. Evaluation of the throughput between two apertures.
The beam consists of all rays (straight lines) through the apertures A (chosen as the reference surface) and A'. Its throughput Θ is evaluated by the double integral of eq. (9.2), integrating or summing over all solid-angle elements $d\omega$ within ω, the solid angle subtended by A' at each point x,y of A, and over all area elements dA within A. Each solid-angle element $d\omega$ makes an angle θ with the normal to the area element dA at the point x,y.

VIGNETTING. Away from the two aperture screens of figure 9.1, across any other plane that intersects the entire beam, there is vignetting so that the solid angle $\omega(x,y)$ at some points x,y in that plane is not defined by just one of the beam-defining apertures. This situation, and the meaning of the term "vignetting", is best explained by reference to figures 9.2 and 9.3. Figure 9.2 shows a plane section through the centers of two parallel coaxial circular apertures, A_1 and A_2. The extreme rays a - a through corresponding points, together with extreme rays b - b through opposite points, bound different regions of the defined beam (made up of all rays that pass through both apertures). When a point x,y of an intersecting plane, such as the plane c - c, lies within one of the central regions, (2), (4), or (6), there is no vignetting because the solid angle $\omega(x,y)$ filled by all rays of the beam that intersect at x,y is a solid angle subtended at that point by just one of the apertures. In regions (2) [as shown] and (6) this is aperture A_2; in region (4) it is A_1. However, except in the planes of the apertures, every such intersecting plane c - c also crosses regions (1), (3), or (5), where things are different. Consider the point x,y on the intersecting plane c - c in figure 9.3, where it is shown in region (5) with the extreme rays bounding the solid angle $\omega(x,y)$: the uppermost ray through the apertures is limited by aperture A_2 while the lowest one is limited by aperture A_1. Thus the rays of the beam that intersect at this particular point x,y do not completely fill either of the apertures, a condition called "vignetting". It is easily seen that the same holds true for all points in the cross-hatched regions, (1), (3), and (5). When vignetting occurs, as described here, it becomes extremely difficult to specify the limits of the solid angle $\omega(x,y)$ in order to evaluate the throughput by eq. (9.2). A good example of the complexities in a case involving just a small amount of relatively simple vignetting is found in [9.5]. Thus, there is a clear advantage, when defining or analyzing any beam and its parameters, to choosing one of the two aperture-screen planes, where there is no vignetting, as the reference surface.

A BEAM-DEFINING APERTURE PAIR. Similar vignetting complications arise if any of the rays from source to receiver, that would otherwise pass through both apertures, are blocked by any obstruction or obstacle at a third location, whether between the apertures or external to them.[1] Accordingly, in the design of instruments and experimental configurations, especially when the highest accuracy is needed, and particularly when it is necessary to compute the value of the throughput, it is desirable to arrange to have a beam defined by just two apertures with no other obstructions to cause vignetting. We'll designate such an unvignetted pair of apertures as a beam-defining aperture pair or, most of the time, as just an aperture pair.

So far, we've looked only at the situation where both apertures and all of the space between them are in the same uniform, isotropic medium so that all rays between them are straight lines. We intend that those conditions be implied by the term aperture pair, except when explicitly stated otherwise. However, there are many optical systems where

[1]See "More on Vignetting" later in this chapter, especially figure 9.8.

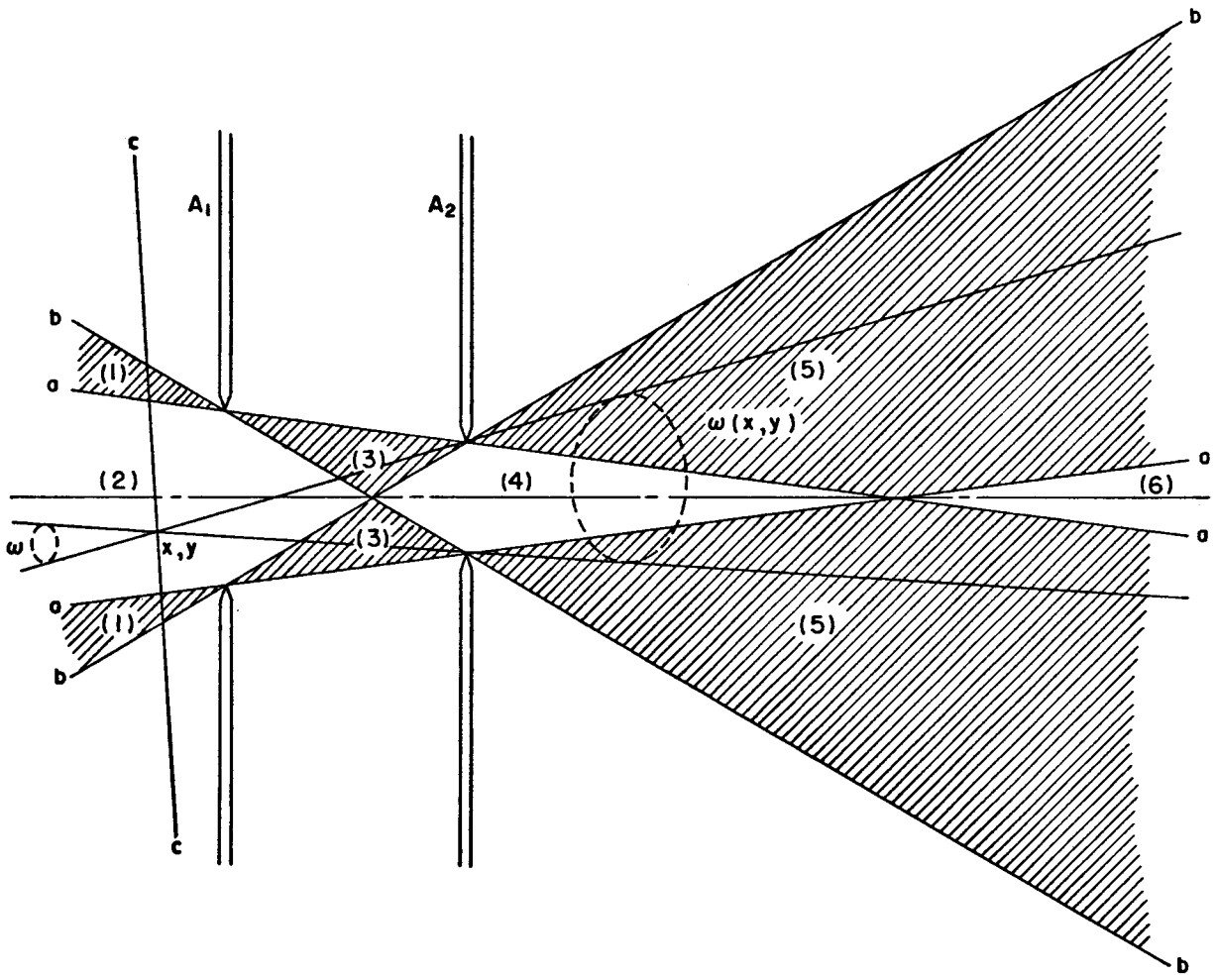


Figure 9.2. Vignetting regions in a beam between two circular apertures.
Vignetting occurs in the cross-hatched regions, (1), (3), and (5), but not in the clear central regions, (2), (4), and (6). At the point x,y on plane c - c, in clear region (2), intersecting rays are limited to the solid angle $\omega(x,y)$ subtended at that point by the aperture A_2.
[See also figure 9.3.]

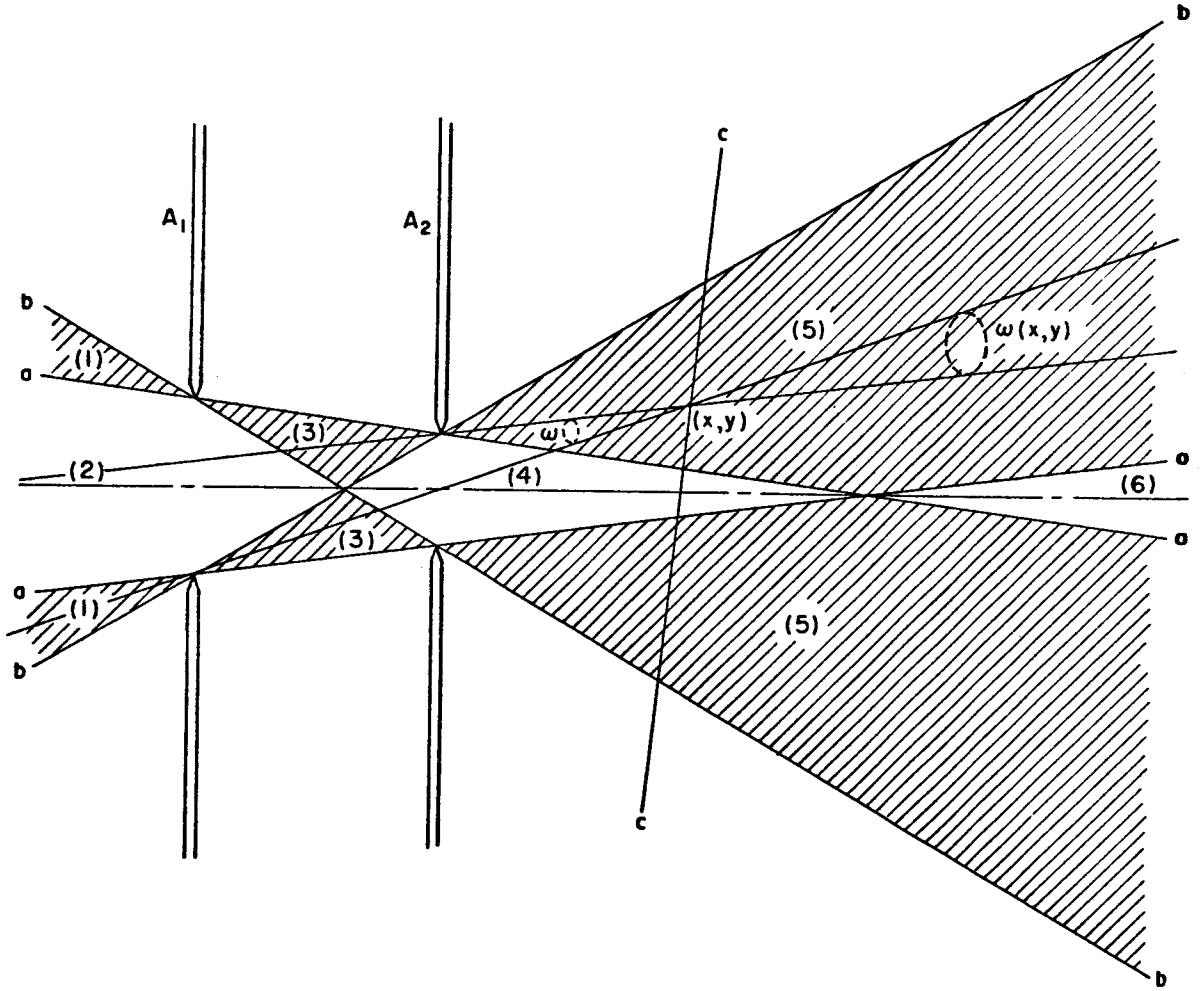


Figure 9.3. An example of vignetting in a beam between two circular apertures. At the point x,y on plane c - c, in the cross-hatched vignetting region (5), note that the upper extreme ray is limited by aperture A_2 but the lower one is limited by aperture A_1, so the solid angle $\omega(x,y)$ of rays of the beam intersecting at x,y does not fill either of the apertures.

there are focusing optical elements that deflect the rays between two beam-defining apertures. In the next section, we analyze such a case. We find that it is always possible, for purposes of analysis and throughput computation, to replace any such pair of beam-defining apertures that has intervening focusing optical elements by an equivalent aperture/aperture-image pair that has no focusing elements so that all rays between this equivalent pair are straight lines.

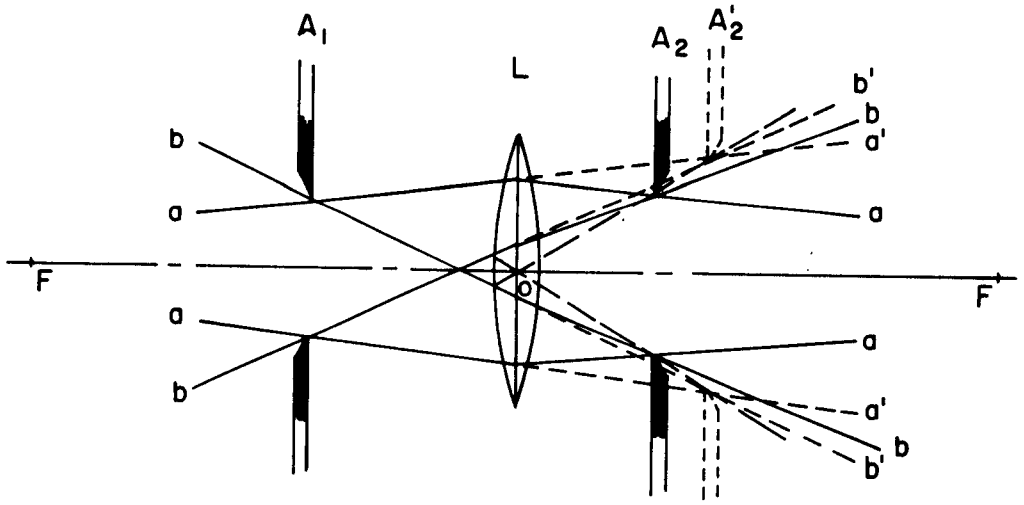
A THIN LENS between TWO APERTURES: BEAM-DEFINING APERTURE/APERTURE-IMAGE PAIRS.

Figure 9.4 depicts two vertical screens with coaxial circular apertures A_1 and A_2 aligned along the horizontal optical axis of a thin circular lens L located between them. In (a), the extreme rays between the margins of the two apertures are again shown by solid lines a - a and b - b, as in figure 9.2, but now they are deflected by refraction as they pass through the lens. On the right-hand side of the lens, the rays from the left-hand side have also been extended through the lens as undeviated straight dashed lines that intersect at points on the margin of a virtual image[1] A_2' (also shown by dashed lines) corresponding to similar points of intersection of the solid lines at the margin of aperture A_2.

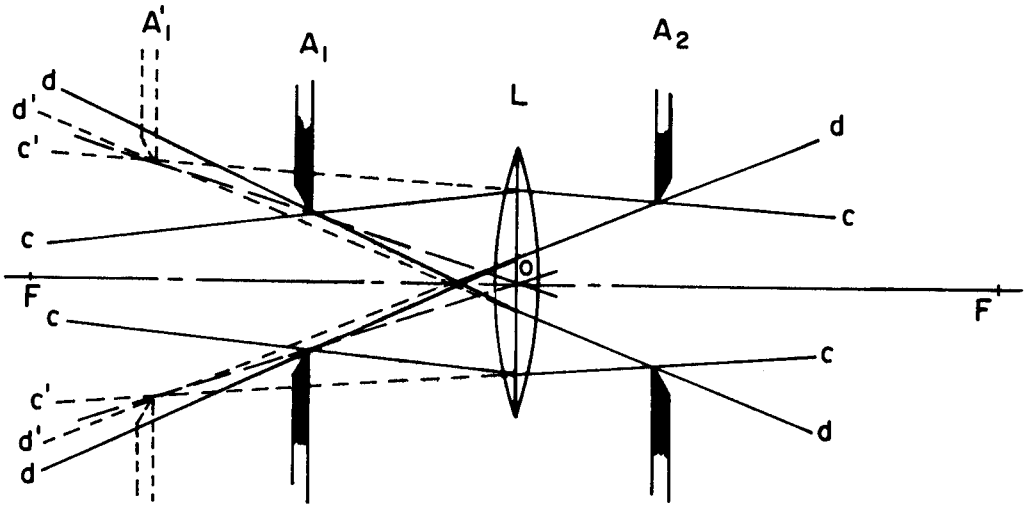
It is clear that the rays of the beam that cross at aperture A_1 are exactly the same there at A_1 as they would be if the lens were removed and the aperture A_2 were replaced by its virtual image A_2'. That is, the rays between A_1 and A_2 that intersect at A_1 to form the subtended solid angles $\omega_1(x_1,y_1)$ at points x_1,y_1 within the aperture A_1 in the expression for the throughput of the aperture-lens combination

$$\Theta = \int_{A_1} \int_{\omega_1(x_1,y_1)} \cos\theta_1 \cdot d\omega_1 \cdot dA_1 \ [m^2 \cdot sr], \tag{9.8}$$

[1]A real image is formed when rays from each object point converge, after deviation by a focusing mirror or lens, to intersect at a corresponding image point. If strong enough visible light is used, such a real image can be seen on a ground glass or other scattering or diffusing material placed at the location of the image. A virtual image is formed when the rays diverge, after deviation by the lens or mirror, and intersect only when extended back, as straight lines, again past the deflecting mirror or lens. Since the intersecting lines at each virtual image point are only extensions of the rays, which do not actually pass through that point, they cannot be seen there on a ground glass like those of a real image. Instead, for example, the virtual image A_2' of the aperture A_2 would be seen, and appear to be located at the position of A_2', only by looking back through the lens from the left-hand side at the rays coming through it from the real object, the aperture A_2. In the vicinity of A_1, looking back through the lens, those rays appear to be coming from A_2', rather than from A_2, because of the refraction by the lens. The most familiar virtual images are, of course, those that we see of ourselves and our surroundings in a common plane mirror or looking glass.



(a) Throughput between A_1 and A_2' (virtual image of A_2).



(b) Throughput between A_1' (virtual image of A_1) and A_2.

Figure 9.4. Two apertures with a thin lens between them. (Vertical scale exaggerated by a factor of 2.)

are exactly the same as those that would come from corresponding points of an aperture at the virtual image A_2' if the lens and A_2 were removed. Consequently, we can evaluate the throughput most readily on that basis -- between A_1 and A_2' in the absence of the lens L and aperture A_2.

In figure 9.4(b), in exactly the same way, is shown the virtual image A_1' of A_1. And it combines, similarly, with A_2, in the absence of A_1 and L, to establish solid angles $\omega_2(x_2,y_2)$, subtended by intersecting straight-line rays from A_1' at points x_2,y_2 across A_2, for evaluating the throughput

$$\Theta = \int_{A_2} \int_{\omega_2(x_2,y_2)} \cos\theta_2 \cdot d\omega_2 \cdot dA_2 \ [m^2 \cdot sr]. \tag{9.9}$$

In Chapter 2 [9.1], in the discussion following eqs. (2.28) and (2.29), we establish the geometrical invariance of throughput for any beam as long as no rays are added or removed. Consequently, the throughput at the two locations, given by eqs. (9.8) and (9.9), must be exactly the same. For those who find a numerical example more satisfying than this analytical conclusion, computations of the two throughputs are carried out, in Appendix 9A, for both aperture/aperture-image pairs, using eq. (A3-15) [9.6], for an assumed configuration of aperture and lens dimensions, positions, and focal length.

Generally, when there are focusing optics between two apertures that, together, define a radiation beam, an equivalent beam-defining aperture pair (joined by undeviated straight-line rays) is formed by either aperture and the image of the other aperture produced by all of the intermediate optical elements.

OBJECT SPACE and IMAGE SPACE. Note that, in figures 9.4(a) and (b), the straight-line-equivalent rays between each of the equivalent beam-defining aperture/aperture-image pairs coincide with the actual rays (solid lines) on just one side of the lens. They are extended as straight dashed lines past the lens and the second stop as though those optical elements were not there, to intersect at the virtual image of that stop. This situation is briefly described in geometrical optics by stating that both the aperture and the aperture image of a pair, and the straight-line rays between them, exist in the same space: the other pair exists in a second space. One "space" is associated with the rays on the left side of the lens, where both actual and equivalent rays coincide; the other is associated with the actual (refracted) rays on the right side of the lens where, again, the actual and equivalent rays coincide.

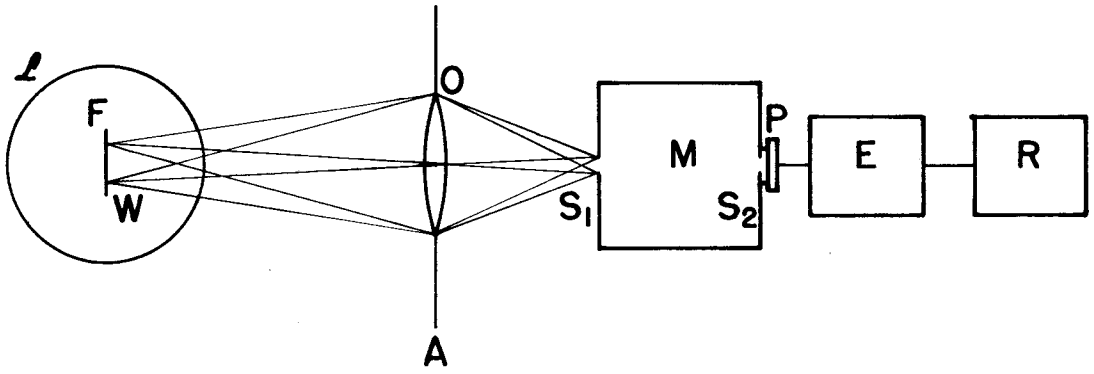
The conventions and terminology of geometrical (ray) optics are, typically, related to situations involving imaging. In radiometry, however, we often use focusing optics in situations where we are not concerned with forming an image of an object or source of radiation but, rather, with defining a beam of radiation from that source. When focusing optics are used for imaging, the space associated with the rays between the object or source and the focusing optics, and the straight-line extensions of those rays past any optical elements as though they were not present, is referred to as object space. Similarly, the

space associated with the deviated (refracted) rays emerging from the last focusing element and their straight-line extensions back past all optical elements as though they were not present, is called image space, since it is these rays that may intersect to form a real image (or their extensions may intersect to form a virtual image) of the source or object. Both spaces overlap completely and coexist on both sides of the optics. The determining factor is the side on which the rays themselves (*not* their straight-line extensions) actually exist.

Customarily, a source or object is illustrated as being on the left of a focusing lens or lenses, with the radiation propagating through from left to right to form a real image, if any, of the source or object on the right-hand side of the optics (or a virtual image on the left side by straight-line extensions of the right-hand-side rays). Accordingly, unless explicitly stated otherwise, we will usually refer to the space associated with the actual rays on the left as object space and that associated with actual rays on the right as image space. For example, the aperture pair A_1-A_2' [figure 9.4(a)] would be said to be in object space, while the pair A_1'-A_2 [figure 9.4(b)] is in image space. Note that this is consistent with the fact that, in visible light, the rays between A_1 and A_2' appear perfectly straight when viewed from the object side of the lens, while A_2' can't even be seen from the image (right) side, i.e., it doesn't "exist" in image space, etc.

Similar relationships exist with respect to a focusing or imaging mirror element, but then we can't speak of the rays on different sides of the element because the reflection returns all rays to the same side. Instead, we must then distinguish between the incident rays in object space (those coming from the object or source toward the element) and the exitent rays in image space (those leaving the element), and their respective straight-line extensions past the element to form virtual images beyond (behind) it.

APERTURE STOPS, FIELD STOPS, and THEIR IMAGES: PUPILS and WINDOWS. In Chapter 5 [9.3] of this Manual, figure 5.1 (repeated here as figure 9.5) shows the physical configuration used in a laboratory measurement of the radiance of the glowing filament of a tungsten-ribbon lamp. In simple cases like this, where a relatively small source on the optic axis (the target area W) is imaged on a receiver (the entrance slit S_1), the aperture that limits the pencil or solid angle of rays from the source which is accepted by the instrument is called the aperture stop, wherever it may be located in the instrument. In this case, the aperture A in the lens mounting limits the size of the accepted beam to which the instrument responds and, hence, the solid angle that it subtends at the target W, so it is the aperture stop of the system. With more complicated multi-element optics, where the aperture stop may be located between some of the elements, we refer to the image of the aperture stop, formed by all elements that precede it (those between it and the source), as the entrance pupil of the system. A second aperture that, along with the aperture stop, determines the beam that can be accepted by the instrument, i.e., its throughput, is called the field stop. It limits the solid angle $\omega(x,y)$ filled by the intersecting rays of the beam at each point x,y across the aperture stop. In this case (figure 9.5), the field stop is the entrance slit S_1. The image of the field stop, formed by all focusing elements that



ℒ -- tungsten lamp with ribbon filament F

W -- target area on emitting surface of F

O -- external focusing optics (for illustrative simplicity we show a single thin lens)

A -- receiving aperture

M -- monochromator, with entrance slit S_1 and exit slit S_2 (internal details not shown)

P -- photocell, responds to radiation emerging from exit slit S_2

E -- electronics (amplification and signal processing)

R -- readout (recording and/or display of output "signal")

Figure 9.5 Diagramatic horizontal section of measurement configuration for spectral-radiance comparison measurements (previously figure 5.1 [9.3]).

Note: The target area W is the entrance window of the entire optical system and the receiving aperture A is the entrance pupil.

precede it, is the entrance window, in this case the target area W.[1] Accordingly, in this simple case (small axial source imaged on receiver) the accepted beam or throughput is defined by the entrance window W in combination with the entrance pupil A, both in object space.

More complicated, multi-element optical systems may also involve repeated reimaging. However, if the instrument is sharply focused to accept radiation only from a well-defined portion of an extended source, this means that some stop or limiting aperture is imaged there, at the source. It is that stop that is designated as the field stop and its image there on the source (in object space) is the entrance window. Similarly, at least in most well-designed optical systems for radiometric instruments, another stop or aperture restricts the accepted solid angle or pencil of rays from any point of the entrance window that can pass through the optical system without obstruction. It determines the transverse dimensions of the beam entering the instrument. It is called the aperture stop and its image in object space is the entrance pupil. Together, the entrance window and entrance pupil form a beam-defining aperture(-image) pair in object space. They establish the maximum accepted beam, the one containing all possible rays that can pass through the optical system without obstruction. The measure of this beam is the throughput Θ of the optical system.

DESIGNATIONS of STOPS, PUPILS, and WINDOWS in RADIOMETRY. The defined throughput or maximum accepted beam of the optical system, just described, exists whether or not there is actually an extended source at the entrance window. In radiometry, it is not uncommon to find situations where focusing or imaging of a source does not take place but where, nevertheless, an optical system serves to define the accepted beam in the way just described, i.e., by two stops whose images in object space form a beam-defining aperture(-image) pair. The member of this pair farthest[2] from the instrument is loosely called the entrance window and the associated stop, the field stop; the other image and stop, similarly, being termed the entrance pupil and the aperture stop, respectively. We say "loosely" because the conventions of geometrical optics apply, strictly, only to situations involving an object and its image. In fact, the rules for assigning the designations of aperture stop and field stop are based on the object-point location and the designations may be different for different locations of an object with respect to the same optical system [9.7,9.8,9.9]. However, for radiometry it isn't really important which member of the beam-defining aperture(-image) pair is the entrance pupil and which is the entrance window. Accordingly, we won't attempt to go into the geometrical-optics rules here. Those who need them will find them in the cited texts [9.7,9.8]. Meanwhile, we will continue, in most instances, to follow the "loose" but convenient practice just described in designating the aperture and field stops of most optical systems. Incidentally, since they are also used occasionally

---

[1]Thus, S_1 determines the target area W or "field" from which the instrument accepts radiation for measurement; hence the designation "field stop".

[2]It may even be "at infinity" -- see next section.

in radiometry, we should mention that the image of the aperture stop formed by all optical elements that follow it (conventionally to the right) is called the exit pupil and, similarly, the image of the field stop formed by all following optical elements is called the exit window. The exit pupil and exit window, thus, form a beam-defining aperture(-image) pair in image space.

OPTICAL SYSTEM FOCUSED on INFINITY. There remains one important special case that is not adequately covered by the foregoing treatment. When an instrument, usually one intended for use in the field, is focused on infinity, the field-stop image that is the entrance window is at infinity. Consequently, it's no longer accessible so we can't evaluate the throughput as determined by the entrance window and entrance pupil as a beam-defining aperture(-image) pair.

Actually, things are made simpler by the fact that, when the system is focused on infinity, the angular spread of the incident rays of the beam that intersect at each point of the entrance pupil is the same for all such points. The solid angle $\omega(x,y)$, containing all rays of the beam intersecting at the point x,y of the entrance-pupil area A is then

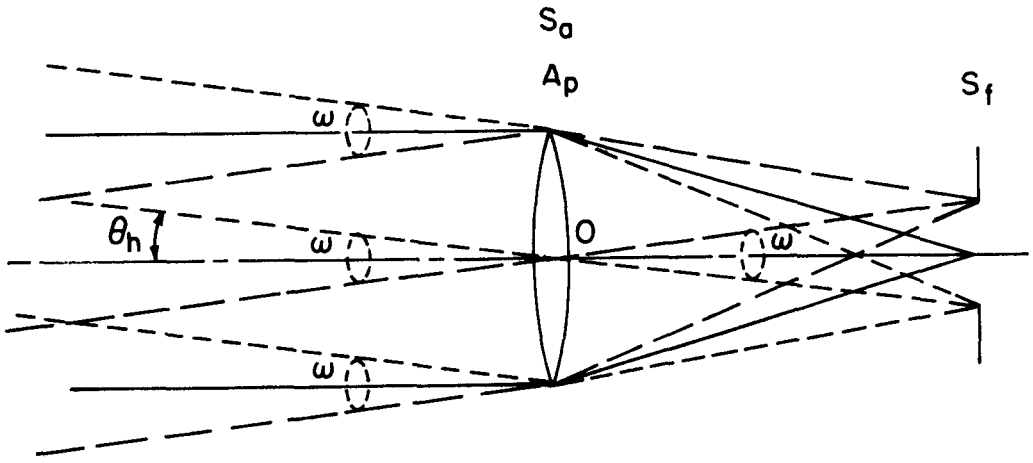


Figure 9.6. A simple optical system focused on infinity (for field use with distant targets or sources).

A_p -- entrance-pupil area [m^2]

S_a -- aperture stop

S_f -- field stop

ω -- field solid angle [sr]

θ_h -- half-field angle [rad]

a constant for all such points: $\omega(x,y) = \omega$, a constant. This is evident in figure 9.6, illustrating how a thin lens focuses incident parallel beams of dashed rays at opposite points on the margin of the field stop in its focal plane. These extreme rays enclose the same solid angle ω wherever they intersect at the entrance pupil which, in this simple case, is the aperture stop formed by the margins of the lens, itself. Also, in this case, evaluation of the solid angle ω is quite simple; it is just the solid angle subtended at the center of the lens 0 by the undeviated rays from the margins of the field stop. More generally, with a thick lens or with multiple primary optical elements, this would be the solid angle subtended at the rear nodal point[1] of the primary optics (in most cases, with all elements in the same medium, air, this is the same as the rear principal point[1]), rather than at the center of the lens 0. Once ω is established, the throughput is quite simply evaluated as the product of the entrance-pupil area A and the projected solid angle $\Omega = \int \cos\theta \cdot d\omega$ {eq. (2.30) [9.1]}. If, with the usual circular field stop, ω is the solid angle at the apex of a right circular cone of half-apex angle θ_h (see figure A2-5 [9.10]), we found in Appendix 2, eq. (A2-14) [9.10] that the corresponding projected solid angle is

$$\Omega = \pi \cdot \sin^2\theta_h \ [\text{sr}]. \tag{9.10}$$

Then, by eq. (2.30) [9.1], the throughput is

$$\Theta = A \cdot \Omega = A \cdot \pi \cdot \sin^2\theta_h \ [\text{m}^2 \cdot \text{sr}]. \tag{9.11}$$

FLEXIBILITY in CHOICE of REFERENCE SURFACE for SPATIAL PARAMETERS. Although we consistently use the receiving aperture or entrance pupil of a radiometer as the logical reference surface for the incident radiation to which an instrument responds, it was pointed out in Chapter 2 [9.1], particularly in connection with the transformation equations, eqs. (2.41) and (2.42) or (2.41a) and (2.42a), that any convenient reference surface that intersects the entire beam may be used for the designation of the spatial parameters of position x,y and direction θ,ϕ. This follows from the invariance of throughput which, accordingly, may be evaluated at any surface that intersects the entire beam. From what we have seen here about the effects of vignetting, the most convenient choices will surely be a stop or beam-defining aperture, or an image of such a stop, but it may be *any* one, depending on where the pertinent lengths and angles are most easily evaluated.

ENTRANCE PUPILS as RECEIVING APERTURES -- SOME PRACTICAL PROBLEMS. We have seen that, in an imaging or focusing optical system, it is the entrance pupil that constitutes the receiving aperture from the standpoint of limiting the transverse dimensions of the incident beam of rays (in object space) that is accepted by the instrument. When the entrance pupil is a virtual image of an aperture stop, inaccesibly located behind a lens, it may be difficult to measure its dimensions and position when these data are needed. Fortunately,

---

[1]See any standard text on geometrical optics [9.7,9.8].

they are not required for solving the measurement equation in the illustrative problems of Chapter 5 [9.3]. For radiance measurements, such as those of Problem 1, it is only necessary to be sure that the dimensions and position of the entrance pupil or receiving aperture remain exactly the same for all of the measurements. They then cancel out, so their values don't enter into the final result, as in eq. (5.11) or eq. (5.20) [9.3]. For irradiance measurements, as in Problems 2 and 3, uniform, isotropic irradiation of the monochromator entrance pupil and constant internal-beam geometry are insured by the use of a diffuser ahead of the focusing optics. An aperture in front of the diffuser is then the receiving aperture for the overall system. In the next section, we discuss some considerations relating to such diffusers and the apertures in front of them. However, when field measurements are made, with an instrument focused on infinity, of the irradiance from distant targets or sources, it is often not feasible to use a diffuser and then it may be quite important to know the exact position of the entrance pupil or receiving aperture.

There are two compelling reasons for not using a diffuser for these field measurements. First, a diffuser reduces the detectivity of the instrument substantially, often beyond the point where it is adequate to respond to the low irradiance from very distant sources. Second, isolation of the radiation of the source or target from that of other unwanted background sources in the uncontrolled field environment, along with an adequately large receiving aperture, can be achieved only with focusing optical elements.

For field measurements of distant sources that subtend very small angles at the radiometer, it may no longer be essential to have isotropic responsivity, even across a relatively narrow field of view of the instrument. However, if the responsivity is not isotropic, it is important to know the value of the responsivity as a function of direction together with the direction of the source within the instrument field at the time of the measurement. This poses some difficult practical problems for moving sources, such as aircraft and missiles, in addition to the problem of dealing with background radiation, e.g., that from the sky, and its contribution to the output of the measuring instrument.

Returning to the problem of the dimensions and position of the entrance pupil or receiving aperture, its area A cancels out and does not appear in the final solution of the measurement equation for irradiance. For example, it appears in both eqs. (5.27) and (5.28), so that it is eliminated when the quotient between them is taken in eq. (5.29). However, the exact position of the entrance pupil may still be very important. An irradiance-standard lamp is usually calibrated and certified for the values of irradiance E^s, or spectral irradiance $E^s_\lambda(\lambda)$, *at a given distance*, so we must be able to verify that the reference surface established by the entrance pupil is, in fact, at that specified distance from the standard source (lamp).[1] As already pointed out, this may be difficult if the entrance pupil is a virtual image of an aperture stop and is inaccessibly located behind a lens or lenses. The simplest solution for precision work is to install an accessible

[1]This problem may not arise with a collimated-beam standard of spectral irradiance if the unknown is also collimated or very far away.

aperture stop in front of the entire optical system, making sure that it becomes the limiting aperture by making it just small enough so that no rays of the measurement beam can quite reach the edges of the internal stop, which is then no longer the aperture stop for that measurement. This will somewhat reduce the flux in the beam but it will probably be a small price to pay for the improved accuracy in establishing the dimensions and location of the aperture stop (serving also as entrance pupil and receiving aperture). It is important to make sure, also, that it *does* form a beam-defining aperture pair with the field stop or its image, i.e., that from every point within the aperture stop rays can pass, without obstruction at some other point, through every point within the field stop.

The field solid angle ω, or, with circular optical elements and stops, the half-field angle θ_h, is also relatively easy to measure directly. A distant small source, at least optically distant (e.g., by means of a collimator), is observed at different angles θ from the optical axis (with the vertex of the angle at the center of the entrance pupil or receiving aperture). The edge of the field solid angle or the angle θ_h is that angle where there is no longer any instrument response. With a steady source, this can also be made a measurement of the directional responsivity characteristic, the relative responsivity as a function of direction $r(\theta,\phi)$. With good symmetry, the responsivity may be independent of ϕ and a function only of the angle from the axis $r(\theta)$. In any case, this information will help to insure that the irradiance measurement is actually made with the source well within a central region of isotropic directional responsivity to the extent that this is required (see Chapter 5, problems 2 & 3 [9.3]).

DIFFUSERS. In problems 2 and 3 of Chapter 5 [9.3], we have already seen that instruments used for measuring irradiance (including spectral irradiance) often employ a diffuser, such as an averaging sphere, as the first optical element in order to achieve the required isotropic directional flux responsivity (or, as an approximate equivalent, a lambertian response that varies as the cosine of the angle of incidence when irradiated by a well-collimated, iso-radiance, incident beam that is wide enough to completely fill the receiving aperture at all times). When used as the first optical element, a diffuser terminates the incident beam and becomes a new source for the exitent beam; there is no ray-to-ray continuity between the two beams. For this purpose, the best diffuser for measurements of optimum accuracy is probably a good integrating or averaging sphere with its entrance port (receiving aperture) defined by a sharp-edged aperture of well-defined position and dimensions (see "More about Vignetting" below, especially figure 9.9). Even then, internal baffles or screens may be needed to insure that all rays of the exitent beam have not been reflected directly from points exposed to the direct incident beam.

Sometimes, when the use of a sphere is not convenient, other forms of diffuser may be employed, such as a flat-plate diffuse reflector or a transmitting ground-glass, opal-glass, or translucent-plastic element. The result may be greater variation in directional responsivity or narrowing of the solid angle within which the response to incident radiation is acceptably isotropic. The problem of adequate baffling to insure that the exitent beam for measurement includes radiation only from a desired portion of the diffusing element may

also be aggravated. On the other hand, the directional responsivity of a transmitting diffuser is sometimes "improved" by adjusting its shape to achieve a more nearly lambertian or "cosine" response, empirically compensating for the fact that it is not a perfect diffuser.

A typical empirically "shaped" diffuser is something like that depicted in figure 9.7(b). The width of a uniform (iso-radiance) collimated beam of parallel rays striking the flat diffuser of figure 9.7(a) is reduced by the factor $\cos\theta$ as the angle of incidence θ is increased. However, most diffusing materials are not perfectly diffusing. Also, reflection loss at the first surface tends to increase with increasing θ. In general, such diffusers become less efficient diffuse transmitters at larger angles of incidence so that the flux reaching the detector beneath the diffuser is reduced by more than the factor of $\cos\theta$. By raising the center of the element and adjusting its shape by trial and error, it is possible to intercept a slightly wider beam at the larger angles of incidence, as shown in figure 9.7(b), to compensate for the losses. In this way, the flux reaching the detector, and hence its response, can be made to follow quite closely a desired "cosine response" as the angle of incidence of a uniform collimated (parallel) beam is changed. However, the beam from a relatively close lamp of finite dimensions can't be expressed as the sum of such conveniently uniform beams of perfectly parallel rays. At present, we don't know how to establish a receiving-area or reference-surface distance or location for such shaped diffusers that will give consistent results for measurements of beams with quite different spatial distributions of flux or beam geometries.

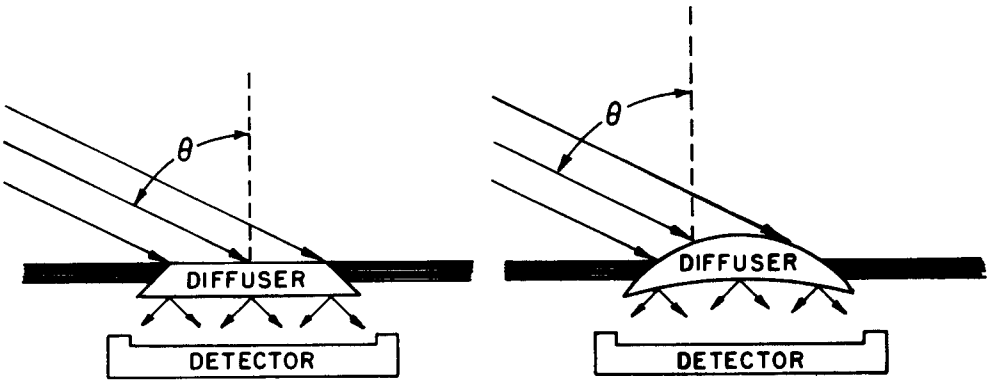


Figure 9.7. Diffuser-detector configurations for measuring irradiance. Note that, for isotropic flux responsivity over the widest possible angle, it is usually desirable to make the detector wide enough to collect as much as possible of the diffusely transmitted radiation coming from the back (underside) of the diffusing element, as shown.

When it is known that the beam geometry and the spatial distribution (both position and direction) of flux within the beam are exactly the same for both the unknown source and the calibration source, there is, of course, no problem. Then any sample of the same portion of each beam can serve as a measurement of the entire beam or any portion of it in each case, so it's only necessary to use the same reference surface consistently for both measurements. The problem of identifying the exact receiving-area distance for the irradiance measurement assumes importance only to the extent that the two beam geometries (unknown-source and calibration-standard spatial distributions of flux) differ. Unfortunately, we have not been able to devise any general method for establishing the receiving area or reference-surface distance in such situations.

In summary, most real diffusers fall considerably short of the ideal, producing truly isotropic flux responsivity for the instrument only over a limited solid angle or range of directions rather than the desired full hemisphere. This makes it particularly important to measure the directional responsivity and to make sure that all significant sources (both unknown and standard) are clearly within the isotropic-responsivity region. On rare occasions, if the beam geometry of both unknown and standard sources is identical, measurements of samples of the beam taken in the same way can establish an adequate comparison regardless of variations in responsivity as a function of direction as long as it is the same for both measurements. However, in almost all instances, with a standard source that has a geometrical configuration quite different from that of the unknown source, accurate irradiance measurements cannot be made without an instrument of adequately isotropic flux responsivity.

BAFFLES. When the defined beam completely fills both of just two beam-defining apertures or stops, with no other obstruction or vignetting, this means that all other apertures through which that beam passes are then large enough so that they do not block any rays that are passed by both beam-defining apertures. The other apertures are then either baffles, introduced to intercept and reduce scattered radiation, or they are supports for the optical elements. Baffles are used to block strong rays that, while they cannot pass directly through all of the stops or beam-limiting apertures, might reach the detector or transducer by reflection or scattering from the internal surfaces of the instrument chamber.

MORE on VIGNETTING. Vignetting takes place, not only at the edges of a beam beyond or between the stops as illustrated in figures 9.2 and 9.3, but also at the nominal stops when some of the rays between points within the nominal stops are blocked by obstructions elsewhere, either between or beyond the nominal stops, as in figures 9.8(a) or (b). For those blocked rays, the nominal stop is no longer the actual stop; instead, the obstacle becomes the stop (beam-limiting-aperture margin) for the blocked rays, so that there is no longer just a beam-defining aperture pair or pair of stops.

Even when we do have just a beam-defining aperture pair or pair of stops, it can be difficult enough to evaluate the integrals in

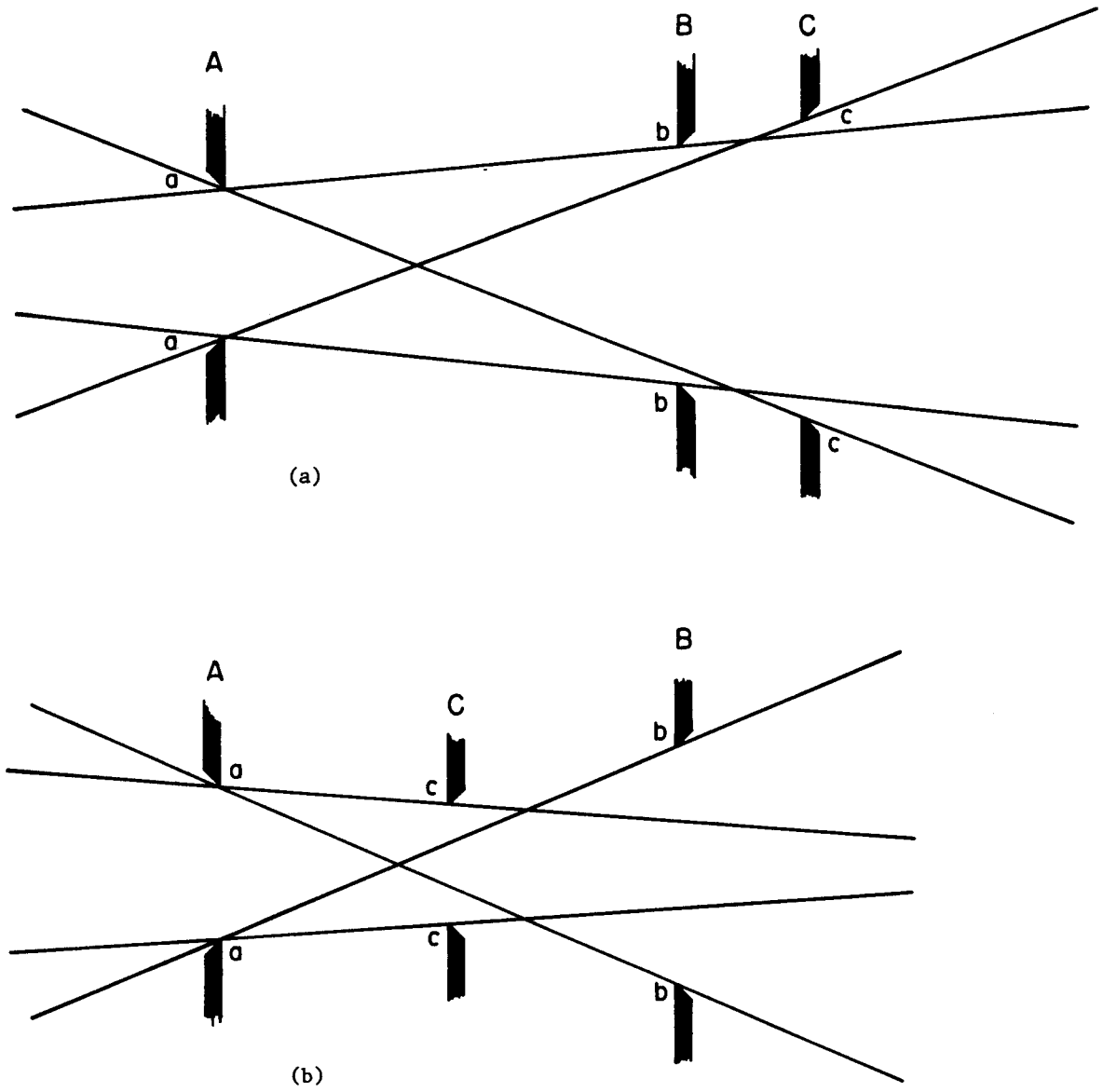


Figure 9.8. Two cases of vignetting.
A and B are the nominal stops (beam-defining aperture pair). Introduction of aperture C causes vignetting where unobstructed rays of the beam through each point of A can no longer reach all points of B, and vice versa. There are now extreme rays a - c on one side and a - b on the other.

$$\Theta = \int_A \int_{\omega(x,y)} \cos\theta \cdot d\omega \cdot dA \ [m^2 \cdot sr] \qquad (9.2)$$

because the solid angle $\omega(x,y)$, filled by rays of the beam that intersect each area element dA (at each point x,y over the area A) is often, as indicated, a significant function of the position coordinates x,y of that element. That's why we're always glad to take advantage of the published tables of configuration factors, as described in Appendix 3 [9.6], to evaluate such throughputs as that between two parallel coaxial circular apertures in eq. (A3-15) [9.6]. When the limits of ω involve the margins of more than one aperture or barrier at different distances (i.e., when there is vignetting), evaluation of the throughput becomes even more complicated and difficult [9.5].

Note also that, unless the stops have thin sharp edges, there can be vignetting of oblique rays at the edges of a stop (see figure 9.9). On the near side, they are limited by the front edge of the stop; on the far side by the back edge. If the edge is rounded, the plane of the limiting margin changes continuously around the stop for an oblique, well-collimated beam. A well-designed stop ordinarily has sharp edges lying in a plane perpendicular to the optical axis.

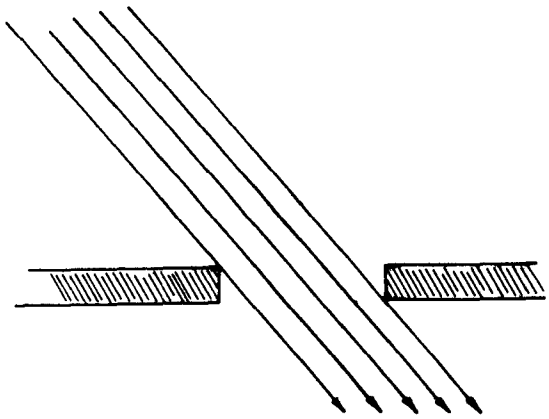


Figure 9.9. Vignetting by an aperture in a thick wall.
The front (upper) edges and the rear (lower) edges form circular stops in different planes. Oblique parallel rays, as illustrated, are limited by the front (upper) edge on the near (left) side and by the rear (lower) edge on the far (right) side. Of course, if these were just slightly diverging rays from a point source, or one smaller than the aperture, the rear (lower) edge would become the limiting aperture if the source were accurately aligned on the axis, perpendicularly above the aperture. However, in general, radiation from a more extended source will include parallel bundles of rays, incident at angles away from the normal, with the effect shown.

RESPONSIVITIES and BEAM GEOMETRY. In an instrument with uniform, isotropic, linear flux responsivity [$R_\Phi(x,y,\theta,\phi) = R_\Phi$, a constant, for all rays that reach the detector], the radiance responsivity R_L is directly related to the flux responsivity R_Φ by the throughput Θ:

$$R_\Phi \equiv S/\Phi = S/(L\cdot\Theta) \ [S\cdot W^{-1}],$$

$$R_L \equiv S/L = \Theta\cdot R_\Phi \ [S\cdot W^{-1}\cdot m^2\cdot sr]. \tag{9.12}$$

Note that, although we don't have a different symbol, these are now overall instrument responsivities. They would be average responsivities if the individual-ray responsivity $R_\Phi(x,y,\theta,\phi) \equiv dS/d\Phi$ were not the same for all rays. Here Φ [$=\int\int L\cdot\cos\theta\cdot d\omega\cdot dA$] is the total flux in the beam to which the instrument responds, S is the resulting output signal, L is the radiance of an iso-radiance incident beam that completely fills the throughput Θ (or the average radiance of such a beam if the radiance is not everywhere the same throughout the beam).

Equation (9.12) is true, regardless of vignetting [as long as the effect of the vignetting is correctly accounted for in the evaluation of the throughput Θ, e.g., by choosing the correct limits for $\omega(x,y)$ as it varies from point to point in eq. (9.1)]. However, the corresponding relation between flux responsivity R_Φ and irradiance responsivity R_E in eq. (5.49) [9.3]

$$R_E \equiv S/E = A\cdot R_\Phi \ [S\cdot W^{-1}\cdot m^2], \tag{9.13}$$

is valid, for the full aperture or entrance-pupil area A, only when there is no vignetting. This is apparent if we stop to think in terms of a uniform incident beam that is well collimated (parallel rays from nearly the same direction). If there is vignetting and such an incident beam swings far enough away from the normal or axial direction, so that rays through part of the nominal receiving aperture or entrance pupil are blocked by the vignetting obstacle elsewhere in the optical system, the result is a reduction of the actual aperture area A and a corresponding reduction in the irradiance responsivity R_E given by eq. (9.13). Accordingly, such instruments should be used, if possible, only for incident beams or sources that lie well within the angular field or entrance window (away from directions that involve the vignetting) where the irradiance responsivity R_E is constant. We won't even attempt to deal here with all of the problems that can arise if vignetting can't be avoided.

SUMMARY of CHAPTER 9. In radiometry, a radiation beam is usually defined physically by two apertures. The beam then consists of all rays that pass through both apertures. They are called beam-defining apertures if rays from each point of the plane across one aperture reach all points within the other aperture *and* if all such rays through both apertures extend for the full length of the beam, from source to receiver, without encountering any

other obstacle. If, in addition, the medium between the apertures is homogeneous (iso-refractive-index), so that all rays between them are straight lines, we call them a beam-defining aperture pair, or just an aperture pair. When focusing optics are used and there are one or more focusing elements between the beam-defining apertures, the rays between them will be refracted (deviated) by the focusing element(s). Then the beam (throughput) at either aperture (chosen as the reference aperture) is identically the same as that defined by an equivalent aperture pair consisting of that (reference) aperture paired with the image of the other aperture formed by the intervening optical element(s) {see figure 9.4(a) and (b)}. The straight-line rays joining this equivalent pair coincide with the actual rays of the beam in the region about the reference aperture and extend, undeviated, past the intervening optical elements as though those elements were not present.

Beam-defining apertures in a focusing optical system are called stops. The stop that, regardless of its location in the system, determines the width of the beam where it enters or is incident on the optical system is designated as the aperture stop. The image of the aperture stop, formed by any optical elements that precede it (elements traversed by the incident beam before reaching it), is called the entrance pupil. If there are no such elements, the stop itself is the pupil (see figure 9.5). The entrance pupil serves as the receiving aperture of the instrument (with respect to incident rays and their undeviated straight-line extensions past any focusing optical elements preceding the entrance pupil). The entrance pupil is the image of the aperture stop in object space, the space associated with the source or object, usually shown to the left of the optical system in illustrations with incident rays coming from the left. Object space also includes the undeviated straight-line extensions of those incident rays past any focusing optical elements into the region to the right of them {as though they were not present, as in figure 9.4(a)}.

Similarly, the stop that, regardless of its location in the optical system, determines the solid angle filled by intersecting incident rays of the beam at each point across the entrance pupil is called the field stop. The image of the field stop formed by all optical elements that precede it, i.e., the image of the field stop in object space, is called the entrance window. The entrance window and entrance pupil, both in object space, form a beam-defining aperture(-image) pair. When the magnitude of the beam throughput is needed, it is often convenient to evaluate it as the throughput of this aperture(-image) pair (entrance window and entrance pupil). A simple illustration of such numerical evaluations is presented in Appendix 9A. In addition, the solid angle subtended by the entrance window at the center of the entrance pupil in object space is called the angular field, or sometimes just the field, of the optical system. For systems with coaxial circular elements and stops, the half-vertex angle of the right circular cone formed by the entrance window as its base and the center of the entrance pupil as its vertex is called the half-field angle of the optical system.

Although not often used in radiometry, our list of related geometrical-optics quantities and concepts also includes the exit pupil, the image of the aperture stop in image space, i.e., the image formed by all optical elements that follow it (aperture stop), and

the exit window, the image of the field stop in image space. They are included here primarily for completeness.

In the special case of an instrument focused on infinity (entrance window at infinity), it is not feasible to evaluate the throughput as that of the aperture(-image) pair formed by the entrance window and entrance pupil in object space. However, in this case the field solid angle has the same value at all points of the entrance pupil (see figure 9.6) so the corresponding projected solid angle is also a constant as a function of position across the entrance pupil. Accordingly, the throughput which, in general, is the integral of the product of projected area and solid angle (or area and projected solid angle) over the entrance-pupil reference surface (see figure 9.1) then becomes simply the product of the entrance pupil area and the constant projected solid angle $\{\Theta = \iint dA \cdot \cos\theta \cdot d\omega = A \cdot \Omega \ [m^2 \cdot sr]\}$.

Vignetting takes place when a beam does not completely fill both of two nominal stops or beam-defining apertures, i.e., when it is limited in one transverse direction by one aperture margin and in other transverse directions by one or more additional apertures or obstacles (see figures 9.3 and 9.8). When vignetting occurs, it becomes very difficult to evaluate the beam throughput or related measures of beam geometry, or even to make sure that they are the same (physically unchanged) for two different radiometric measurements.

In most of this chapter, the geometrical-optics quantities and concepts are treated in terms of throughput and the limits of the throughput integral(s), because they are the quantitative measures of beam geometry. It is only in these terms that an adequate understanding of their significance can be achieved, including the ability to confidently decide, in all cases, whether or not a numerical evaluation of the throughput may occasionally be required.

In an instrument with uniform, isotropic, linear flux responsivity, where $R_\Phi(x,y,\theta,\phi) = R_\Phi$ is constant for all rays that reach the detector (are not blocked by the optical system), the radiance responsivity R_L is directly related to the flux responsivity R_Φ by the throughput Θ:

$$R_\Phi \equiv S/\Phi = S/(L \cdot \Theta) \ [S \cdot W^{-1}]$$

$$R_L \equiv S/L = \Theta \cdot R_\Phi \ [S \cdot W^{-1} \cdot m^2 \cdot sr]. \qquad (9.12)$$

This is true regardless of vignetting, as long as the effect of the vignetting is correctly accounted for in the evaluation of the throughput Θ. However, the corresponding relation between flux responsivity R_Φ and irradiance responsivity R_E,

$$R_E \equiv S/E = A \cdot R_\Phi \ [S \cdot W^{-1} \cdot m^2], \qquad (9.13)$$

is valid, for the full aperture or entrance-pupil area A, only when there is no vignetting.

Customarily, when the defined beam completely fills both of just two beam-defining apertures or stops, with no other obstruction or vignetting, all other apertures through which that beam passes are either baffles or supports for the optical elements. Baffles are employed to reduce or eliminate scattered radiation by blocking strong rays that, while they cannot pass directly through all stops or beam-limiting apertures, might reach the detector or transducer by reflection or scattering from the internal surfaces of the instrument chamber.

The receiving aperture or entrance pupil of a radiometer is consistently used, here, as the logical reference surface for incident radiation to which the instrument responds. However, *any* unvignetted stop or beam-defining aperture (of a pair), or an image of same, i.e., any pupil or window, may be the reference surface for the spatial parameters of position and direction, depending on where the pertinent lengths and angles are most easily evaluated.

Diffusers are often used to achieve the isotropic flux responsivity needed for irradiance measurements (see problems 2 and 3 of Chapter 5 [9.3]). They are usually located just behind a well-defined receiving aperture where they terminate the incident beam and become a new source for the internal beam through the remaining optical components. Ideally, the radiance of every ray in that internal beam is directly proportional to the total incident flux and is not affected by any changes in the spatial distribution of that flux in the externally incident beam. The use of diffusers involves a number of special problems that are discussed in some detail in this chapter, including the uncertainties in the receiving-aperture location when empirically "shaped" diffusers are used in attempts to compensate for imperfectly diffusing materials.

The treatment of geometrical-optics quantities and concepts in this chapter is all in terms of ideal optical elements that produce perfectly sharp images and shadow boundaries. In addition, we've ignored attenuation losses, since we've already seen how, at least in principle, to introduce the propagance (see Chapter 2 [9.1]) in order to account for them. However, real instruments also suffer from aberrations that produce imperfect images and from scattering and diffraction that produce attenuation of some rays and augmentation of others, resulting in a spatial redistribution of flux. We hope to treat the resulting problems in future chapters of this manual. Accordingly, we must re-emphasize the statement at the beginning of this chapter that the material presented here, while it is very important for design purposes and as a starting point for analysis, is no more than that -- a starting point or first approximation -- and must never be considered as establishing exact values or absolute limits (e.g., integration limits in the measurement equation) that cannot be exceeded.

Finally, although the geometrical-optics concepts of pupils and windows are conventionally defined in terms of an object (source) and its image, they are defined here, for radiometry, with respect to a radiation beam which may or may not involve imaging of the radiation source. These designations are rather widely used in the literature on radiometry

and such an extension of the customary geometrical-optics definitions is needed to provide a consistent basis for distinguishing between aperture stops and pupils and field stops and windows when there is no source imaging involved. However, the really important consideration in radiometry is to clearly recognize and understand the significance of beam-defining apertures and equivalent aperture/aperture-image pairs, regardless of which is designated as the aperture stop or pupil and which the field stop or window.

Appendix 9A. Numerical Evaluation Confirming Invariance of Throughput for Beam through Thin Lens between Two Beam-Defining Apertures

by Fred E. Nicodemus

In Chapter 9, we have treated the case of a thin lens between two beam-defining apertures. We find two aperture/aperture-image pairs, one in object space and one in image space, that define the beam of all rays through both apertures. The corresponding throughput values are given, respectively, by eqs. (9.8) and (9.9). The argument for the invariance of throughput in a given beam, when no rays are added or removed, which follows eqs. (2.28) and (2.29) in Chapter 2 [9.1], is of general validity and requires no further proof as the basis for asserting that the throughput given by eqs. (9.8) and (9.9) must be exactly the same. Nevertheless, some readers may find it satisfying to corroborate this analytical conclusion numerically. Also, carrying out the numerical throughput calculations for both aperture/aperture-image pairs, using eq. (A3-15), may help to bring out more clearly the way in which apertures and aperture images enter into the evaluation of the throughput as well as to physically define a beam.

Both figures, 9.4(a) and (b), are drawn approximately to scale for the dimensions given below except that, for clearer illustration of the relationships, the vertical dimensions are exaggerated by a factor of two. We choose the lens focal length as $f = 7.00$ [cm] and its radius as $r \approx 1$ [cm] (large enough to pass all necessary rays without obstruction). A_1 is a circular aperture of radius $r_1 = 0.500$ [cm] in a vertical plane at a distance $x_1 = 3.00$ [cm] to the left of the center of the lens, which we designate as the origin 0. A_2, similarly, has a radius $r_2 = 0.600$ [cm] at a distance $x_2 = 2.000$ [cm] to the right of 0. In order to determine the size and position of each virtual image, we make use of the thin-lens formula[1]

$$\frac{1}{x_o} + \frac{1}{x_i} = \frac{1}{f} \ [\text{cm}^{-1}], \tag{9A.1}$$

where x_o is the object distance (from 0), x_i is the image distance (also from 0), and f is the focal length of the lens. We also need the fact that image size is in direct proportion to distance from 0, since each image point must lie on the undeviated ray (drawn with long dashes) through 0 and the corresponding object point. Thus A_2' has a radius $r_2' = 0.840$ [cm] and is located at a distance $x_2' = 2.800$ [cm] to the right of 0; A_1' has a radius $r_1' = 0.875$ [cm] and is located $x_1' = 5.250$ [cm] to the left of 0.

A_1 and A_2' then constitute a circular aperture pair of radius $r_1 = 0.500$ [cm] and $r_2' = 0.840$ [cm], respectively, separated by a distance $D_1 = x_1 + x_2' = 5.800$ [cm]. Using eq. (A3-15) [9.6], the throughput for this aperture pair is evaluated as $\Theta_1 = 0.0503$ [$\text{cm}^2 \cdot \text{sr}$]. In the same way, A_1' and A_2 constitute a circular aperture

[1] See any standard text on geometrical optics [9.7,9.8], or almost any general-physics text.

pair of radius $r_1' = 0.875$ [cm] and $r_2 = 0.600$ [cm], respectively, separated by $D_2 = x_1' + x_2 = 7.250$ [cm]. The throughput for this second aperture pair is evaluated similarly as $\Theta_2 = 0.0507$ [$cm^2 \cdot sr$].

Since the thin-lens formula is an approximation, derived for paraxial rays (rays deviating from the axial direction by angles small enough so that we can use the approximations $\sin\theta \approx \theta$, for θ measured in [rad], and $\cos\theta \approx 1$), this agreement to less than one per cent for the two throughput values is gratifying confirmation of the validity of this approach.

References

[9.1] Fred E. Nicodemus and Henry J. Kostkowski, "Distribution of Optical Radiation with Respect to Position and Direction -- Radiance", Chapter 2 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech. Note 910-1, 93 pages (March 1976) pp. 10-44.

[9.2] Fred E. Nicodemus, "More on the Distribution of Optical Radiation with Respect to Position and Direction", Chapter 4 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech. Note 910-2, 118 pages (Feb. 1978) pp. 1-57.

[9.3] Henry J. Kostkowski and Fred E. Nicodemus, "An Introduction to the Measurement Equation", Chapter 5 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech. Note 910-2, 118 pages (Feb. 1978) pp. 58-92.

[9.4] John B. Shumaker, "Distribution of Optical Radiation with Respect to Polarization", Chapter 6 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech Note 910-3, 62 pages (June 1977).

[9.5] "Optical Radiation Measurements: The 1973 Scale of Spectral Irradiance", Nat. Bur. Stand. (U.S.), Tech. Note 594-13, 36 pages (Apr. 1977), Appendix, pp. 20-28.

[9.6] Fred E. Nicodemus, "Projected Solid Angles, Throughputs, and Configuration Factors", Appendix 3 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech. Note 910-2, 118 pages (Feb. 1978) pp. 93-100.

[9.7] Francis A. Jenkins and Harvey E. White, "Fundamentals of Optics", 3rd ed. (McGraw-Hill, New York, 1957).

[9.8] Miles V. Klein, "Optics" (Wiley, New York, 1970).

[9.9] R. H. Anderson, "Closeup Imaging of Documents and Displays with Lens Arrays," Appl. Opt. 18, No. 2 (15 Feb 1979) 477-484; there is a good example of the dependence of the entrance pupil on object position described on p. 480.

[9.10] Fred E. Nicodemus, "Spherical Coordinates and Geometrical Relationships", Appendix 2 of "Self-Study Manual on Optical Radiation Measurements: Part I--Concepts", Nat. Bur. Stand. (U.S.), Tech. Note 910-1, 93 pages (March 1976) pp. 62-80.

ERRATA and ADDENDA for NBS Tech Notes 910-1 and 910-2:

1. Table A1-3 (p. 58 of TN 910-1): Add note: A more complete version of this table appears as table 4-1 on p. 3 of TN 910-2.

2. Table A1-4 (p. 59 of TN 910-1): Add note: This table also appears as table 4-2 on p. 5 of TN 910-2.

3. Table A1-6 (p. 61 of TN 910-1): Add note: This table also appears as table 4-3 on p. 6 of TN 910-2.

4. Table A1-6 (p. 61 of TN 910-1) *and* table 4-3 (p. 6 of TN 910-2): The modifier for the first three entries should be changed from "photon" to "photon-flux", just as in the fifth and succeeding entries. Note that this change is particularly important in the case of the first entry, photon-flux energy Q_p $(= \int \Phi_p \cdot dt)$, because of the very wide and well established use of the term "photon energy" for $h\nu$, the energy per photon, i.e. the energy of a single photon or quantum.

5. Table A1-4 (p. 59 of TN 910-1) *and* Table 4-2 (p. 5 of TN 910-2): Add alternative term for the first quantity listed so that it reads, "luminous energy; quantity of light".

6. Table A1-6 (p. 61 of TN 910-1) *and* Table 4-3 (p. 6 of TN 910-2): Add alternative term for the first quantity listed so that it reads, "photon-flux energy; number of photons".

ADDENDUM to tables A1-6 (p. 61 of TN 910-1) and 4-3 (p. 6 of TN 910-2),

Units for Photon-Flux Radiometry

The relationship of the photon-flux quantities of Tables A1-6 and 4-3 to the corresponding radiometric quantities of Tables A1-3 and 4-1 may be stated explicitly as follows:

$$\Phi_p = \int_{\lambda_1}^{\lambda_2} \Phi_{e,\lambda}(\lambda) \cdot (h \cdot c/\lambda)^{-1} \cdot d\lambda \quad [q \cdot s^{-1}],$$

where Φ_p $[q \cdot s^{-1}]$ is the photon flux in the spectral interval λ_1 to λ_2 [nm] in a beam characterized by the spectral radiant flux $\Phi_{e,\lambda}(\lambda)$ $[W \cdot nm^{-1}]$, and where Planck's constant $h = (6.626\ 176 \pm 0.000\ 036) \times 10^{-34}$ $[J \cdot s]$, the vacuum speed of electromagnetic radiation $c = (2.997\ 924\ 580 \pm 0.000\ 000\ 012) \times 10^{17}$ $[nm \cdot s^{-1}] \approx 3 \times 10^8$ $[m \cdot s^{-1}]$, and λ [nm] is the wavelength, so that the energy per quantum (per photon) is $(h \cdot c/\lambda)$ $[J \cdot q^{-1}]$.* Similar

*Strictly, the energy per quantum is $(h \cdot c/\lambda_o) = \{h \cdot c/(n \cdot \lambda)\}$ $[J \cdot q^{-1}]$, where $\lambda_o = n \cdot \lambda$ [nm] is the vacuum wavelength and n [dimensionless] is the index of refraction of the medium in which the wavelength is λ [nm]. Also, $\nu = c/\lambda_o$ [THz] for $c \approx 3 \times 10^5$ $[km \cdot s^{-1}]$ and λ_o in [nm]; and $\sigma = 1/\lambda_o$ $[cm^{-1}]$ only when λ_o is given in [cm].

relations exist for all of the other corresponding quantities and their spectral distributions, e.g., for photon-flux (surface) density $E_p [q \cdot s^{-1} \cdot m^{-2}]$, where the spectral irradiance is $E_{e,\lambda}(\lambda)$ $[W \cdot m^{-2} \cdot nm^{-1}]$,

$$E_p = \int_{\lambda_1}^{\lambda_2} E_{e,\lambda}(\lambda) \cdot (h \cdot c/\lambda)^{-1} \cdot d\lambda \quad [q \cdot s^{-1} \cdot m^{-2}].$$

If the spectral distribution is with respect to frequency ν [THz], this becomes

$$E_p = \int_{\nu_1}^{\nu_2} E_{e,\nu}(\nu) \cdot (h \cdot \nu)^{-1} \cdot d\nu \quad [q \cdot s^{-1} \cdot m^{-2}],$$

where the energy per quantum is now given as $h \cdot \nu$ $[J \cdot q^{-1}]$. Similar relations can also be given for spectral distributions in terms of wave number σ $[cm^{-1}]$.

ERRATA: Several errors have been noted in the labelling in Figure A2-4 on page 68 of NBS TN 910-1. It should be corrected to appear as follows:

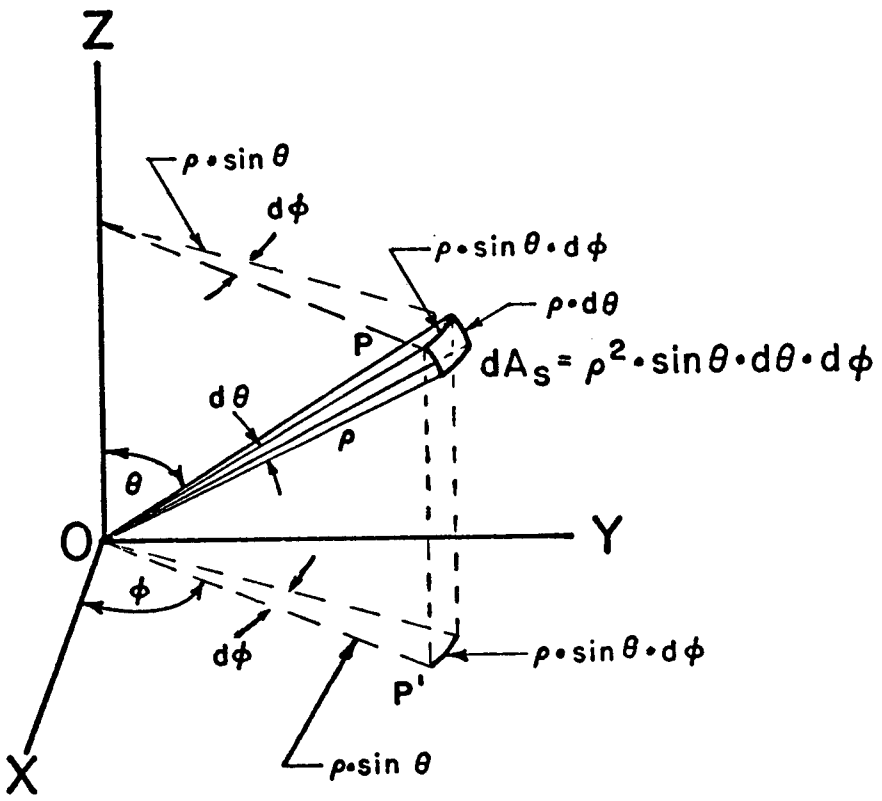


Figure A2-4. An element of solid angle

$d\omega = dA_s/\rho^2 = \sin\theta \cdot d\theta \cdot d\phi \quad [rad^2]$