



City Research Online

City St George's, University of London

Citation: Landau, U. M. (1980). Application of signal processing to IR spectrophotometric instrumentation. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/37053/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

APPLICATION OF SIGNAL PROCESSING TO
IR SPECTROPHOTOMETRIC INSTRUMENTATION

by Uri Mair Landau

A thesis submitted for the Degree of
Doctor of Philosophy

The City University, London
Department of Systems Science

August 1980

Contents

Acknowledgements	ix
Declaration	x
Abstract	xi
List of symbols and abbreviations	xiii
<u>0 Introduction</u>	1
<u>1 Principles of IR Absorption and Spectrophotometry</u>	7
1.1 The nature of ir absorption	8
1.1.1 The spectrophotometer transfer function	12
1.1.2 Infra Red detectors and noise sources	16
1.2 The dual beam technique	26
<u>2 The Differential Equation of the Spectrophotometer Slit Transfer Function</u>	32
2.1 Introduction	32

2.1.1 A simple solution of the discrete presentation of an analogue process	35
2.1.2 Some features of the discrete presentation of an analogue process	36
2.2 The digital - analogue differential equation of the spectrophotometer slit transfer function	40
2.2.1 The digital - analogue differential equation of some class of symmetric finite impulse response systems	51
2.3 A discrete simulation of a gaussian impulse response system	56
2.4 Summary	57
<u>3 A Discrete Realisation of the Spectrophotometer Slit Transfer Function</u>	59
3.1 Introduction	59
3.2 Derivation of a simple deconvolution formula for a triangular impulse response transfer function	64
3.3 Another derivation of the deconvolution formula for a triangular impulse response system	68
3.3.1 Accuracy and limitation of the deconvolution formula found in 3.3	70
3.4 Developing the partial deconvolution concept	73
3.4.1 A discrete realisation of the partial deconvoluiton formula	80

3.4.2 The accuracy and the limitation of the partial deconvolution formula found in 3.4.1	81
3.5 A partial deconvolution formula in the case of a gaussian impulse response system	83
3.5.1 A discrete realisation of the partial deconvolution formula	87
3.5.2 A discrete realisation of a gaussian impulse response transfer function by a fourier series	90
3.6 Summary and Conclusions	91
<u>4 The Estimation Problem</u>	98
4.1 Introduction	98
4.2 Modelling of the beam processing in the spectrophotometer	103
4.3 Some features of the error function	109
4.4 The solution in the case of small perturbation and a perfect resolution spectrophotometer	121
4.4.1 Derivation of the estimation variance	128
4.4.2 Interpretation of the expression of the variance	132
4.4.3 The modified signal power	139
4.5 The general concentration estimation problem (imperfect resolution)	142

4.5.1 The modified signal power in the case of imperfect resolution	147
4.6 The effect of the resolution on the estimation error	149
4.7 Summary and conclusions	153
<u>5 Simulation and Results</u>	155
5.1 introduction	155
5.2 The estimation algorithms	157
5.2.1 The maximum likelihood estimation algorithm	160
5.2.2 The sequential estimation algorithm (for high resolution signal)	164
5.3 Simulation and results	174
5.3.1 introduction	174
5.3.2 Accuracy of the estimation as a function of the S/N ratio	177
5.3.3 The estimation variance as a function of the cell length	184
5.4 Creating the bank of signal and the supporting package of programs	187
5.5 Summary and conclusions	196

6 Conclusions and Recommendations 197

6.1 Summary and conclusions 197

6.2 Recommendations for further research 203

6.3 Practical applications in spectrophotometric instrumentation 206

6.3.1 Some considerations in implementing the sequential estimation algorithm 207

References 212

Appendices

Appendix 1 Derivation of the set of equations for implementing the inverse transfer function with a triangular impulse response

Appendix 2 Derivation of the set of equations for implementing the inverse transfer function with a gaussian impulse response

Appendix 3 Tables for deconvolution formulas in the case of gaussian and triangular impulse response systems

Appendix 4 The estimation problem in the linear case

Appendix 5 The formula for increasing the slit width

Appendix 6 Interval of confidence

LIST OF TABLES

Chapter 1

1	Some characteristics of ir detectors	21
---	--------------------------------------	----

Chapter 2

1	Differential equations, in operator form, of some symmetric impulse response systems	55
---	--	----

Chapter 5

1	Error in estimation 2. and 20. PPM of CH ₄ in the presence of 1000. PPM of H ₂ S	179
2	Comparison of the error, in estimation 2. and 20. PPM of H ₄ , between the MLE and SE algorithms	181

Appendix 3

Tables for deconvolution formulas in the case of gaussian and triangular impulse response systems

1		3 - 2
2		3 - 4
3.1		3 - 6
3.2		3 - 7
4		3 - 9

Appendix 5

1	Coefficients of the operator that multiply the slit width by M	5 - 2
---	--	-------

LIST OF FIGURES

Chapter 1

1	Schematic diagram of the spectrophotometer	12
2	Three states of the entry slit image demonstrate how the slit transfer function is found out	14
3	A schematic diagram of an analogue divider which is equivalent to the spectrophotometer servo loop	29
4	The servo loop in the dual beam spectrophotometer	29

Chapter 2

1	Some general symmetric finite impulse response	41
---	--	----

Chapter 3

1	The spectrophotometer response to an impulse (sharp absorption line)	60
2	Error graph of deconvolution formula for a triangular impulse response transfer function	72
3	Error graph for the partial deconvolution formula (triangular impulse response)	82
4	Error graph of partial deconvolution formula (gaussian impulse response)	89
5	Increasing the resolution of a given signal by applying a partial deconvolution formula (gaussian impulse response)	96
6	The proposed concentration estimation algorithm	97

Chapter 4

1	A schematic diagram of the estimation process	98
2	Block diagram of the process in the spectrophotometer	104
3	The error in the gradient is smaller the smaller is the ratio $\nabla E_r / \nabla E_d^2$	116
4	Geometrical interpretation for the estimation problem in the linear case	137

Chapter 5

1	A general simulation model	156
2	Block diagram of the simulated MLE algorithm	163
3	Block diagram of the simulated SE algorithm	173
4	rms error in estimating the concentration of CH ₄ in the presence of H ₂ S (1000. PPM) using the MLE technique	180
5	rms error in estimating the concentration of CH ₄ in the presence of H ₂ S (1000. PPM) using the SE technique	182
6	Comparison of the rms error and the estimation variance in applying the MLE	183
7	Comparison of the rms error and the estimation variance in applying the SE	183
8	The error graph (estimation of CH ₄ in the presence of H ₂ S) indicates the location of the optimum cell length	185
9	Recording the spectrum on an analogue tape recorder (7"/sec, B.W.: DC -1.25kHz)	189
10	The set up for transferring the analogue signal to a digital tape	190
11	Schematic diagram of the sampled signal being transferred from the PDP 11 to a digital tape	190
12	A plot of the stored signal after unpacking it	191
13	A plot of the stored signal after unpacking it	192
14	The result output of applying GETSLIT and CONVOLV to a stored signal (convolving with 2.1/cm triangular slit)	194
15	The result output of applying GETSLIT and CONVOLV to a stored signal (convolving with 14.1/cm triangular slit)	194
16	Adding gaussian noise to a stored spectrum (using GETNOIS)	195
17	Simulating spectra with different concentrations (using ATTEN)	196

Chapter 6

1	Connecting a u-processor unit to the spectrophotometer for performing concentration estimation	208
2	Timing of the inter-connection between the u-processor and the spectrophotometer	209

ACKNOWLEDGEMENTS

I would like to extend my thanks to my thesis supervisor, Professor L. Finkelstein, not only for his advice and guidance but also for suggesting this exciting research topic to me. His encouragement and support throughout the work have been very much appreciated.

I am indebted to [REDACTED] for his invaluable discussions and suggestions.

Thanks are also due to the SIRA Institute for allowing me to use their facilities, and especially to [REDACTED] for his generous help in the SIRA laboratories and his general advice.

Also, I would like to express my thanks to Fischer Control Ltd for their financial support and the interest they have shown in this research.

Finally, I am in especial debt [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] for their kind and generous help in financing my studies throughout all this research period, with particular thanks to [REDACTED] [REDACTED] [REDACTED]

DECLARATION

The author grants powers of discretion to the City University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

The main objectives of the research programme were :

(1) Developing a discrete mathematical model describing the electro optic process in the spectrophotometer which includes three subprocesses :

- (a) The absorption process (Beer - Lambert law)
- (b) The slit transfer function
- (c) The detector noise

The model provides the theoretical tool in creating signal processing techniques that are applied to spectrophotometry instrumentation, in particular to deficiencies like low resolution, noise and distortion.

(2) Studying and developing signal processing techniques that estimate the concentration of one compound in a mixture of several.

The estimation techniques that are based on the model created in (1), exploit well known concepts in statistical estimation theory (i.e. The Maximum Likelihood principle) and give as a result the vector $\hat{C}$ where each component $\hat{C}_j$ is the estimate of the compound j .

A discrete formulation has been chosen for the above in order to achieve easy computer implementation both in testing the technique itself and in later practical use of the algorithms developed.

The first step in modelling the electro optic process in the spectrophotometer was to derive the differential equation of a process given its impulse response in a polynomial form or in a gaussian pulse shape. The differential equation found is given in operator form using operators like D the differentiator or δ the central difference.

The second step was implementation of the differential equation in a discrete form. A standard form of finite series in the discrete operator (the average operator) $j=1,10$ was chosen and derived. The implementation was carried out on both the forward transfer function and the inverse transfer function (with both gaussian and triangular impulse response). The implemented transfer functions were tested for their accuracy as a function of the signal frequency. The concept of the partial deconvolution formula was suggested and implemented. The idea that underlies it is to implement a discrete transfer function that reduces the slit width by only a certain amount and thus allows a trade off between the noise level and the resolution.

In the second part of the work two estimation algorithms are proposed. The two algorithms are based on the Maximum Likelihood Principle which in our case (noncorrelated gaussian noise components) yields the minimum rms error. The first algorithm (called the MLE) is treated as an optimization problem in n dimension. The result, the estimation vector $\hat{C}$, gives the minimum rms error. The algorithm is

based on an iteration process where a correction vector ΔC is found at each step until no further correction is possible. The equation for the correction C and for the error (the estimation variance) $\Delta C \Delta C'$ are derived. It is shown that the expression of the error has the same form as in the linear case.

The algorithms were tested with a signal spectrum of a mixture of two gas compounds (two concentrations) and with a perfect resolution (the effect of the spectrophotometer slit width is ignored) over a range of several S/N ratios and concentrations. Also, the calculated rms error was compared to the value given by the theoretical expression derived. It was shown that the two are in good agreement.

The second algorithm (called SE) was developed as a sequential algorithm in order to provide easy and fast hardware implementation. The algorithm based on linearization of the estimation problem and on the Maximum Likelihood principle, processes the signal samples as they enter the signal processing unit, and thus avoids the need for storing the signal.

Since both algorithms are based on the MLE principle it means that the algorithms have a built-in filtering mechanism. Therefore, the spectrophotometer scanning speed is not limited by the time constant of an additional filter but only by the detector time response.

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols

a_n	general scalar coefficient
A_i	twice the error between the absorption given by the current estimated concentration and the absorption given by the actual concentration at a wave number v_i
ACR_{ijm}	accumulative element R_{ij} based on a signal with m samples
ACR_{yjm}	accumulative element R_{yj} based on a signal with m samples
b_n	general scalar coefficient
c_j	concentration of compound j in the mixture
C	vector of concentrations
C^0	vector of actual concentrations
$\hat{C}$	vector that estimates C^0
D	differentiator operator
e	error vector between the actual sampled spectrum and the simulated spectrum
e_i	component of e at wave number v_i
$E^2(C)$	error function
$E_d^2(C)$	deterministic part of the error function
$E_r(C)$	random part of the error function
f	frequency
$f(v_i, C)$	simulated absorption at a wave number v_i
$g(v_i)$	modified signal at wave number v_i
$h(v)$	impulse response function (discrete or analogue)
$h_a(v)$	impulse response of an analogue system
$I(v)$	intensity of the ir beam at wave number v
ir	infra red

$\left. \begin{matrix} i \\ j \\ k \\ l \\ m \\ n \end{matrix} \right\}$	general subscripts
jw	imaginary number with absolute value of w (j , in this case, is the square root of (-1))
$\left. \begin{matrix} L \\ M \\ N \end{matrix} \right\}$	constant integers
J	Jacobian matrix
L	cell length
N_0	noise power
n_i	measurement noise (random variable) at a wave number v_i
q_i	gradient of the simulated absorption ($f(v_i, C)$) with respect to C
R	correlation matrix in the general case (non-linear)
R_{ij}	element ij of R
R_{ij}^*	element of R^* , the correlation matrix in the linear case
S_i	modified signal power of compound i
S_i^*	signal power of compound i in the linear case
T	sampling period
u_j	average operator that operates on two samples with T_j sampling periods between them
v_i	wave number v at the sampling number i
w	angular frequency
$x(v_i)$	absorption at wave number v_i
$y(v_i)$ y_i	measured spectrum at wave number v_i
Y	measured signal (spectrum) as a vector with components y_i
z	complex number

$\alpha_j^0(v_i)$	absorption coefficient of compound j at wave number v_i
$\alpha_j(v_i)$	absorption coefficient multiplied by the cell length
λ	wave length
Δ	increment operator
δ	central difference operator
∇	gradient operator
ρ	correlation coefficient
σ^2	variance of a gaussian distribution
$\prod_{i=1}^N a_i$	product of all the a_i , $i = 1, \dots, N$

Abbreviations

S/N	Signal to Noise ratio
ir	infra red
MLE	Maximum Likelihood Estimator
SE	Sequential Estimator

Special symbols

'	transpose of a matrix or vector
AB	when A and B are vectors it is understood that the operation between them is a product in the usual sense, namely between the corresponding components of each vector. The result is a scalar.

Throughout this work an integral or summation sign without boundaries means integral or summation over all the non-zero elements

0. Introduction

This thesis represents an account of a theoretical and experimental study of the application of signal processing to infra red (ir) spectrophotometry.

Ir spectrophotometry has become an indispensable tool for the analytical chemist in studying, from both quantitative and qualitative aspects, the chemical properties of a substance(s). With the tremendous development of minicomputers and microprocessors, ir spectrophotometry has been advanced as well. Thus it is not uncommon to find nowadays an on-line computer data system - equipped with a programs package designed to deal with the different aspects of chemical analysis - linked to a spectrophotometer and performing a variety of analytical tasks like wave number calibration, base line normalization, finding peak absorption etc.

The main objectives of the research programme were :

(1) To establish a discrete mathematical model describing the electro optic process in a spectrophotometer; the description includes three subprocesses :

- (a) The absorption process (Beer - Lambert law),
- (b) The slit transfer function,
- (c) The detector noise.

The model aimed to provide a theoretical tool for use in further

research in spectrophotometry instrumentation, and into spectrophotometer deficiencies like low resolution, noise and distortion.

(2) To study and develop a technique that estimates the concentration of one compound in a mixture of several.

The estimation technique, based on the model created in (1), exploits well known concepts in statistical estimation theory (i.e. the maximum likelihood estimation) and gives as a result the vector $\hat{C}$ where each component $\hat{C}_j$ is the estimate of the compound j .

A discrete formulation has been chosen for the above in order to achieve easy computer implementation : both in testing the technique itself and in later practical use of the algorithms developed.

Ir spectroscopy has been applied before in measuring gas concentration. Most of the commercial instruments are based on non dispersive spectroscopy [12], where the power of the transmittance spectrum of the gas with the unknown concentration (target gas) is compared to the power of the transmittance spectrum of the reference (the same gas with known concentration). The fact that the comparison is made between the power of the two spectra rather than their shape means :

(i) There is no need for a dispersive system. Therefore the measuring system is fairly simple.

(ii) The S/N ratio at the output of the measuring system is much higher than in the case of a dispersive system. Consequently, the accuracy in

estimating the concentration is improved.

(iii) The main disadvantage of this technique is its inability to deal with more than one absorbent gas in a mixture and its lack of accuracy in measuring the concentration of one compound in the presence of another.

A laboratory technique that has shown a very good sensitivity in measuring a very low concentration is called masking correlation [3,9,13]. In this technique a spatial mask allows the dispersed spectrum to reach the detector only in specific wavelengths where the transmittance spectrum is not zero (match position). Misalignment of the mask (by shifting it slightly across the ir beam) will reduce substantially the amount of ir power that reaches the detector (unmatch position). By shifting it, repeatedly, between the two positions - match and unmatch - the detector output is modulated. The 'depth' of the modulation is a sensitive function of the number of molecules that absorb the ir beam power, and, consequently, provides a mean to measure the concentration of the gas.

The technique developed here is different from the techniques currently used in the following aspects :

(i) The technique is based on a software signal processing algorithm that accepts a regular spectrophotometer output, processes it, and gives as a result the concentration of its compound in the mixture. No hardware implementation is needed.

(ii) The technique developed is based on the knowledge of the spectrum

of the gases in the mixture rather than on some characteristic absorption lines of the compound. Consequently the technique is fairly general and can be applied to any mixture of gases that have spectra in the measured wavelength range.

The technique developed is a product of the theoretical investigation which is reported here. Among other objectives of the research, the technique was simulated on a computer in order to verify some of the theoretical results.

In measuring the concentration of a single compound in a mixture three main difficulties arise. Firstly, random noise is generated at the detector and is added to the signal output which represents the transmittance of a given mixture. Secondly, the transmittance at a given wave length is the product of the partial transmittance of each compound, which in turn, depends exponentially on its concentration level. Therefore, the transmittance in a given wave length is a nonlinear function of several variables (concentrations). In other words, in the general case of concentration estimation, the other gases in the mixture mask the desired signal and attenuate it. This attenuation is a nonlinear function of the concentration(s). Thirdly, the output from the spectrophotometer, due to the finite slit width represents some weighted average (convolution) of the transmittance over a local range (slit width) of wave length. That is, the output of the spectrophotometer represents a contour - distorted by the instrument transfer function - of the true transmittance. Other inaccuracies are caused by: aberration, diffraction, improper alignment of the

different optics and the nonlinearity of the detector and the electronic process.

An earlier study which underlies the research work has included an investigation and modelling of the dual beam spectrophotometer, its slit transfer function and its noise sources. A bank of signals and a programs package have been developed. The bank of signals consists of several transmittance spectra of some gases, energy and noise profiles of a spectrophotometer. The package consists of programs that perform the following functions: retrieving, packing, unpacking a data file and simulating a variety of cases of different resolutions, concentrations and noise levels.

The investigation undertaken was a development in the application of the principles of signal processing and includes contributions which can be summarized as follows:

- (a) formulating the discrete estimation equation of a single compound in a mixture of several compounds. This equation includes the effect of the spectrophotometer slit transfer function, the absorption process (Beer - Lambert law) and the noise generated in the detector.
- (b) Solving the above equation for the case of perfect resolution spectrum (the effect of the slit transfer function is neglected) and deriving the formula of the estimation variance (estimation rms error).
- (c) Extending the solution of the estimation equation to include the effect of a wide slit width (low resolution) transfer function on the

spectrum.

(d) Estimating procedures based on two search algorithms the first based on the well known Gauss search technique ,the second, based on a sequential technique that was developed here. These techniques have been investigated, in off-line simulations, with a variety of signal to noise ratios (S/N-Ratio), and concentrations.

(e) Developing of a convolution/deconvolution technique based on control of the slit width of a given transfer function. This technique (partial deconvolution) enables us to increase the resolution of a given signal by a desired degree, and thus to control the effect of the increasing noise which, usually, degrades the signal in operation such as deconvolution.

1. Principles of IR Absorption and Spectrophotometry

In this chapter the absorption process of absorbing ir energy by gas molecules is discussed as well as some principles that are implemented in the spectrophotometer.

The approach adopted here is the approach sometimes called a systems approach where the actual mechanism in the process is non relevant but only its final effect (in our case) on the ir beam. This approach is extremely useful in analysing linear systems where the ratio of the input/output intensities, as a function of the frequency (wavelength), gives a full description of the system. As will be shown, the process in the spectrophotometer can be considered as a linear process.

The same approach is used for describing the dual beam spectrophotometer as an instrument. Its main feature is the elegant solution for two main sources of error : random noise at low frequencies (drift) and the dependence of the output beam intensity on the input beam intensity.

Also, a classification of ir detectors and noise sources is given and discussed.

Ch. 1

1.1

The nature of ir absorption

In ir spectroscopy the absorption or the emission of radiation is primarily due to changes in the vibrational energy of the system through which the radiation is passing (absorption) or from which it is emanating (emission). The energy, E , of the light quantum absorbed or emitted is given by Einstein's formula

$$E = h\nu$$

where h is Plank's constant (6.625×10^{-34} Joule sec) and ν is the frequency of the radiation. Wave mechanics shows that the quantum energy must exactly equal the difference between the initial and the final energy levels of the system. Thus existence of absorption or emission bands implies that we have a method of observing directly the difference between energy levels, and therefore, providing the analytical chemist with clues to the molecular structure of the matter from both quantitative and qualitative aspects [1].

The ir radiation range covers the electromagnetic wave range of $1\mu\text{m}$ – $1000\mu\text{m}$. The relation between the velocity - c , the wave length- λ and the frequency- f is given by the well known formula :

$$c = \lambda f$$

the wave number - ν is defined by :

$$\nu = 1/\lambda = f/c \quad [1/\text{cm}]$$

The ratio between an incident beam intensity and an output beam intensity from a substance, at a wave length , may be regarded as the transfer function of the absorption process. In general (excluding from our discussion phenomena like emission spectroscopy and Raman effect) an output at a wave length may occur only if there is an input at this wave length. In other words, an input at a given frequency does not generate other frequencies - this may be considered as a definition of a linear system. Thus the absorption process may be included in the optical transfer function of the spectrophotometer which is also described by a linear function [2].

The linearity feature of the absorption process is exploited in the spectrophotometer. By exposing a substance to a beam consisting of the sum (integral) of the intensities over a given range of frequencies, but measuring the response (output) at a single frequency (in practice, a narrow band of frequencies) , the superposition principle, which holds only in a linear systems, is applied.

The mathematical expression for the input-output relation of the absorption process is given by Beer - Lambert law [3]

$$i_{out}(v) = i_{in}(v) \exp(-ca(v)) \quad (1.)$$

where

$i_{out}(v)$ - is the output intensity at a wave number v ,
 $i_{in}(v)$ - is the input intensity at a wave number v ,
 c - is the concentration of the absorption material,

$$a = a^{\circ}L \quad (1a.)$$

where :

$a^{\circ}(v)$ - is the material absorption coefficient at wave number
 v

L - is the cell length

That is, it is assumed that the ir beam is absorbed by molecules in a cell with a length - L . The dependence of the relation on v - the wave number (or f - the frequency) only, reflects the linear nature of this

process.

Assuming an input beam with a constant power density, we have

$$i(v) = i : v_1 < v < v_2 \quad (2.)$$

In the case of a mixture with several compounds, concentrations $c_1, \dots, c_n$ and absorption coefficients $a_1(v), \dots, a_n(v)$, the output beam, after absorption has the intensity at a wave number v

$$i_{out}(v) = i \cdot \exp(-\sum_j c_j a_j(v)) \quad (3.)$$

Sampling the signal at v_1, v_2, v_M , we have :

$$i_{out}(v_i) \equiv i_i = i \cdot \exp(-\sum_j c_j a_j(v_i)) \quad (3a.)$$

$$i = 1, M$$

1.1.1 The spectrophotometer transfer function

In limiting our consideration to ir absorption (assuming that emission in the wavelength of interest is negligible) only, we have suggested in 1.1 that the input/output relation, of the absorption process, at a wave number ν , may be considered as a linear process such as may the optical process in the spectrophotometer. Consequently, the spectrophotometer, as a whole, has a linear transfer function. Thus, in the spectrophotometer a substance is exposed to a wide wave length beam, the output beam is dispersed and a narrow exit slit allows a narrow wave length beam to reach the detector. The detector converts the radiation to an electric signal (see Fig.1).

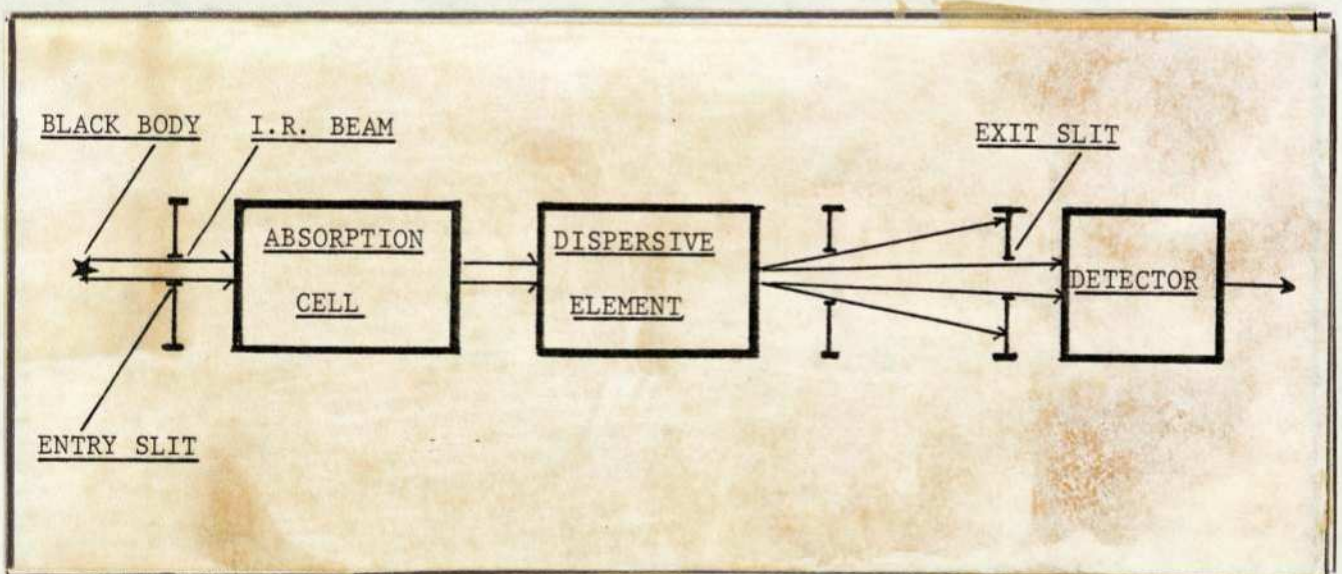


Fig. 1 Schematic diagram of the spectrophotometer

By shifting the exit slit across the dispersed beam the output, which is a function of the slit location corresponding to some wave length, describes the transmittance of the material as a function of the wave length.

The transfer function of the optical part in the spectrophotometer is the output/input ratio as a function of the wavelength for an input which consists of a single frequency [4]. In the case of equal entry and exit slits the output is maximum when the image of the entry slit covers exactly the exit slit, since the entire input beam intensity (minus optical losses) reaches the detector (see Fig.2). Shifting the exit slit by, say, a half slit width so that half the entry slit image, only, reaches the detector, the output will be reduced to half. Shifting it by the full slit width, no output will reach the detector. It can be concluded that the power reaching the detector varies linearly with the position of the exit slit - from maximum to zero.

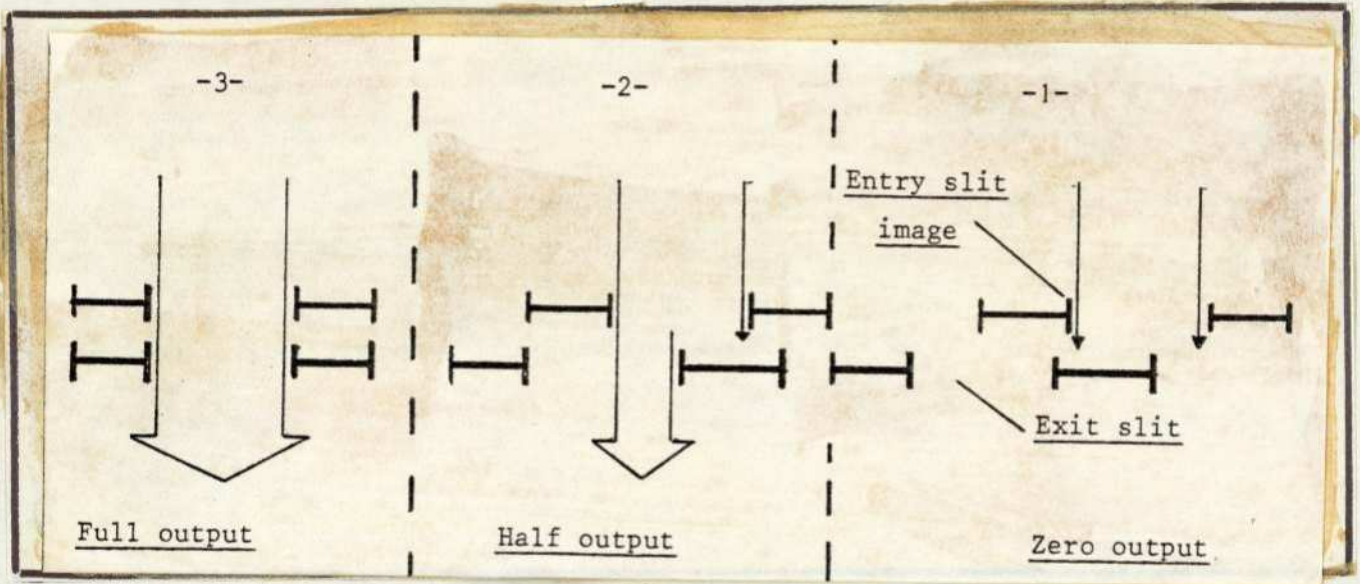


Fig. 2 Three states of the entry slit image demonstrate how the slit transfer function is found out

Since the shifting may occur in both directions, with the same result, we conclude that the transfer function of the optical system in a spectrophotometer (free of aberration and diffraction) with an equal entrance and exit slit width is a triangular function given by

$$h(v) = \begin{cases} (v_s + v) \frac{1}{v_s^2} & -v_s \leq v \leq 0 \\ (v_s - v) \frac{1}{v_s^2} & 0 \leq v \leq v_s \\ 0 & |v| > v_s \end{cases} \quad (4.)$$

In the case of non equal slits the transfer function is a symmetrical trapezoidal function, with a lower base equal to the sum of the slits width and the upper base equal to the difference (in absolute value) of the slits. A more rigorous treatment should include phenomena of aberration and diffraction, but in the practical case of a large slit width compared to the wave length, these phenomena are insignificant. Abberation and misalignment of the optics can be described with a gaussian transfer function [4]

$$h(v) = \frac{2}{w_s + k} \left(\frac{\ln 2}{\pi} \right)^{1/2} \exp \left(- \frac{4 \ln 2}{[w_s + k]^2} v^2 \right) \quad (5.)$$

w_s - is the mechanical slit width (both equal)

k - a constant varies slightly with the wave length.

The linearity of the process in the spectrophotometer implies that the input to the detector is given by the convolution integral

$$i_o(v) = \int_{-v_s}^{v_s} h(s) i(v-s) ds \quad (6.)$$

Substituting (3), we have

$$i_o(v) = i \int_{-V_s}^{V_s} h(u) \exp[-\sum_j c_j a_j (u-v)] du \quad (7.)$$

1.1.2 Infra red detectors and noise sources

The ir beam in the spectrophotometer undergoes some optical process, reaches the detector, is then converted into electronic signal, amplified, filtered and recorded. In these processes a noise is added to the signal. The signal reaches its weakest level at the input to the detector, where only small part (see Fig.1) of the signal passes the exit slit. Any noise added at that point will have major effect on the output, in other words, the signal to noise ratio at the output (S/N) is defined at the detector.

Any physical effect that converts an ir beam into an electric signal may be used in ir detection. The theory that deals with ir detectors is a field in its own right [5,6], we shall touch this field briefly, from aspects which concern this work only.

Ir detectors can be divided to two main classes

(a) Thermal detectors

In this class the power of the ir beam is converted to heat that raises the temperature of a device which, in turn, converts the difference in the temperature to emf. Some of the phisical effects that are exploited are

(i) Bolometers

In this group the change in the temperature causes a change in the conductivity of the device.

(ii) Thermocouples

In this group, two different materials create two junctions. One junction is kept in constant temperature, while the other absorbs ir beam and therefore its temperature is changed accordingly which, in turn, generetes an emf.

(iii) The Golay Cell

In this group, the absorbed energy of the ir beam changes the volume of a gas cell which, in turn, causes a deflection in a membrane.

This class of detectors are non selective, that is, their response is independent of the wave length, their impedance is low (1-100.ohm), their time constant, due to their heat capacity is high (milli- sec) and they have a very low noise level. The Golay

Cell is commonly used in spectrophotometry.

(b) Photoconductive detectors (quantum detectors)

In the ir range where the photon energy is insufficient to produce photon- emission, that is, the absorption of a photon does not lead to the ejection of an electron - the mechanism that leads to detection of ir beam is the creation of the electron- hole pair by a photon. With the development of the semiconductor materials in the last two decades, this class of the detector has reached high level of sophistication. A variety of semiconductors has been developed for this purpose. By controlling the level of impurities of different materials in the crystal structure of Germanium or Silicon the designer may achieve different charactatristic behaviour of the detector.

Compared to the thermal detector, the photoconductive detector has a short time constant (μ -sec), high impedance (meg-ohm), higher noise level and they are wave length dependent. This last feature makes them, still, unsuitable to application in general spectrophotometry (see table 1).

In order to compare the quality of different detectors, the following parameters are defined

The detector responsivity - defined as

$$R = u_s / P$$

where

u_s - is detector output

P - is the ir beam input power.

Given the slit area and the beam intensity, the signal output is given by the above equation. The noise power is translated to the equivalence beam power by

$$NEP = P \frac{u_n}{u_s} \frac{1}{\Delta f}^{1/2}$$

The noise is measured over a specific band width of frequency. In the ideal case the noise power density $u_n^2/\Delta f$ is constant (white noise).

The noise in most of the detectors is proportional to the square root of the detector area. This can be explained by regarding each unit area as an independent detector that generates the same noise power. The total noise power is the sum over the whole area (sum over independent random variables).

D^* (D-star) is defined as :

$$1/D^* = NEP/(Ad)^{1/2}$$

where Ad is the detector area

$1/D^*$ is the equivalent of the square root of beam power that is generated by a detector with a unit area over a unit band width.

Table 1. is a comparison of some of the ir detectors

Detector-material	Useful wave-length range [um]	Resistance [ohm]	Time-constant [u sec]	D* [cm Hz 1/W]
Thermocoupl	1-40	1-10	25000	6-15E8 (5)*
Golay Cell	5-1000	-	20000	5-10E8 (10)
Silicon	.5-1.05	.1-1.10E6	100.	1-5E12 (90)
Lead Selnide	.9-4.6	1-10E6	2	1-4E9 (800)

*

At his frequency (Hz)

Table 1. Some characteristics of ir detectors

From the above table it is evident that solid-state detectors can detect much weaker ir source, but only at a certain wave length. The audio frequency range that the thermal detectors are effective is only a few Hz. This imposes a restriction into the scanning speed with which a range of wave length may be scanned.

Although the noise in the thermal detectors is mainly due to a thermal noise, the noise level of the first electronic stage (coupling

and preamplifier) has to be taken into account as well. Therefore, noise source mechanisms active in semiconductors has also be consiered. Some of the main noise sources are [2]

(a) Johnson noise

Johnson noise or thermal noise is charachteristic of the flow current through an electrical resistance. The noise rms current - i_n is given by :

$$i_n = 4kT\Delta f/R$$

where

k - is Boltzman's constant

T - is the absolute temperature

Δf - is the band width of frequencies

(b) Generation - recombination noise

This noise is due to the statistical fluctuation of the number of carriers in a semiconductors.

$$i_n = kI_o \frac{T\Delta f}{N[1 + f T]}$$

where

I_0 - is the average current

$\bar{N}$ - is the average number of carriers

T - is the carrier life time

f - is the audio frequency

(c) Current noise

Also known as 1/f-noise, its power is proportional to the current

$$i_n^2 = k_1 I^\alpha \frac{\Delta f}{f} \beta$$

k_1 , α and β are constants describing the size and the shape of the detector.

Its precise mechanism is not known but it is associated with lattice dislocation, surface texture and similar features.

The noise in the detector and semiconductors is a random process which can be described by its statistical features. Each noise mechanism is the result of large number of independent phenomena, like generation - recombination of electron hole-pair, that contribute to some average value and to a fluctuation from this average. Consequently, the Central Limit theorem [7] - which states that in such a case the probability density function which describes this process is given by a gaussian distribution - may be applied.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \quad (8.)$$

where

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x) dx \quad (8a.)$$

We assume

$$\bar{x} = \int_{-\infty}^{\infty} x p(x) dx = 0$$

We assume that the noise has a stationary nature (weak sense). That is, its mean value is zero everywhere and its correlation is, therefore, a function of the time interval between two points and not their location [8]

$$R(t_1, t_2) = R(t_2 - t_1) = R(\tau) \quad (9.)$$

The power density of the noise is given by

$$S(j\omega) = \int_{-\infty}^{\infty} R(s) \exp(-j\omega s) ds \quad (9a.)$$

Also, the ergodicity of the mean and the correlation is assumed (A good discussion of the subject is given in [8])

$$\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt \quad (10.)$$

$$\overline{x(t)x(t+\tau)} = R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) dt \quad (10a.)$$

That is, the statistical mean and the statistical correlation can be studied by the time mean and time correlation.

1.2

The dual beam technique

The dual beam technique is a technique that incorporates in a servo-loop two functions in order to eliminate two main errors :

(i) Drift (low frequency noise)

In practice, it means that over a period of time, the reading of the detector fluctuates randomly, slowly. It is caused by the 1/f-noise mechanism (1.1.2).

(ii) Dependence of the output on the input beam intensity

From (3a.) it follows that the output from the detector is a function of the input intensity :

$$i_o(v) = i(v) \exp(-\sum_j c_j a_j(v))$$

The input intensity is provided by ir source which varies as a function of wave length, the age of the source, and also from one ir source to another

These inaccuracies are reduced by the following process :

(i) Converting the 'DC' input beam to 'AC' by 'chopping' it. Thus the main beam flux is shifted from 'DC', where the noise level is too high, to the chopping frequency (first harmonic), where the detector noise level is lower. It may be noticed, that the effect of the absorption on the beam intensity is to modulate it :

$$i_{out}(v, t) = [i(v) \sin(w_0 t)] \exp(-\sum_j c_j a_j(v))$$

The fourier transform of the signal above is shifted to the frequency w_0 . Since the frequency spectrum (audio frequencies which depend on the spectrophotometer scanning speed) of the signal, without modulation, has only low frequencies around DC, the modulated signal has a band around the frequency w_0 .

Band pass filtering the signal with a filter of $2w_b$ bandwidth the noise can be represented now as:

$$S_o(jw) = \begin{cases} S_n(jw_0) & |w - w_0| \leq w_b \\ 0 & |w - w_0| > w_b \end{cases}$$

The noise power is given by:

$$N_o \cong 2w_0 S(jw_0) \quad (11a.)$$

Thus by modulating the signal its power was transferred from DC to the frequency w_0 (more precisely, to a bandpass around w_0) where the detector noise is weaker [7].

More sophisticated techniques are based on converting the input beam, from a 'DC' flux, to a sum of audio frequencies, where each frequency is assigned to a different wave length (frequency or coding multiplexing). The exit slit is no longer needed. The spectrum, over a wave length range, is accepted now simultaneously, a fact, that contributes, substantially, to the improvement of the output S/N (Fillgett Advantage) [9].

By reducing ω the noise is reduced proportionally (11a.), which also imposes a restriction on the speed the wave length range may be scanned.

(ii) Dividing the output intensity by the input intensity. Thus eliminating the dependence of the signal output on the input intensity. An analogue servo loop that performs division is described in Fig. 3. the output of the multiplier in the feedback loop is controlled such that it is equal to the product- xA (x times A). If one of the multiplier input is ' x ' its second input is the unknown factor ' A '. In the spectrophotometer ' x ' is the input intensity and ' xA ' is the output intensity. Fig. 4 demonstrates how this divider is implemented in the spectrophotometer [14].

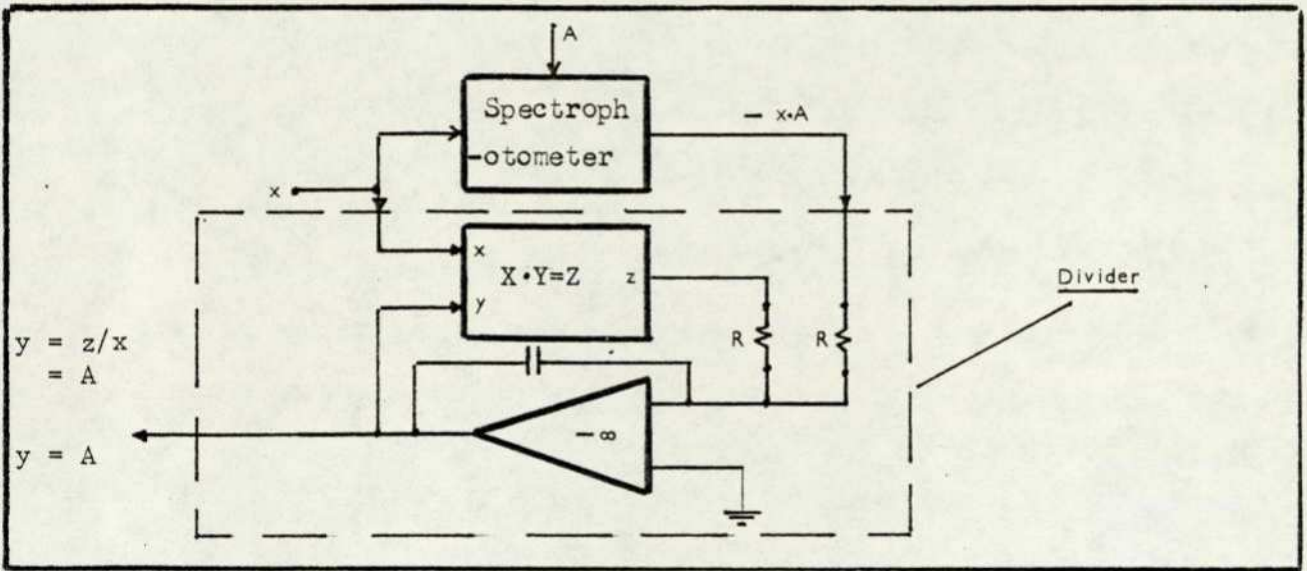


Fig. 3 A schematic diagram of an analogue divider which is equivalent to the spectrophotometer servo loop

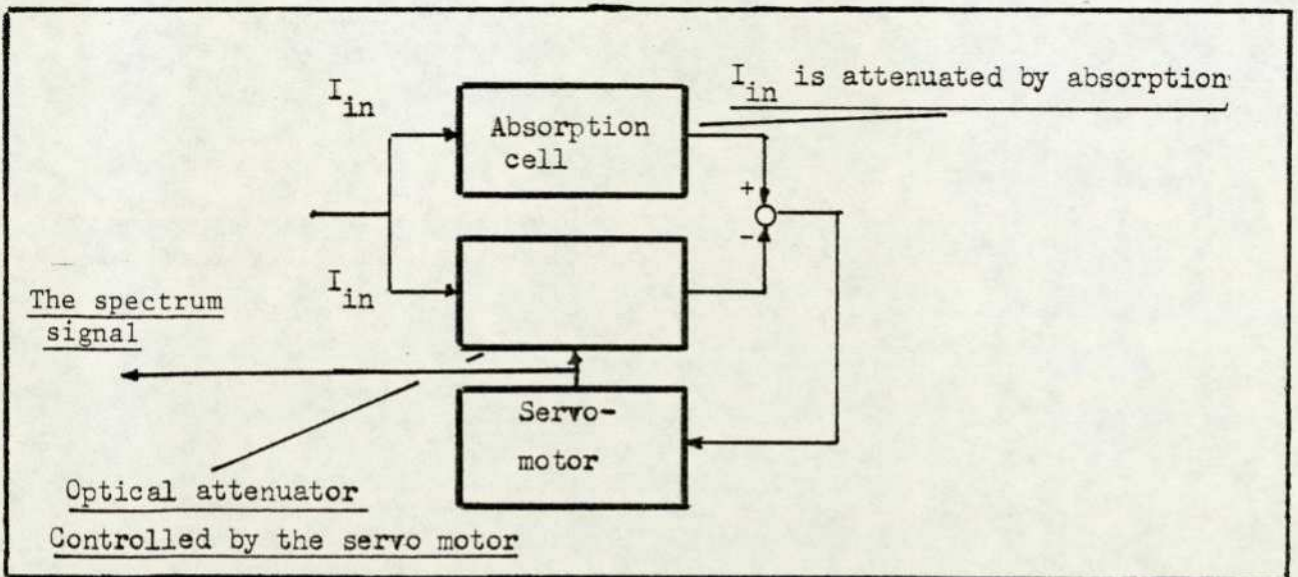


Fig. 4 The servo loop in the dual beam spectrophotometer

The analogue loop performs division and therefore enables to find the unknown ratio $x_A/x = A$ (see Fig.3). In a steady state the multiplier output - Z is equal x_A , since the input to the integrator is zero (virtual ground). Therefore the y input to the multiplier is

$$y = z/x = -x_A/x = -A$$

This relation applies, also, when the input is changed. In the spectrophotometer loop the servo motor replaces the integrator, the optical attenuator replaces the multiplier and the summing point is replaced by the demodulator (which has to be used, since the signal was modulated).

Neglecting the noise component, the output is :

$$\text{Spectr. (output)} = i_{\text{out}}/i_{\text{in}} = \exp(-\alpha(v)) \quad (11b.)$$

It may be noticed that although in using the dual beam technique, the input intensity does not affect the output, it does affect the gain loop which, in turn, affects the time response of the loop. Thus in a case of strong absorption the gain loop is reduced substantially. As a result, the instrument responds slowly to a change in the absorption. The same effect is observed when the slit width is reduced. A detail

analysis of this loop should include this non-linearity (multiplier) feature.

2. The Differential Equation of the Spectrophotometer Slit Transfer Function

2.1

Introduction

It is well known that the output - $y(t)$ of a linear system with an impulse response $h_a(t)$ and an input $x(t)$ is given by:*

$$y(t) = \int h_a(t-s) x(s) ds \quad (1.)$$

In our case, the function $y(t)$ (the spectrum, the signal from the spectrophotometer) is sampled every T sec, the impulse response $h_a(t)$ of the spectrophotometer is known (see Sec.1.1.1). Given a true spectrum $x(s)$, the output from the spectrophotometer is :

$$y(nT) = \int h_a(nT-s) x(s) ds \quad (1a.)$$

$$n = 1, \dots, L$$

We would like to find a discrete process that produces the same series $y(T), y(2T), \dots, y(1T)$ for a given input series $x(T), x(2T), \dots, x(1T)$:

$$y(nT) = \sum_j h(T(n-j)) x(jT) \quad (2.)$$

*

See Special symbols p. xv

If a discrete system with such an impulse response exists, it is said that the system is a digital simulator of the analogue system described by $h_a(t)$.

A digital simulator plays an important role in simulating and processing, digitally, analogue signals. In our case, where the signal is a spectrum of a mixture of gases the digital simulator has two main functions to fulfil :

(i) In the estimation of concentration level(s) of compound(s) in a mixture of gases.

Beer - Lambert law, which describes the dependence of the spectrum on the concentration(s) of the compound(s) in the mixture, is valid for a monochromatic ir beam. In practice this demand may be satisfied if the given spectrum is a spectrum with a high resolution. That is, the spectrophotometer transfer function affects very little the shape of the absorption lines, consequently, the difference between $[x(T), x(2T), \dots, x(MT)]$ and $[y(T), y(2T), \dots, y(MT)]$ is negligible. Therefore, given a high resolution discrete record of a spectrum signal, Beer - Lambert law may be applied directly, to the given signal, in order to find the concentrations of the compounds in the mixture. In other cases, where the effect of the transfer function cannot be neglected, a digital simulator of $\bar{h}_a^{-1}(s)$ has to be found and the given record is processed in order to simulate $x(nT)$ - the high resolution spectrum. The high resolution spectrum is then used for the purpose of concentration estimation by application of Beer - Lambert law.

(ii) In controlling the spectrum resolution.

In our case, it is difficult to achieve an exact simulator of $\bar{h}_a^{-1}(s)$. Far more accurate results should be obtained with the simulation of $hh(s) = \bar{h}_a^{-1}(s) * h_1(s)$. Where $hh(s)$ represents the transfer function resulting from convolution of $h_1(s)$ - the spectrophotometer transfer function with the desired slit width and $\bar{h}_a^{-1}(s)$ is the inverse transfer function of the actual spectrophotometer that produces $y(nT)$. The net effect of applying $hh(s)$ on the signal $y(nT)$ is to simulate a new discrete record with the desired resolution. This approach allows a trade off between resolution and S/N - Ratio, an important feature, since in an attempt to achieve a too high resolution signal, a signal with too low S/N ratio may come out.

In the following subsections of this chapter the differential equation of a process (or a system) with a triangular impulse response is found. The technique used is generalised so that the differential equation of a process with an impulse response, in a polynomial form, can be derived. Also, the differential equation of a process with a gaussian impulse response is found. This is the first step in finding a digital process that enables us to simulate, in a discrete form, the spectrophotometer transfer function.

In the next subsection a simple approach, with limited accuracy, for finding a digital presentation of an analogue process is shown.

2.1.1 A simple solution of the discrete presentation of an analogue process

It can be noticed that if T - the sampling period - is small enough, we have :

$$y(nT) = \int h_a(nT-s)x(s)ds = \sum_j h_a(nT-jT)x(jT)T$$

Comparing the last result with (2.) we conclude, that for a given error we can find some sampling rate so that the series $Th_a[jT]$ presents the discrete simulator of $h_a(t)$ within the given error.

Summarizing :

In case the sampling rate is high enough so that the convolution integral may be presented by a sum, the actual values - $h_a(nT)$ - are used to present the digital simulator of $h_a(s)$. But in the general case $h[n]$ has, still, to be found.

The problem of simulating an analogue process by a discrete process is treated in [14]. The approach and its conclusions are presented ,briefly, in the following section.

2.1.2 Some features of the discrete presentation of an analogue process

The approach to the problem of discrete presentation of an analogue process, below, is given in [14]. Its conclusion are discussed and compared, later on, with the results derived here.

Given an input $\exp(j\omega t)$ and an analogue system with an impulse response $h_a(t)$ the output is (see 1.)

$$y(t) = \int h_a(t-s) \exp(j\omega s) ds = \int h_a(s) \exp(j\omega(t-s)) ds$$

$$y(t) = \exp(j\omega t) \int h_a(s) \exp(-j\omega s) ds = \exp(j\omega t) H_a(j\omega)$$

Therefore, the response of the analogue system to an input $\exp(j\omega t)$ is :

$$y(t) = \exp(j\omega t) H_a(j\omega) \quad (3.)$$

Applying the same input, $\exp(j\omega t)$, to a discrete linear system with an impulse response given by $h[n]$, we have (see 2.) :

$$y(t) = \sum_k h(t-kT) \exp(j\omega kT) = \sum_k h(kT) \exp(j\omega(t-kT))$$

$$y(t) = \exp(j\omega t) \sum_k h(kT) \exp(-j\omega kT) \quad (4.)$$

$H(z)$, the z -Transform of $h[n]$ is defined as

$$H(z) = \sum_n h(nT) z^{-n} \quad (5.)$$

Taking $z = \exp(-j\omega T)$ and substituting into (4.), we have

$$y(t) = \exp(j\omega t) \sum_n h(nT) \bar{z}^n = \exp(j\omega t) H(z)$$

$$y(t) = \exp(j\omega t) H(\exp(j\omega T)) \quad (6.)$$

Comparing (6.) and (3.) - the response of a discrete and an analogue system to a harmonic $\exp(j\omega t)$ - it can be noticed that the input signal $\exp(j\omega t)$ is affected in different ways. Bearing in mind that every signal is described uniquely by its Fourier Transform, we require that both systems, the analogue and the digital simulator will have the same response to a harmonic $\exp(j\omega T)$

$$H(\exp(j\omega T)) = H_a(j\omega) \quad (7.)$$

$H(\exp(j\omega T))$ is the digital simulator that simulates the analogue system $H_a(j\omega)$ for a given set of harmonics that satisfy the above equation.

Clearly, $H(\exp(j\omega T))$ the response of the discrete system to $\exp(j\omega T)$ is a periodic function with a period $\omega = \frac{2\pi}{T}$, where $H_a(j\omega)$ the response of the analogue system is not necessarily a periodic function. We maintain that if (7.) holds for a range of $|\omega| < \frac{\pi}{T} = \omega_b$, it holds for the class of signals which is band limited by ω_b , where $1/T$ is the sampling rate. That is, given an analogue signal with a Fourier Transform $F(j\omega) \neq 0$ in the range $\omega < \omega_b$ only, the output $y(j\omega)$ is given by :

$$Y(j\omega) = F(j\omega)H(\exp(j\omega T)) = F(j\omega)H_a(j\omega)$$

for every ω .

We conclude that for such signals, $H(\exp(j\omega T))$ is the digital simulator of H_a .

(7.) may be written

$$H_a(j\omega) = H(\exp(j\omega T)) = \sum_n h(nT) \exp(-j\omega nT)$$

for $w < \frac{\pi}{T}$

$h[n]$ are, therefore, the Fourier coefficients in the expansion of $H(\exp(jwT))$:

$$h[n] = \frac{2\pi}{T} \int_{-w_b}^{w_b} H(\exp(jwT)) \exp(jwnT) dw = \frac{2\pi}{T} \int_{-w_b}^{w_b} H(w) \exp(jwnT) dw \quad (8.)$$

The above equation is a formula for finding a digital simulator for an analogue system which has a frequency response - $H_a(jw)$ and for a class of band limited signals

Our approach, presented in the following sections, is based on operations in the time domain only. By introducing time domain operators (like D the differentiator and δ the central difference), it is possible to find a closed form, in the time domain, for a digital simulator to the analogue system presented by a triangular function. The result is generalized to some class of analogue systems.

The representation, we found, is, actually, the differential equation presented by a mixture of analogue (like D the derivative) and discrete (like δ the central difference) operators. Thus, given an impulse response (with a triangular impulse response in the case of the spectrophotometer) we, first, find the differential equation that the given impulse response represents, and secondly, a discrete form to the analogue equation is found by implementing the differential equation.

2.2 The digital - analogue differential equation of the spectrophotometer slit transfer function

In this section, the differential equation of the spectrophotometer transfer function (free of aberration and diffraction) with equal entry and exit slits - is developed. It was shown (see Sec.1.1.1) that such a transfer function has a triangular impulse response. For convenience, we describe the impulse response in a co-ordinate system as in Fig. 1. We assume first a general impulse response, symmetric about M , bounded and non zero only in the range $0 < M < 2M$.

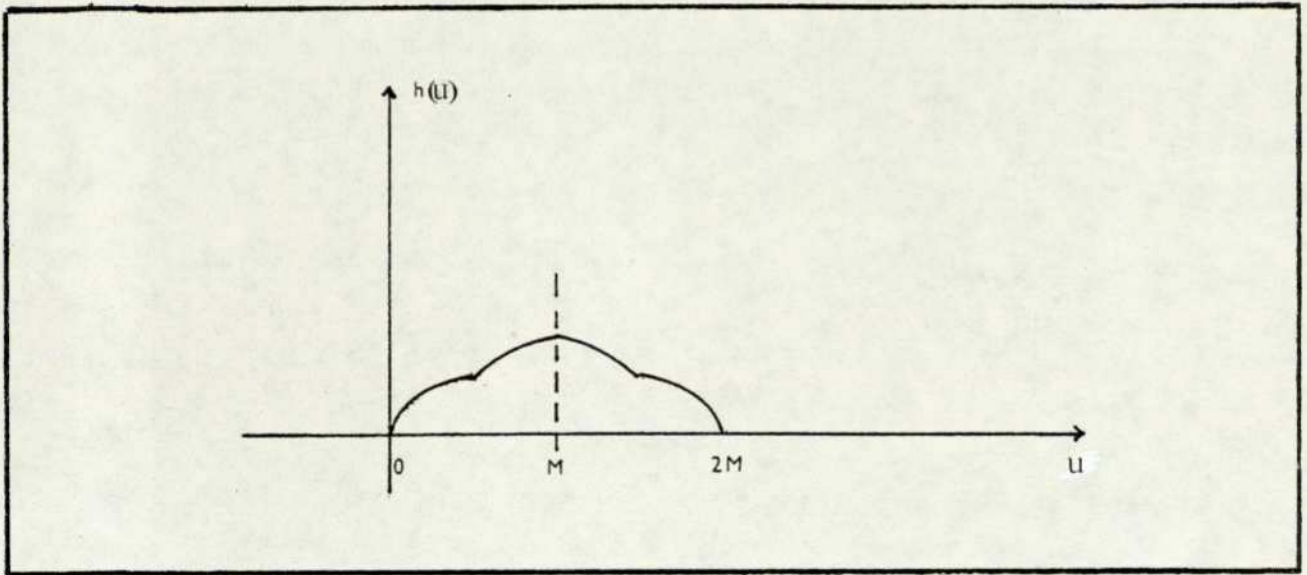


Fig. 1 Some general symmetric finite impulse response

$$h(M-u) = h(M+u) \quad (9.)$$

The output, from such system is given by :

$$y(nT) = \int_0^{2M} x(nT-u) h(u) du = \int_0^M x(nT-u) h(u) du + \int_M^{2M} x(nT-u) h(u) du$$

Which is also :

$$y(nT) = \int_0^M x(nT-u) h(u) du + \int_0^M x(nT-M-u) h(M+u) du \quad (9a.)$$

Substituting in the first term above :

$$M-s = u, \quad -ds = du$$

$$u = 0, \quad s = M$$

$$u = M, \quad s = 0$$

We have

$$\int_0^M x(nT-u) h(u) du = -\int_M^0 x(nT-(M-s)) h(s) ds = \int_0^M x(nT-M+s) h(s) ds$$

Changing integration variable from s to u and substituting in (9.a) we have :

$$y(nT) = \int_0^M x(nT-M+u) h(M-u) du + \int_0^M x(nT-M-u) h(M+u) du$$

Since $h(u)$ is symmetric function about M (see 9.), we have :

$$y(nT) = \int_0^M h(M+u) [x(nT-M+u) + x(nT-M-u)] du \quad (10.)$$

for asymmetric function, where :

$$h(M-u) = -h(M+u) \quad (11.)$$

we have

$$y(nT) = \int_0^M h(m+u) [x(nT-M+u) - x(nT-M-u)] du \quad (12.)$$

Using Taylor expansion, $x(u)$ is expanded around $(nT-M)$:

$$x(nT-M+u) = \sum_k D^k x(nT-M) \frac{u^k}{k!}$$

$$x(nT-M-u) = \sum_k D^k x(nT-M) \frac{(-u)^k}{k!}$$

Substituting in (10.), the odd terms cancel each other leaving only even terms :

$$y(nT) = 2 \int_0^M h(M+u) \sum_k D^{2k} x(nT-M) \frac{u^{2k}}{(2k!)} du$$

We assume that changing the order of operations is allowed :

$$y(nT) = 2 \sum_k \frac{D^{2k}}{(2k)!} x(nT-M) \int_0^M h(M+u) u^{2k} du \quad (13.)$$

Replacing the integration variable u by :

$$M + u = s, \quad du = ds$$

$$u = 0, \quad s = M$$

$$u = M, \quad s = 2M$$

we have :

$$\int_0^M h(m+u) u^{2K} du = \int_M^{2M} h(s) (s-M)^{2K} ds$$

and using again u as variable instead of s , we have :

$$y(nT) = 2 \sum_k D^{2k} x(nT-M) \frac{1}{2k!} \int_M^{2M} h(u) (u-M)^{2k} du \quad (14.)$$

$h(u)$ is given by :

$$h(u) = \begin{cases} u & 0 \leq u \leq M \\ 2M-u & M < u \leq 2M \\ 0 & \text{elsewhere} \end{cases} \quad (15.)$$

Substituting (15.) in 14.) :

$$\begin{aligned} \int_M^{2M} h(u) (u-M)^{2k} du &= \int_M^{2M} (2M-u) (u-M)^{2k} du = \\ &= \int_M^{2M} (M(u-M)^{2k} - (u-M)^{2k+1}) du \\ &= \left. \frac{M(u-M)^{2k+1}}{2k+1} - \frac{(u-M)^{2k+2}}{2k+2} \right|_M^{2M} \\ &= \frac{M^{2k+2}}{2k+1} - \frac{M^{2k+2}}{2k+2} = \frac{M^{2k+2}}{(2k+1)(2k+2)} \end{aligned}$$

Hence

$$\int_M^{2M} h(u) (u-M)^{2k} du = \frac{M^{2k+2}}{(2k+1)(2k+2)} \quad (16.)$$

Substituting 16. in 14. we have :

$$\begin{aligned}
 y(nT) &= 2 \sum_k D^{2k} x(nT-M) \frac{M^{2k+2}}{(2k+1)(2k+2)} \\
 y(nT) &= 2M^2 \sum_k \frac{(MD)^{2k}}{(2k+2)!} x(nT-M) \quad (17.)
 \end{aligned}$$

The first term ($k=0$) is $M^2 x(nT-M)$ where M^2 is the area under $h(u)$ given in 15.

Expanding $\cosh(z)$ as a power series :

$$\cosh(z) = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots + \frac{z^{2k}}{(2k)!} + \dots$$

and we easily notice :

$$\frac{\cosh(z)}{z^2} - \frac{1}{z^2} = \frac{1}{2!} + \frac{z^2}{4!} + \dots + \frac{z^{2k}}{(2k+2)!} + \dots \quad (18.)$$

$$\frac{\cosh(z)-1}{z^2} = \sum_k \frac{z^{2k}}{(2k+2)!} \quad (18a.)$$

Comparing 18a. and 17. $y(nT)$ can be written :

$$y(nT) = 2M^2 \frac{\cosh(MD)-1}{(MD)^2} \{x(nT-M)\} \quad (19.)$$

Where $\cosh(MD)$ is interpreted, as merely, the series given in 18. in manipulating expressions like $\cosh(MD)$, where an expression is a function of the operator D (or any other linear operator), we shall understand this expression, as merely, a power series in the given operator. Consequently, equality between two expressions means equality between their power series. Operations, like addition etc. between two expression are performed between their two power series.

Thus the power series which defines $\cosh(MD)$ can be expressed as a linear combination of two other power series :

$$\cosh(MD) = \frac{1}{2}\exp(MD) + \frac{1}{2}\exp(-MD)$$

When operating on some function $g(t)$, we have :

$$\cosh(MD)\{g(t)\} = \left(\frac{1}{2}\exp(MD) + \frac{1}{2}\exp(-MD)\right)\{g(t)\}$$

The fact that the above expressions are power series of linear operators enables us to 'open bracket' and study the effect of each expression on $g(t)$ separately :

$$\cosh(MD)\{g(t)\} = \frac{1}{2}\exp(MD)\{g(t)\} + \frac{1}{2}\exp(-MD)\{g(t)\}$$

But $\exp(MD)$ is the Taylor series in operator form :

$$\exp(MD)\{g(t)\} = \sum_{k=0}^{\infty} (MD)^k \frac{M^k}{k!} \{g(t)\} = g(t+M) \quad (20.)$$

Therefore, the effect of the operator $\cosh(MD)$ on $g(t)$ is :

$$\cosh(MD)\{g(t)\} = \frac{1}{2}(g(t+M) + g(t-M)) \quad (20a.)$$

The average operator u_{2M} is defined :

$$u_{2M}\{g(t)\} = \frac{1}{2}(g(t+M) + g(t-M)) \quad (21.)$$

$$\cosh(MD) = u_{2M} \quad (21a.)$$

We note that u is a discrete operator, that is, if $M = LT$ where T is a sampling period and L is some integer, u can be applied easily:

$$u_{2M}\{g(nT)\} = \frac{1}{2}(g((n+L)T) + g((n-L)T))$$

Thus some expressions have a simple meaning when applied to a discrete function.

The following identities hold :

$$2\sinh(MD)\{g(t)\} = (\exp(MD) - \exp(-MD))\{g(t)\}$$

From (20.):

$$2\sinh(MD)\{g(t)\} = g(t+M) - g(t-M) \equiv \delta_{2M}\{g(t)\}$$

defined above is the discrete central difference operator, which is given by :

$$\delta_{2M} = 2\sinh(MD) \quad (21b.)$$

Using the following identity, we have :

$$\begin{aligned} \cosh(MD) - 1 &= \frac{1}{2} \left(\exp\left(\frac{MD}{2}\right) - \exp\left(-\frac{MD}{2}\right) \right)^2 \\ &= 2\sinh^2\left(\frac{MD}{2}\right) \end{aligned}$$

From (21b.) :

$$\frac{1}{2}\delta_M = \sinh\left(\frac{MD}{2}\right)$$

Substituting the above :

$$\cosh(MD) - 1 = \frac{\delta_M^2}{2} \quad (22.)$$

Finally, substituting the above to 19. we have :

$$y(nT) = M^2 \frac{\delta_M^2}{(MD)^2} \{x((n-M)T)\}$$

Bearing in mind that the above expression is derived for a system with a triangular impulse response shifted by MT (see Fig.1 and Eq.15), we conclude that for a triangular function symmetric around y -axis (as given in 1.4) we have :

$$y(nT) = M^2 \frac{\delta_M^2}{(MD)^2} x(nT) \quad (23.)$$

Equation (23.) expresses the effect of a linear system with a triangular impulse response (width of $2M$ and area of M^2) on an input signal $x(nT)$.

The spectrophotometer impulse response is taken, usually, as a triangular function with unit area (see 1.4).

$$y(nT) = \frac{\delta_M^2}{(MD)^2} \{x(nT)\} \quad (24.)$$

Comments

(i) In order to simulate an output from a system with a triangular impulse response with any width (which is a multiple of the sampling period) it is enough to perform, first, a double integration then applying the discrete operator can be easily performed by using (22.)

:

$$\delta_M^2\{g(nT)\} = 2(\cosh(LD) - 1)\{g(nT)\}$$

$$\delta_M^2\{g(nT)\} = g((n+L)T) + g((n-L)T) - 2g(nT)$$

For simulating a different slit width only the last operation, with a different L , has to be repeated.

(ii) Assuming an input function :

$$x(nT) = \exp(jwnT)$$

The output is :

$$y(nT) = A \frac{\delta_M^2}{(MD)^2} \exp(jwnT) = A \delta_M^2 \frac{\exp(jwnT)}{(jMw)^2}$$

Where A is the area under the triangular impulse response. Using (21b.), we have :

$$\begin{aligned} \delta_M^2 \frac{\exp(\exp(jwnT))}{(jMw)^2} &= \exp(jwnT) \left(\frac{\exp(j\frac{wMT}{2}) - \exp(-j\frac{wMT}{2})}{(jwM)^2} \right)^2 \\ &= \exp(jwnT) (2j)^2 \frac{\sin^2(\frac{wMT}{2})}{(jwM)^2} \end{aligned}$$

$$y(nT) = A \exp(jwnT) 4 \frac{\sin^2(\frac{wMT}{2})}{(MW)^2} \quad (25.)$$

Taking a triangular with $2M$ width and a unit height, the area $A = M^2$, therefore :

$$y(nT) = \exp(jwnT) 4 \frac{\sin^2(\frac{wMT}{2})}{(Mw)^2} = \exp(jwnT) H_a(jw) \quad (26.)$$

$$H_a(jw) = \left(\frac{\sin(\frac{wMT}{2})}{\frac{Mw}{2}} \right)^2 \quad (26a.)$$

Where $H_a(jw)$ is the Fourier transform of a triangular function - the analogue system.

The above result holds for every w . In view of the discussion in 2.1 and and Eq. (7.) we conclude that the digital - analogue simulator given in (24.) has no restriction on the signal band width. In practice, the discrete simulation of a pure integration will impose a restriction on the class of signal that may be applied to the simulator.

2.2.1 The digital - analogue differential equation of some class of symmetric finite impulse response systems

The result achieved in 2.2 is extended, in the following section, to some class of transfer functions.

A symmetric impulse response of order L is defined as followed :

$$h_L(u) = \begin{cases} u^L & 0 \leq u \leq M \\ (2M - u)^L & M \leq u \leq 2M \\ 0 & \text{elsewhere} \end{cases} \quad (27.)$$

The input/output relation of a system with $h(u)$ as an impulse response is (see eq. 14.) :

$$y(nT) = 2 \sum_{k=0}^{\infty} D^{2k} x(nT-M) \frac{1}{2k!} \int_M^{2M} h(u) (u-M)^{2k} du \quad (28.)$$

Substituting $h_L(u)$ in the equation above, the following integral has to be solved :

$$\int_M^{2M} h_L(u) (u-M)^{2K} du = \int_M^{2M} (2M-u)^L (u-M)^{2K} du$$

Integrating by parts :

$$\begin{aligned} \int_M^{2M} (2M-u)^L (u-M)^{2k} du &= \left. \frac{(2M-u)^L (u-M)^{2k+1}}{2k+1} \right|_M^{2M} + \\ &+ \frac{L}{2k+1} \int_M^{2M} (2M-u)^{L-1} (u-M)^{2k+1} du \\ &= \frac{1}{2k+1} \int_M^{2M} (2M-u)^{L-1} (u-M)^{2k+1} du \end{aligned}$$

Repeating this operation L times :

$$\begin{aligned} (2M-u)^L (u-M)^{2k} &= \frac{L!}{(2k+1), \dots, (2k+L)} \int_M^{2M} (u-M)^{2k+L} du \\ &= \frac{L!}{(2k+1), \dots, (2k+L)} (u-M)^{2k+L+1} \bigg|_M^{2M} \end{aligned}$$

$$\int_M^{2M} (2u-M)^L (u-M)^{2k} du = \frac{L! M^{2k+L+1}}{(2k+1)!, (2k+L+1)}$$

Substituting the above in (28.) :

$$\begin{aligned} y(nT) &= 2L! M^{L+1} \sum_{k=0}^{\infty} \frac{(MD)^{2k} x(nT-M)}{(2k+L+1)!} \\ &= (L+1)! A \sum_{k=0}^{\infty} (MD)^{2k} \frac{x(nT-M)}{(2k+L+1)!} \\ y(nT) &= (L+1)! A \sum_{k=0}^{\infty} (MD)^{2k} \frac{x(nT-M)}{(2k+L+1)!} \end{aligned} \quad (29.)$$

where A is the area under $h_a(u)$ given in (27.)

$$A = \frac{2M^{L+1}}{L+1}$$

Again, as in 2.3, expanding $\cosh(z)$ as a power series, we have :

$$\cosh(z) = 1 + \frac{z^2}{2!} + \dots + \frac{z^{2k}}{2k!} + \dots$$

It is easily noticed that for $L = 2i + 1$, we have :

$$\frac{\cosh(z)}{z^{L+1}} - \sum_{j=0}^i \frac{z^{2(j-i-1)}}{2j!} = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k+L+1)!}$$

$$L = 2i + 1$$

Taking $z = MD$ and substituting the above in (29), we have :

$$y(nT) = (L+1)! A \frac{\cosh(MD)}{(MD)^{L+1}} - \sum_{j=0}^i \frac{(MD)^{2(J-i-1)}}{(2j)!} \{x(nT)\} \quad (30.)$$

The equation above describes the input/output relation of a system with an impulse response given by (27.)

In order to find a presentation to a symmetric L - order function, where L is an even number, $\sinh(z)$ is expanded as a power series :

$$\sinh(z) = z + \frac{z^3}{3!} + \dots + \frac{z^{2k+1}}{(2k+1)!} = \sum_{k=0}^{\infty} \frac{z^{2k+1}}{(2k+1)!}$$

Taking $L = 2i$, and following the same procedure above :

$$\frac{\sinh(z)}{z^{L+1}} = \sum_{k=0}^{i-1} \frac{z^{2(k-i)}}{2k!} = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k+L+1)!}$$

$$L = 2i, \quad i > 0$$

Taking $z = MD$, and substituting the above to (29.) :

$$y(nT) = (L+1)! A \frac{\sinh(MD)}{(MD)^{L+1}} - \sum_{k=0}^{i-1} \frac{(MD)^{(k-i)}}{2k!} \{x(nT)\} \quad (30a.)$$

$$i > 0, A = \frac{2M^{L+1}}{L+1}, \quad L = 2i$$

The differential equations, in operator form, of some of the symmetric impulse response systems are given in Table 1.

L	Function
0	$2M \frac{\sinh(MD)}{(MD)}$
1	$M^2 \frac{\cosh(MD)}{(MD)^2} - \frac{1}{(MD)^2}$
2	$2 \frac{M^3}{3} \frac{\sinh(MD)}{(MD)^3} - \frac{1}{(MD)^2}$
3	$2 \frac{M^4}{4} \frac{\cosh(MD)}{(MD)^4} - \frac{1}{(MD)^4} - \frac{1}{2(MD)^2}$

Table 1. Differential equations, in operator form, of some symmetric impulse response systems.

In much the same way presentation to an asymmetric impulse response can be achieved. This, in turn, will enable to present an impulse response with a general polynomial form by separating it to its symmetric and asymmetric part.

2.3 A discrete simulation of a gaussian impulse response system

Using a gaussian impulse response of the form :

$$h(u) = \frac{1}{\sigma\sqrt{2\pi}} \int x(nT-u) \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) du$$

and an input signal $x(u)$, the output $y(nT)$ is given by (14.) :

$$y(nT) = \frac{1}{\sigma\sqrt{2\pi}} \int x(nT-u) \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) du$$

Expanding $x(nT-u)$ as a Taylor series around nT :

$$y(nT) = \frac{1}{\sigma\sqrt{2\pi}} \int \sum_k \frac{D^k\{x(nT)\}}{k!} (-u)^k \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) du$$

The odd terms are canceled in the integration over infinite range.

Changing the order of operations :

$$y(nT) = \frac{1}{\sigma\sqrt{2\pi}} \sum_k D^{2k} \left\{ \frac{x(nT)}{2k!} \int u^{2k} \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) du \right\} \quad (31.)$$

The integral above is given by [15]

$$\frac{1}{\sigma\sqrt{2\pi}} \int \frac{u^{2k}}{2k!} \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2\right) du = \frac{\sigma^2}{2} \frac{1}{k!}$$

Substituting to (31.)

$$y(nT) = \sum_{k=0}^{\infty} D^{2k} x(nT) \left(\frac{\sigma^2}{2}\right)^k \frac{1}{k!}$$

In operator form :

$$y(nT) = \sum_k \left(\frac{1}{2}(\sigma D)^2\right)^k \frac{1}{k!} \{x(nT)\} = \exp\left(\frac{1}{2}(\sigma D)^2\right) \{x(nT)\}$$

$$y(nT) = \exp\left(\frac{1}{2}(\sigma D)^2\right) \{x(nT)\}$$

Where, again as in 2.2.3, $\exp\left(\frac{1}{2}(\sigma D)^2\right)$ is understood as merely, the power series in (σD) .

2.4

Summary

In this chapter the differential equation of some general process was derived. The derivation was, first, limited to a system with a finite symmetric impulse response, given in a polynomial form. It was shown that the same approach can be applied in finding the differential equation of a system with asymmetric finite impulse response. Consequently, the differential equation of a system with any finite impulse response, in a polynomial form, can be derived.

Ch. 2

The approach developed used only operations in the time domain (rather than operations in the frequency domain). It was shown that some expressions have a simple meaning when applied to a discrete signal e.g. the transfer function :

$$T(j\omega) = \sin(\omega T)$$

can be expressed in the time domain as an operator :

$$F(D) = \sinh(DT)$$

where D is the differentiator operator.

When the above operator is applied to a discrete signal it is, simply, half the central difference operator :

$$\sinh(DT) = \delta_2$$

Thus, in some cases the presentation of an analogue process in a discrete form can have a simple form.

In the next chapter a discrete presentation of the differential equation in the case of a triangular response and a gaussian impulse response is derived.

3. A Discrete Realisation of the Spectrophotometer Slit Transfer Function

3.1

Introduction

In Ch.2 it was shown that the output- $y(t)$ of a spectrophotometer with a triangular shaped impulse response is given by applying the following operator to the input $x(t)$:

$$y(t) = \frac{\delta_M^2}{(MD)^2} \{x(t)\}$$

$$y(t) = T(D) \{x(t)\}$$

$$T(D) = \frac{\delta_M^2}{(MD)^2}$$

Where t , the time axis may, with a change in the scaling factor, be regarded as the wave length axis $[1/\text{cm}]$.

$T(D)$ may be regarded as the system time domain transfer function (or the system operator). The above is justified by applying $T(D)$ to a harmonic $\exp(j\omega t)$. the result is (see Eq.12) :

$$\frac{\delta_M^2}{(MD)^2} \{\exp(j\omega t)\} = \frac{\sin^2\left(\frac{\omega M}{2}\right)}{\left(\frac{\omega M}{2}\right)^2} \exp(j\omega t)$$

the above result is the well known frequency response of a system with a triangular impulse response (see Fig.1) given by [15] :

$$H(j\omega) = \frac{\sin^2(\frac{\omega M}{2})}{(\frac{\omega M}{2})^2}$$

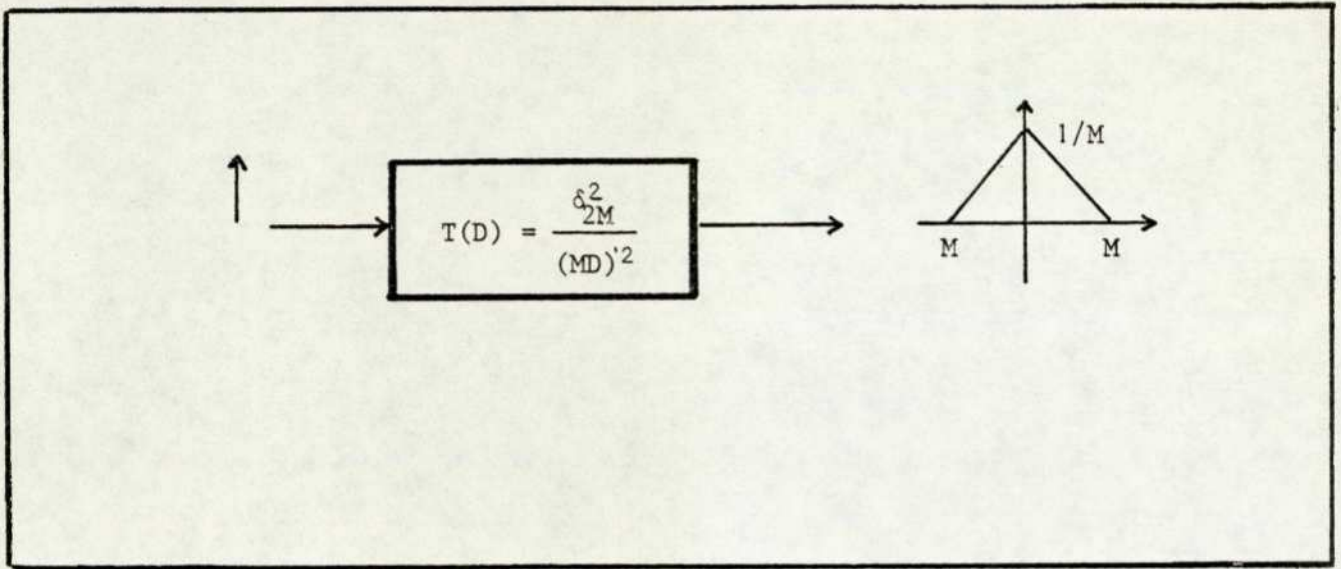


Fig. 1. The spectrophotometer response to an impulse
(sharp absorption line)

The frequency response of the spectrophotometer transfer function given by $[\sin(\omega M/2)/(\omega M/2)]^2$ reveals what is well known, namely, that any phenomenon with a duration less than the slit width is not resolved, and the fine structure of the spectrum diminishes with little possibility of recovery. In fact, the spectrophotometer transfer function acts as a low pass filter.

This chapter has two objectives. Firstly, as one of the aims of the thesis we wish to find a signal processing technique that enables the user to improve the resolution of a given spectrum signal. This implies that the inverse transfer function has to be realized such that it is possible to apply it to a discrete signal. As in the frequency domain, the $1/(\text{forward time domain transfer function})$ gives this operator. Thus if the time domain transfer function $T(D)$ is given by (see Fig.1) :

$$T(D) = \frac{\delta_M^2}{(MD)^2}$$

the inverse time domain transfer function is given by :

$$IT(D) = \frac{(MD)^2}{\delta_M^2}$$

Later it will be shown that a discrete realisation of the above transfer function has the form of the deconvolution formula.

As a part of the second goal, which is the estimation of gas concentration in a mixture, the forward transfer function has to be realized (see Sec.3.6 of this chapter). As is shown above, the time domain and the frequency domain transfer function are given, and may be manipulated in order to fulfil the above goals. A time domain implementation was chosen because of the prospect of developing an on-line fast signal processing technique. This is particularly important in solving the estimation problem, where a search loop for finding the concentration estimate is applied repeatedly.

In order to be able to apply either the time domain transfer function or the frequency domain transfer function the transfer function has to be expressed in a series of discrete operators so it may be applied to a discrete signal. In the time domain it may take the form of :

$$T(D) = \sum_n a_n u_n$$

Where $\{a_n\}$ has to be found, and $T(D)$ is the given transfer function.

Implementation as above can be applied to the given discrete signal directly with a minimum of computing effort, since the effect of the operator on the signal (see Ch.2, Eq.21) is given by :

$$u_n\{x(t)\} = \frac{1}{2} [x(t + \frac{n}{2}) + x(t - \frac{n}{2})]$$

Two time domain transfer functions which are most commonly used in the literature to describe the spectrophotometer transfer function were tested. The one has a triangular shape impulse response (see Fig.1), the other one is with a gaussian impulse response.

In the case of the transfer function with a triangular shape response the following implementation approaches have been tried :

(i) A simple derivation of a discrete deconvolution formula (Sec.3.1)

This formula has a limited accuracy, since only those samples with a slit width separation from one another are used. This formula is interesting, since it is similar to a correction formula that has been obtained by others [16]

(ii) Another discrete derivation of a deconvolution formula

A deconvolution formula (Sec. 3.3) which exploits all the available samples around a point is developed. The error graph as a function of the frequency is given.

(iii) Development of the partial deconvolution concept (Sec.3.4)

The inverse time domain transfer function (the deconvolution formula) is generalized such that it enables us to control the signal resolution (a partial deconvolution formula). Given a spectrophotometer and an output, this formula enables us to construct a signal with a desired resolution (desired slit width) from the given signal, within the accuracy limitation of this formula.

It is shown that this formula includes the deconvolution formula as a special case of a zero slit width. Since this formula allows either decreasing or increasing the slit width, it may be used not only in improving signal resolution, but also in solving the estimation problem, where the implementation of the forward transfer function is needed. Furthermore, it is shown

that in some cases of increasing the slit width (decreasing the resolution) this formula may have an exact form in a finite series, and therefore, will not introduce an error in applying it.

In the case of a gaussian impulse response, it is shown that the formula that controls the slit width (and consequently the resolution) and the deconvolution formula have the same form. Consequently, only one formula has to be implemented, for different slit widths.

The formula is tested for its accuracy as a function of the signal frequency.

Finally, an approach based on expanding the operator series (the implementation of the transfer function) to a fourier series is suggested.

3.2 Derivation of a simple deconvolution formula for a triangular impulse response transfer function

In this subsection a deconvolution formula (inverse transfer function) is realised. The formula incorporates only samples which are separated by M , where M is the slit width. By taking only the first two terms of the formula it is recognized as a correction formula that has been given by others. The difference is that the one given here may be applied directly to a discrete signal, whereas the other is given as a polynomial in the frequency domain.

Ch. 3

Let an output be given $y(T), y(2T), \dots, y(nT)$, the input $x(T), x(2T), \dots, x(nT)$ is given (see Ch.2, Eq. 24.) by:

$$x(jT) = \frac{(MD)^2}{\delta_M^2} \{y(jT)\} \quad (1.)$$

where M is the slit width of a triangular transfer function with a unit area. A simple discrete realisation for band limited signal is found as follows :

from (Ch.2, 21b.)

$$\delta_M = 2 \sinh\left(\frac{MD}{2}\right) \quad (2.)$$

The series expansion of $\sinh^{-1}(z)$ is [17]

$$\sinh^{-1}(z) = z - \frac{1}{2.3} z^3 + \frac{3}{2.4.5} z^5 - , + , \quad (3.)$$

$$z < 1$$

Substituting (3.) in (2.) :

$$\begin{aligned} \frac{MD}{2} &= \left(\frac{\delta_M}{2}\right) - \frac{1}{2.3} \left(\frac{\delta_M}{2}\right)^3 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2}\right)^5 - , + , \\ &= \left(\frac{\delta_M}{2}\right) \left(1 - \frac{1}{2.3} \left(\frac{\delta_M}{2}\right)^2 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2}\right)^4 - , + , \right) \end{aligned}$$

$$(\text{MD})^2 = (\delta_M)^2 \left(1 - \frac{1}{2.3} \left(\frac{\delta_M}{2} \right)^2 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2} \right)^4 - , + , \right)^2 \quad (4.)$$

Substituting (4.) to (1.), we have the following realisation

$$\begin{aligned} x(nT) &= \frac{(\text{MD})^2}{\delta_M^2} \{ y(nT) \} = \frac{\delta_M^2}{\delta_M^2} \left(1 - \frac{1}{2.3} \left(\frac{\delta_M}{2} \right)^2 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2} \right)^4 - , + , \right)^2 \{ y(nT) \} \\ x(nT) &= \left(1 - \frac{1}{2.3} \left(\frac{\delta_M}{2} \right)^2 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2} \right)^4 - , + , \right)^2 \{ y(nT) \} \end{aligned} \quad (5.)$$

Eq. (5.) is a discrete realisation of the inverse transfer function of a system with a triangular impulse response, base of $2M$ and a unit area. In the above formula only $y(M), y(2M), y(jM)$, are used. The formula only applies to signal with band width of

$$w < \left| \frac{\pi}{M} \right| \quad (6.)$$

Mathematically, taking $y(nT) = \exp(jwnT)$ and bearing in mind that the series converges only in the open unit circle (see 3.), we have :

$$\begin{aligned} x(nT) &= \left(1 - \frac{1}{2.3} \left(\frac{\delta_M}{2} \right)^2 + \frac{3}{2.4.5} \left(\frac{\delta_M}{2} \right)^4 - , + , \right)^2 \{ \exp(jwnT) \} \\ x(nT) &= \left(1 + \frac{1}{2.3} \sin^2 \left(\frac{wMT}{2} \right) + \frac{3}{2.4.5} \sin^4 \left(\frac{wMT}{2} \right) + , + , \right)^2 \exp(jwnT) \quad (16a.) \end{aligned}$$

Consequently, the series converges only for :

$$\left| \sin \left(\frac{wMT}{2} \right) \right| < 1$$

that is :

$$\left| \frac{wMT}{2} \right| < \frac{\pi}{2}$$

Taking T the sampling period as a unit, Eq. (6.) follows.

Taking the first two terms in (5.)

$$\begin{aligned} x(nT) &\cong \left(1 - \frac{1}{2.3} \left(\frac{\delta_M^2}{2} \right) \right)^2 \{y(nT)\} \\ &= \left(1 - \frac{2}{2.3} \left(\frac{\delta_M}{2} \right)^2 + \frac{1}{2.3} \left(\frac{\delta_M}{2} \right)^4 \right) \{y(nT)\} \\ x(nT) &\cong \left(1 - \frac{\delta_M^2}{12} + \frac{\delta_M^4}{36.16} \right) \{y(nT)\} \end{aligned} \quad (6b.)$$

Since in the formulas above (Eq.5,6b) only powers of δ_M^2 appear, where the effect of δ_M on the signal is given by :

$$\delta_M \{g(t) = g(t+MT) - g(t-MT)\}$$

we conclude that only samples with MT separation from each other are involved.

A similar correction formula has been given by others [16]

It is worth noting that if the analogue signal is first filtered with a band pass filter of $f < 1/2M$, degradation of the S/N level should be less than the degradation caused by a second order differentiator. Since, after all, in the realisation of (1.) only second order derivative is involved.

3.3 Another derivation of the deconvolution formula for a triangular impulse response system

In (3.2) a discrete deconvolution formula- based on samples of distance M from each other- was given. A realisation that takes into account all the samples around a points is as follows :

Substituting (2.) in (1.), the inverse spectrophotometer transfer function is expressed in term of the operator D :

$$\begin{aligned} \frac{(MD)^2}{\delta_M^2} &= \frac{1}{4} \frac{(MD)^2}{\sinh^2\left(\frac{MD}{2}\right)} = \frac{1}{4}(MD)^2 \frac{2}{\cosh(MD)-1} \\ \frac{(MD)^2}{\delta_M^2} &= \frac{1}{2} \frac{(MD)^2}{\cosh(MD)-1} \end{aligned} \quad (7.)$$

Eq. (7.) has to be realised by a series of discrete operators. We choose a realisation for (7.) in the form

$$\frac{1}{2} \frac{(MD)^2}{\cosh(MD)-1} \cong a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND)$$

where the effect of $\cosh(nD)$ on a discrete signal is given by (see Ch. 2. Eq.21)

$$\cosh(nTD)\{x(t)\} = \frac{1}{2} (x(t + nT) + x(t - nT))$$

$$\frac{1}{2} (MD)^2 = (\cosh(MD) - 1) \sum_{n=0}^N a_n \cosh(nD) \quad (8.)$$

Expanding the above series as a power series in D we have (see App. 1)

$$\sum_{k=0}^N \sum_{l=0}^{j-1} \frac{D^{2j} M^{2(j-1)}}{2^{j-1} (j-1)! 2!^k} k^{2l} a_k = \frac{(MD)^2}{2}$$

$$j = 1, \dots, N+1$$

$$k = 0, \dots, N$$

$$k^{2l} = \begin{cases} 1 & k = 0, l=0 \\ 0 & k = 0, l \neq 0 \end{cases}$$

Table 1. , App. 3 gives the coefficients of the series in (9.).

3.3.1 Accuracy and limitation of the discrete deconvolution formula found in 3.3

In this subsection the discrete realisation found in 3.3 is tested for its accuracy and limitation by applying a harmonic $\exp(j\omega t)$ to both, the given transfer function and its realisation in a discrete form. The difference between the two values (normalized to the value given by the true transfer function) is plotted as function of the frequency (see Fig. 2.).

The discrete deconvolution formula found has a form of finite series (see Table 1 App.3.).

The series includes 11 elements and is given for different slit width expressed in terms of the sampling period which is assumed to be equal unity. Applying this operator series to a signal is done easily by using (21.), Ch.2.

$$x(nT) = \frac{(MD)^2}{\delta_M^2} \{y(nT)\} \cong a_0 + a_1 \cosh(TD) + \dots + a_{10} \cosh(10TD) \{y(nT)\}$$

$$x(nT) = a_0 y(nT) + \frac{1}{2} \sum_{i=1}^N a_i (y(nT-iT) + y(nT+iT)) \quad (10.)$$

Applying $\exp(j\omega nT)$ to (10.)- the discrete realisation of the deconvolution formula

$$x_1(nT) = \left(\sum_{n=0}^{10} a_n \cos(\omega nT) \right) \exp(j\omega nT)$$

The same signal $\exp(j\omega nT)$, when applied to the analogue deconvolution formula given in (1.) leads to the following equation.

$$x(nT) = \frac{(\Delta M)^2}{\delta_M^2} \{ \exp(j\omega nT) \} \quad (11.)$$

From Eq.(25.), Ch.2 we have

$$\frac{\delta_M^2}{(\Delta M)^2} \exp(j\omega nT) = A \frac{\sin^2\left(\frac{\omega MT}{2}\right)}{\left(\frac{\omega M}{2}\right)^2} \exp(j\omega nT)$$

where A is the area under the triangular function (see Fig.1, Ch.2).

With $A=1$, we have

$$x_2(nT) = \frac{\left(\frac{\omega M}{2}\right)^2}{\sin^2\left(\frac{\omega MT}{2}\right)} \exp(j\omega nT)$$

Defining the error in the discrete realisation as:

$$Er = \left| (x_1(nT) - x_2(nT)) / x_2(nT) \right|$$

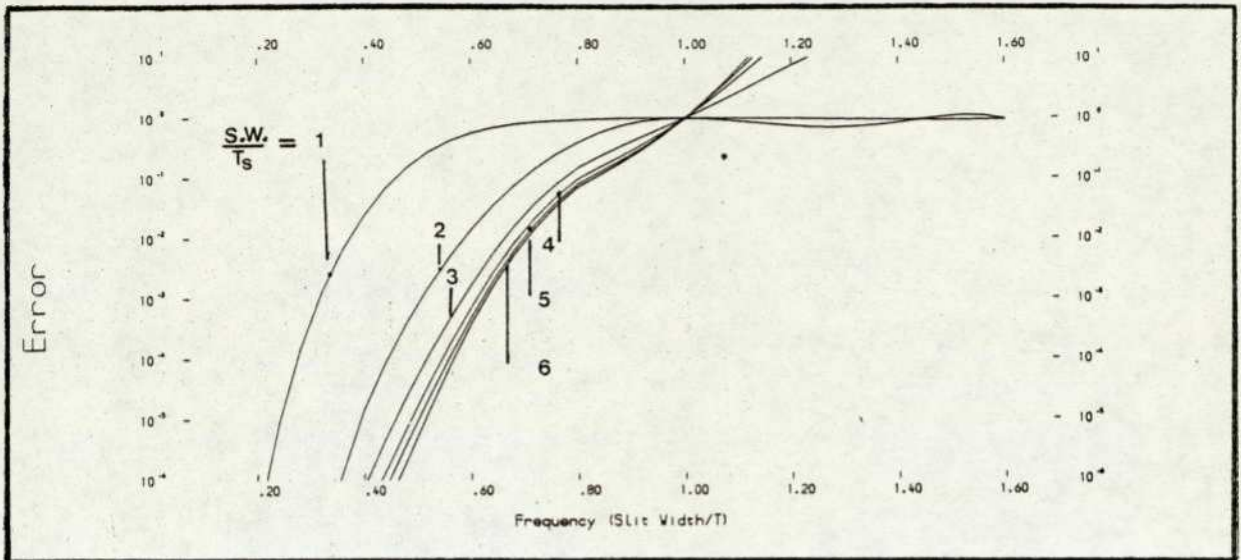


Fig. 2 Error graph of deconvolution formula for a triangular impulse response transfer function

Fig. 2. shows the error (E_r) in applying the discrete realisation (10.) to $\exp(j\omega nT)$. The error curves are plotted for a unity slit width. The graph parameter is (slit width)/ T_s , where T_s is the sampling period. The frequency is also normalized to the slit width. Thus increasing the sampling rate means higher ($S.W./T_s$) and therefore, smaller error. From Fig. 2 it is obvious that the discrete realisation may be used for frequencies in the range $f < 1/M$, where M is the slit width (in units of sampling period). Moreover, the error increases substantially for $f > 1/M$. Consequently, the analogue signal must be band limited in the range of $f < 1/M$ (or the discrete signal should have a lobe with a bandwidth $f < 1/M$) in order to be able to use the above discrete realisation.

Looking at (12.)- the analogue deconvolution equation, it can be shown that the transfer function has poles at

$$\omega = \frac{n}{MT} 2\pi$$

$$f = \frac{n}{MT}$$

From the above, it is clear that the realisation by expansion to a power series [18] is valid in the open circle $|jf| < 1/M$.

3.4 Developing the partial deconvolution concept

By operating, in cascade, on a discrete output $y(nT)$ with $(MD)^2 \delta_M^2$ and then with $(LD)^2 \delta_L^2$, the result emerges as the deconvolved output that again is convolved with a time domain transfer function with a different slit width- L . The 'net' operator is given by :

$$\begin{aligned} x(nT) &= \frac{(MD)^2}{\delta_M^2} \frac{\delta_L^2}{(LD)^2} \{y(nT)\} = \left(\frac{M}{L}\right)^2 \frac{\delta_L^2}{\delta_M^2} \{y(nT)\} \\ x(nT) &= \left(\frac{M}{L}\right)^2 \frac{\delta_L^2}{\delta_M^2} \{y(nT)\} \end{aligned} \quad (13.)$$

A realisation of the above enables us to simulate a discrete record processed with a slit function with the desired slit width- L from a given output of a spectrophotometer with a slit width M .

We shall show that the above equation includes the deconvolution formula as a special case of a time domain transfer function with a zero slit width.

From (22.), Ch.2 we have

$$\delta_M^2 = 2(\cosh(MD) - 1) \quad (13a.)$$

Substituting to (13.) we have

$$x(nT) = \left(\frac{M}{L}\right)^2 \frac{\cosh(LD) - 1}{\cosh(MD) - 1} \{y(nT)\} \quad (13b.)$$

Let L decrease to zero, we have

$$\left(\frac{M}{L}\right)^2 \frac{\cosh(LD) - 1}{\cosh(MD) - 1} = \left(\frac{M}{L}\right)^2 \frac{1 + \frac{(LD)^2}{2!} + \frac{(LD)^4}{4!}}{\cosh(MD) - 1}$$

$$\left(\frac{M}{L}\right)^2 \frac{\cosh(LD) - 1}{\cosh(MD) - 1} = M^2 \frac{\frac{D^2}{2} + L^3 \frac{D^4}{4!}}{\cosh(MD) - 1}$$

$$\lim_{L \rightarrow 0} \left(\frac{M}{L}\right)^2 \frac{\cosh(LD) - 1}{\cosh(MD) - 1} = \frac{(MD)^2}{2(\cosh(MD) - 1)}$$

Using (13a.), we have

$$\lim_{L \rightarrow 0} \left(\frac{M}{L}\right)^2 \frac{\delta_L^2}{\delta_M^2} = \frac{(MD)^2}{\delta_M^2} \quad (14.)$$

Thus we establish mathematically :

(i) Equation (13.) includes the deconvolution formula as a special case of a zero slit with L

(ii) It is impractical to implement accurately the deconvolution

formula (in the case of a triangular function), since that will mean a zero output power (a zero slit width). Consequently, the deconvolution equation can be used only as a correction formula when applied to the spectrophotometer output.

By substituting $jw = D$ (which is equivalent to apply the operator to a harmonic $\exp(jwt)$) we establish that the resultant transfer function is a periodic function, with a period of $2\pi/L$ for $L = nM$, where n is some integer

$$H(jw) = (M/L)^2 \frac{\cosh(LD)-1}{\cosh(MD)-1} \Big|_{D=jw} = (M/L)^2 \frac{\cosh(jw)-1}{\cosh(jMw)-1}$$

$$H(jW) = (M/L)^2 \frac{\cos(Lw)-1}{\cos(Mw)-1} \quad (15.)$$

Since $H(jw)$ - the analogue transfer function is a periodic function, it is possible to find a digital simulator for this function for any given signal [14].

Example 1. (increasing the slit width)

$$L = 2M$$

$$H_2(D) = (1/2)^2 \frac{\cosh(2MD)-1}{\cosh(MD)-1} = 1/4 \frac{2(\cosh^2(MD)-1)}{\cosh(MD)-1}$$

$$H_2(D) = 1/2 \frac{(\cosh(MD)-1)(\cosh(MD)+1)}{\cosh(MD)-1}$$

$$H_2(D) = \frac{1}{2}(\cosh(MD) + 1)$$

Consequently, given a record which is the output of a spectrophotometer with a slit M , a simulation of a record which is the output of a spectrophotometer with $2M$ slit width is given by (see Eq. 21., Ch.2)

$$y(nT) = \frac{1}{2}(\cosh(MTD) + 1)\{x(nT)\} = \frac{1}{2}\left(\frac{1}{2}(x(nT+MT) + x(nT-MT)) + x(nT)\right)$$

A record 0,0,0,1,0,0,0,

may be considered as the sampled record of the impulse response of a triangular function with $M=1$ (base of $2M=2$, and a unit area). Apply the above formula to this record, the result is :

(undefined),0,1/4,1/2,1/4,0,(undefined)

Which may be considered as the impulse response of a system with a triangular impulse response with $M = 2$ (base of $2M=4$, and a unit area).

The result above can be generalized. It has been shown above that in order to double the slit width from M to $2M$ ($L=2M$) the following transfer function has to be applied :

Ch. 3

$$H_2 = \frac{1}{2} \frac{\delta_{2M}^2}{\delta_M^2} = 1/2 \cosh(MD) + 1/2 \quad (15a.)$$

In order to increase the slit width from $M=1$ to $L=2^n$ we apply Eq.13 :

$$H_{2^n} = \frac{1}{2^{2n}} \frac{\delta_{2^n}^2}{\delta_1^2} = \frac{1}{4} \frac{\delta_2^2}{\delta_1^2} \frac{1}{4} \frac{\delta_{2.2}^2}{\delta_2^2} \frac{1}{4} \frac{\delta_{2.2.2}^2}{\delta_{2.2}^2}, \dots, \frac{1}{4} \frac{\delta_{2^n}^2}{\delta_{2^{n-1}}^2} \quad (15b.)$$

But each term above is also (see Eq.15a):

$$\frac{1}{4} \frac{\delta_{2^j}^2}{\delta_{2^{j-1}}^2} = \frac{1}{2} \cosh(2^{j-1}D) + \frac{1}{2}$$

Substituting Eq.15a in to Eq.15.b :

$$H_{2^n} = \frac{1}{2^n} (\cosh(D)+1)(\cosh(2D)+1), \dots, (\cosh(2^{n-1}D)+1) \quad (15c.)$$

The above transfer function describes the time domain operator that has to be applied to an output from a spectrophotometer with a slit width of unity in order to simulate an output from a spectrophotometer with a slit width of 2^n units. In App.5 the above equation is brought to a standard form.

Example 2. (reducing the slit width)

$$M = 2L$$

The result is $1/H_2(D)$ given in Example 1

$$H_2^{-1}(D) = 1/H_2(D) = \frac{2}{\cosh(MD)+1}$$

$$H_2^{-1}(D) = \frac{4}{\exp(MD)+\exp(-MD)+2} = \frac{4}{\left(\exp\left(\frac{MD}{2}\right) + \exp\left(-\frac{MD}{2}\right)\right)^2}$$

Substituting $D=jw$

$$H_2^{-1}(jw) = \frac{4}{\left[\exp\left(\frac{jwM}{2}\right) + \exp\left(-\frac{jwM}{2}\right)\right]^2} = \frac{4\exp(jwM)}{(\exp(jw)+1)^2}$$

$$H_2^{-1}(jw) = 4\exp(jwM) \left(\sum_n \exp(jwnM) (-1)^n \right)^2$$

Thus, the exact discrete realisation of a transfer function that reduces the slit width has the form (in the frequency domain) of a non convergent fourier series.

From the above examples it may be concluded that an exact realisation, in a form of a finite series is possible when a simulation of a function with a larger slit width than the given one is desired. When a simulation of a transfer function with a smaller slit width is desired the result is infinite series. This is established mathematically as follows:

Eq. (15.) the general equation for controlling the slit width has poles at :

$$\cos(Mw) - 1 = 0$$

$$\exp(j\omega MT) + \exp(-j\omega MT) - 2 = 0$$

$$\exp(-j\omega MT) (\exp(j\omega MT) - 1)^2 = 0$$

Thus all the poles are on the unit circle at

$$\omega_{Pn} = \frac{\pi}{M} + \frac{(2n+1)\pi}{M} \quad n = 0, 1, \dots, \frac{2n+1}{M} < M$$

Consequently, the denominator may be written

$$\frac{\exp(j\omega M)}{(\exp(j\omega M) - 1)^2} = \frac{\exp(j\omega M)}{\prod_n (\exp(j\omega M) - \exp(j\omega_{Pn}))^2} \quad (16.)$$

Eq. (15.) has zeros on the unit circle at

$$\omega_{Zn} = \frac{\pi}{L} + \frac{2n+1\pi}{L} \quad n = 0, 1, \dots, \frac{2n+1}{L} < L$$

Therefore (15.) may be written

$$H_2^{-1}(j\omega) = (M/L)^2 \exp(j\omega(M-L)) \frac{\prod_n (\exp(j\omega L) - \exp(j\omega_{Zn}))^2}{\prod_k (\exp(j\omega M) - \exp(j\omega_{Pk}))^2} \quad (17.)$$

$$n = 0, 1, \dots, \frac{2n+1}{L} < L$$

$$k = 0, 1, \dots, \frac{2k+1}{M} < M$$

We conclude that in the case $L > M$ where a spectrophotometer with a larger slit width than the given one is simulated, Eq. (15.) has more zeros than poles. Therefore it is possible that all the poles will be cancelled out as in Example 1.

In case $L < M$, that is, the simulated spectrophotometer has a smaller slit width than the given one, Eq. (15.) has more poles than zeros. Consequently, some of the poles, at most, are canceled out. Since all the poles are on the unit circle, the expansion to a power series in $\exp(j\omega nT)$ yields non convergent series as in Example 2.

In the subsequent sections we shall find a discrete realisation to (15.) and check its accuracy.

3.4.1 A discrete realisation of the partial deconvolution formula

In 3.3 a formula that enables us to simulate a spectrophotometer with a triangular impulse response with different slit width is suggested (see 15.). In this section a discrete realisation to this formula is found as follows :

Substituting (13a.) in (15.)

$$(M/L)^2 \frac{\delta_L^2}{\delta_M^2} = (M/L)^2 \frac{\cosh(LD)-1}{\cosh(MD)-1} \quad (18.)$$

Taking the same representation as in 3.2

$$(M/L)^2 \frac{\delta_L^2}{\delta_M^2} \approx a_0 + a_1 \cosh(D) + \dots + a_N \cosh(ND) \quad (18a.)$$

$$(M/L)^2(\cosh(LD)-1) = (\cosh(MD)-1) \left(\sum_{n=0}^N a_n \cosh(nD) \right)$$

The right hand side is exactly as in (8.). The left hand side has the following power series

$$\cosh(LD)-1 = \frac{(LD)^2}{2!} + \frac{(LD)^4}{4!} + \dots + \frac{(LD)^{2k}}{2k!}$$

Consequently, we have the following set of equations in $\{a_0, a_1, \dots, a_N\}$.

$$\sum_{k=0}^N \sum_{l=0}^{j-1} \frac{D^{2j} M^{2(j-1)}}{2^{j-1} 2l!} k^{2l} a_k = (LD)^{2j} (M/L)^2 \quad (19.)$$

$$\begin{array}{l}
 j = 1, \dots, N+1 \\
 k = 0, \dots, N \\
 T = 1
 \end{array}
 \quad
 k^{2l} = \begin{cases} 1 & k = 0, l = 0 \\ 0 & k = 0, l \neq 0 \end{cases}$$

Table 2. , App.III gives the coefficients $a_0, a_1, \dots, a_N$ for different slit width - M (the given slit width) and different slit width - L (the simulated slit width) where L and M are measured in sampling period units.

3.4.2 The accuracy and the limitation of the discrete realisation found in 3.4.1

In this section the discrete realisation of the partial deconvolution formula given in (18a.) is tested as function of the frequency. Applying $\exp(j\omega T)$ to both sides of (18a.), we have

$$H_1(j\omega) = (M/L)^2 \frac{\cosh(\omega LT) - 1}{\cosh(\omega MT) - 1}$$

$$H_2(j\omega) = \sum_{n=0}^{10} a_n \cos(\omega n T)$$

Where $\{a_0, a_1, \dots, a_{10}\}$ are given in Table 2, App.3 for given L and M

The error is defined

$$E_r(j\omega) = \left| \frac{H_1(j\omega) - H_2(j\omega)}{H_1(j\omega)} \right|$$

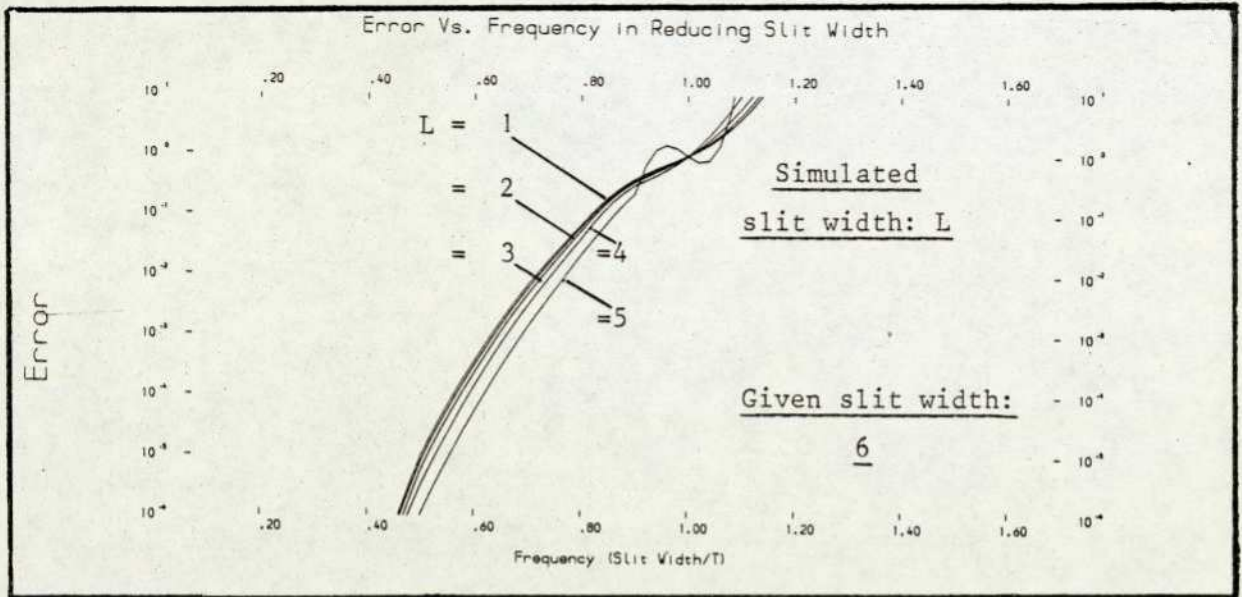


Fig. 3 Error graph for the partial deconvolution formula (triangular impulse response)

Fig. 3. shows the error as a function of the frequency (normalized to the slit width) for different values of L . The graph parameter is L/M (simulated slit width/given slit width), where M is 6 units, and $L = 1, 5$. From the error graphs it can be concluded that a frequency higher than $1/M$ is emphasised considerably beyond what is required and causes substantial error. Consequently, the discrete realisation is good for band limited signals with frequency range of

$$f < \frac{1}{M}$$

3.5 A partial deconvolution formula in the case of a gaussian impulse response system

Till now a spectrophotometer with a triangular impulse response transfer function was assumed. In this subsection, a spectrophotometer with a gaussian impulse response transfer function is assumed. It is shown that in this case, the problem of controlling the slit width has a natural solution. That is, there is no need to work out, as in the preveious case, a new transfer function.

The output response, $y(nT)$, of a system with a gaussian impulse response and an input $x(nT)$ is given by Eq. (31a.), Ch.2

$$y(nT) = \exp\left(\frac{1}{2}(MD)^2\right)\{x(nT)\}$$

where M is measured in T-units.

Comments

(i) The inverse transfer function (in the time domain) is given by :

$$x(nT) = \exp\left(-\frac{1}{2}(MD)^2\right)\{y(nT)\}$$

Substituting $D = j\omega$ (which is equivalent to apply the operator to a harmonic $\exp(j\omega t)$) we have

$$\exp\left(-\frac{1}{2}(MD)^2\right)\Big|_{D=j\omega} = \exp\left(\frac{1}{2}(M\omega)^2\right)$$

We note that the above transfer function does not have poles in the finite complex plane. Consequently, The discrete realisation found has a frequency range of existence which is as high as the sampling frequency (see Introduction, Ch.2) , compared to the range of (sampling frequency)/slit-width where the discrete realisation of the inverse triangular impulse response is valid (see 3.2.1)

(ii) The output $y(nT)$ of a system with a gaussian impulse response and an input $x(nT)$ is :

$$y(nT) = \exp\left(\frac{1}{2}(MD)^2\right)\{x(nT)\}$$

Now, operating on $y(nT)$, the output, with $\exp(-\frac{1}{2}(LD)^2)$ (the inverse transfer function), we have :

$$\exp\left(-\frac{1}{2}(LD)^2\right)\{y(nT)\} = \exp\left(-\frac{1}{2}(LD)^2\right)\exp\left(\frac{1}{2}(MD)^2\right)\{x(nT)\}$$

$$\exp\left(\frac{1}{2}(M^2 - L^2)D^2\right)\{x(nT)\}$$

The above transfer function can be regarded as transfer function with a gaussian impulse response and a variance of $(M^2 - L^2)$. As in the case of a system with a triangular impulse response (see Sec 3.3), taking $M = L$ yields a gaussian slit with a zero variance— an impractical solution. On the other hand taking, say, $L < M/2$ and applying the operator, twice, to the given signal, the net result is a system with an impulse response reduced by $2L$. We conclude that the variance of the gaussian impulse response may be controlled by applying the same transfer function repeatedly (see Fig. 5).

Again, as in the previous section, we would like to realise the above analogue time domain transfer function by a series of discrete operators :

$$\sum_{n=0}^N a_n \frac{\delta_{2n}}{2} = \sum_{n=0}^N a_n \cosh(nDT)$$

where T the sampling period is taken as a unit and $\{a_0, a_1, \dots, a_N\}$ have to be chosen such that :

$$\sum_{n=0}^N a_n \cosh(nD) \cong \exp\left(-\frac{1}{2}(LD)^2\right) \quad (19a.)$$

The approach we have used was developing both side of the equation above to a power series in (D) and to find $\{a_0, \dots, a_N\}$ such that the first $N + 1$ terms of both side will be the same . Although this approach did not guarantee a minimum error, the error decrease as the frequency of the signal input decreases. Some-what better results should be obtained by using a throw back technique [19]. In this technique some of the series terms with order higher than N are presented, with some error, in the set of equations that solve $\{a_0, a_1, \dots, a_N\}$ by using Chebicheve polynomials. Thus the effect of higher terms than $N + 1$ is taken in to account when solving the set of equations above.

An approach that guarantees a minimum rms error is as follows :
Substituting $jw = D$ we have

$$\sum_{n=0}^N a_n \cosh(nD) \Big|_{D=jw} = \sum_{n=0}^N a_n \cos(nw)$$

$$\exp\left(-\frac{1}{2}(LD)^2\right) \Big|_{D=jw} = \exp\left(\frac{1}{2}(Lw)^2\right)$$

where the sampling period T is taken as a unit.

The discrete realisation taken has the form of Fourier series in the frequency domain and has a period :

$$w_b = \frac{2\pi}{T}$$

Consequently, the discrete realisation is valid for frequency $|w| < w_b$.

The coefficients $(a_0, a_1, \dots, a_N)$ are given by [15]

$$a_n = \frac{2}{\pi} \int_0^{\pi} \exp\left(\frac{1}{2} (Lw)^2\right) \cos(nw) dw \quad (20.)$$

$$n = 1, \dots, N$$

$$a_0 = \frac{1}{\pi} \int_0^{\pi} \exp\left(\frac{1}{2} (Lw)^2\right) dw$$

Where T is taken as a unit. A similar results are given in the Introduction of Ch.2. This approach yields an infinite series when the inverse transfer function with a triangular response is to be implemented (see Ex.2, Sec.2) and therefore impractical solution.

3.5.1 A discrete realisation of the partial deconvolution formula for a gaussian impulse response transfer function

In this subsection the transfer function with a gaussian impulse response is implemented by a series of discrete operators. Again, applying both the discrete realisation of the transfer function and the transfer function to an input $\exp(j\omega t)$, and taking the normalized difference between the two output, we have the error as function of the frequency (see Fig.4)

In expanding both sides of (19.) to a power series in (D) we have (see App.II)

$$\sum_{n=0}^N a_n \cosh(nD) = \sum_{j=0}^N \sum_{n=0}^N a_n \frac{(nD)^{2j}}{(2j)!}$$

$$\exp\left(-\frac{1}{2}(LD)^2\right) = \sum_{j=0}^N \frac{(LD)^{2j}}{j!} \left(\frac{-1}{2}\right)^j$$

Where, again, the sampling period is taken as a unit and the variance is measured in number of samples.

We have to find $a_0, a_1, \dots, a_N$ such that

$$\sum_j a_n \frac{(nD)^{2j}}{2j!} = \sum_j \frac{(LD)^2}{j!} \left(\frac{1}{2}\right)^j$$

$$j = 0, \dots, N$$

As in the previous cases (see Sec. 3.3.2) the error for a Harmonic $\exp(j\omega t)$ is

$$E_r(j\omega) = \frac{\left| \sum_n a_n \cos(\omega n) - \exp\left(\frac{1}{2}(L\omega)^2\right) \right|}{\exp\left(\frac{1}{2}(L\omega)^2\right)} \quad (21.)$$

Table 3. gives the coefficient $\{a_0, a_1, \dots, a_N\}$ of the discrete realisation (19.) for $N = 10$ and for various values of variance (L). Figure 4. describes the error as a function of the frequency normalized by L (see Eq.19). The graph parameter is the sampling rate normalized by L . From this graph, we see a reduction in the error as the sampling rate increases, but only to some extent. Fig.5 demonstrates the effect of the transfer function on a given signal.

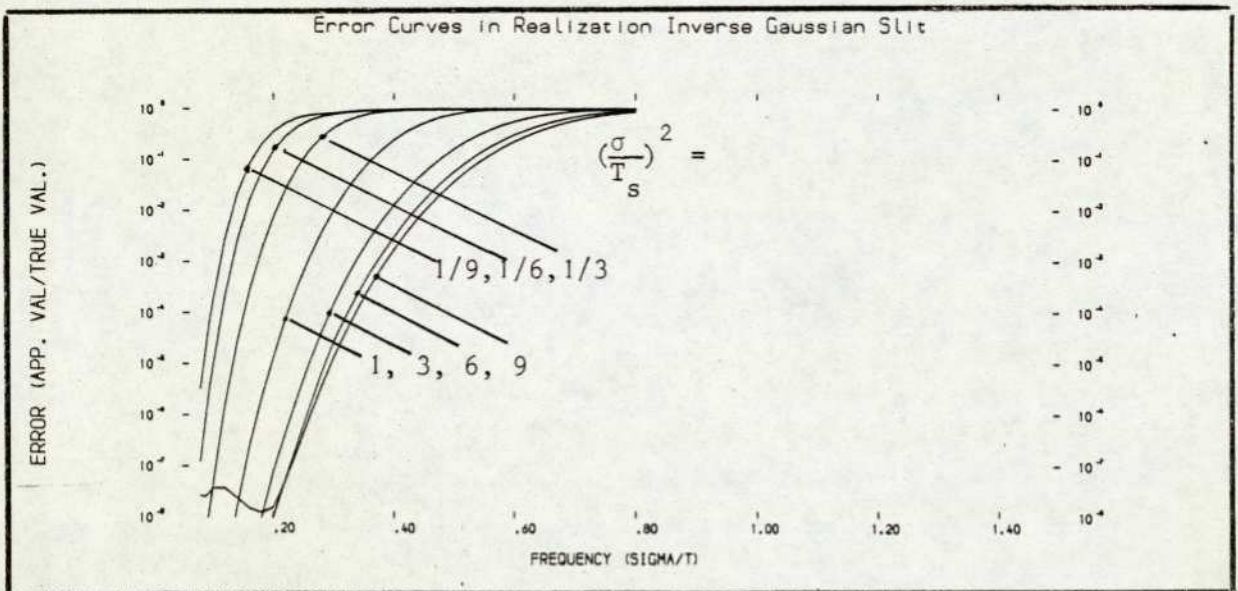


Fig. 4 Error graph of partial deconvolution formula
(gaussian impulse response)

3.5.2 A discrete realisation of a gaussian impulse response transfer function by a fourier series

The realisation of $\exp(-\frac{1}{2}(LD)^2)$ is given by

$$\exp(-\frac{1}{2}(LD)^2) = \sum_{n=0}^N a_n \cosh(nD)$$

where a_n is given by (20.)

$$a_n = \frac{2}{\pi} \int_0^{\pi} \exp(\frac{1}{2}(Lw)^2) \cos(nw) dw \quad (22.)$$

$$n = 1, \dots, N$$

$$a_0 = \frac{1}{\pi} \int_0^{\pi} \exp(\frac{1}{2}(Lw)^2) dw$$

App.3-

Table 4. shows the coefficients calculated. A look in the table shows that the coefficients decrease very slowly with n and therefore a longer series is needed in order to realise the transfer function

The rms error is given by [15]

$$ER^2 = \frac{1}{2} \int_0^{\pi} \exp((Lw)^2) dw - \sum_{n=0}^N a_n^2$$

In this chapter, some techniques of implementing the spectrophotometer transfer function, found in the last chapter, were tested. An input signal in a discrete form and a transfer function in analogue form (or analogue discrete form), impose constraints on the processing of the signal. The given transfer function either in the time or frequency domain has, first, to be expressed in terms of discrete operators. Since the intention of this work is to devise an on-line signal processing technique, we have chosen a realisation involving discrete time operators rather than frequency domain operators. The latter would require transferring, first, the signal to the frequency domain, applying the proposed operator and then transferring it again into the time domain, which may cause substantial computation. This would be the case when the user scanning a spectrum with, say, unsatisfactory noise level or resolution, wishes to improve the signal quality and have the processed signal be displayed on the vdu.

In realising, in the time domain, the inverse transfer function with a triangular impulse response, we have started by suggesting (see Sec.3.2) a formula that exploits only those samples with a slit width separation from each other. This is an interesting case, since a similar correction formula has been obtained by others using a different approach [16] :

$$y(t) = \left(1 - \frac{(MD)^2}{12}\right) \{x(t)\}$$

where the correction derived here is (see Eq.16b.) :

$$y(nT) = (1 - \frac{1}{12} \delta_M^2) \{x(nT)\}$$

It is worth noting that inclusion of higher order correction may not improve the formula significantly, since its range of existence is only defined for frequencies less than $1/\text{slit width}$. Although, adding higher correction terms should not degrade the S/N ratio, as, the formula is only a discrete realisation to D^2/δ_M^2 which should not degrade the signal more than a second order differentiator.

A formula that exploits all the available samples around a point is developed in 3.3. An error graph is given (Fig.2) that shows its accuracy as a function of the signal frequency. Judging from this graph, it can be seen that a signal with a band limited frequency range less than $1/\text{slit width}$ can be recovered.

In analysing the deconvolution formula (the inverse transfer function), a partial deconvolution (convolution) concept is developed by a generalization of the spectrophotometer transfer function (see Eq.13) :

$$T = \left(\frac{M}{L}\right)^2 \frac{\delta_L^2}{\delta_M^2}$$

where L is the simulated slit width,
and M is the given slit width.

A given output $x(nT)$ of a spectrophotometer (with a slit width M) can be used in order to simulate an output of a spectrophotometer with a slit width L . L may be larger or smaller than M .

In case $L > M$ the resolution of the simulated signal decreases (partial convolution formula).

In case $L < M$ the resolution of the simulated signal increases (partial deconvolution formula).

In case $L = M$ $T = 1$

In case $L = 0$ we have (see Eq.14) :

$$\lim_{L \rightarrow 0} T(D) = \frac{(MD)^2}{\delta_M^2}$$

The right hand side of the above equation is recognized as the deconvolution formula in the case of a transfer function with a triangular shaped impulse response. Consequently, applying the deconvolution formula to a signal, means simulation a signal output of a spectrophotometer with a zero slit width. Inevitably, that means a distorted output with poor S/N ratio. The partial deconvolution formula above enables us to control the signal resolution by choosing the desired slit width L .

e.g. in order to double the slit width (in case of a system with a triangular shaped impulse response), the formula is given by (see Ex.1) :

$$y(nT) = \frac{1}{2}x(nT) + \frac{1}{4}x(nT-MT) + \frac{1}{4}x(nT+MT)$$

The above equation is exact. In the above formula $x(nT)$ is the output of the spectrophotometer with a slit width M . and $y(nT)$ is the simulated signal output from a spectrophotometer with $2M$ slit width. This formula has been tested for some simulations of slit width less than the given one (see Sec.3.4.1). That is, the formula is used in this case as a partial deconvolution formula in order to increase the signal resolution. The results are given in a graph as a function of the signal frequency. Again, the accuracy of the formula is satisfactory in the frequency range less than $1/\text{slit width}$.

In the case of the transfer function with a gaussian impulse response, it is shown that the transfer function :

$$T(D) = \exp\left[-\frac{1}{2}(LD)^2\right]$$

will simulate a signal with a control resolution, from a given signal with a given resolution, by varying the parameter L (see Eq.3.5 and comment ii) . By taking L negative the slit width will increase and by taking L positive the slit width will decrease . The accuracy of the formula is tested for a range of slit widths (see Fig.4).

Finally, another approach to the realisation problem is tested (see Sec.3.5.2). In this approach the operator series (see Eq.19) is transferred to a fourier series, the coefficients of which $\{a_n\}$ have to be found so that the series will describe the given transfer function. It is shown that the coefficients $\{a_n\}$ found in this way decline to zero very slowly, and a large number has to be included in the equation (more than 100).

We conclude, the time domain transfer functions of a spectrophotometer, developed in Ch.2, were implemented in this chapter in a discrete form. The implementation has the form of a finite series (11-terms), and it is used as a deconvolution formula.

The partial deconvolution concept was introduced. By this concept we mean a linear process that aims to increase the signal resolution by a certain level only. Again the partial deconvolution formula was implemented by 11 terms series. It is shown (see Error Graphs Fig.2,3 and 4) that a discrete signal which has a band width (its main lobe) of less than $1/(\text{slit width})$ may be recovered by using the above formulas. Fig.5 demonstrates the application of the formula (partial deconvolution) by resolving the shape of two gaussian pulses.

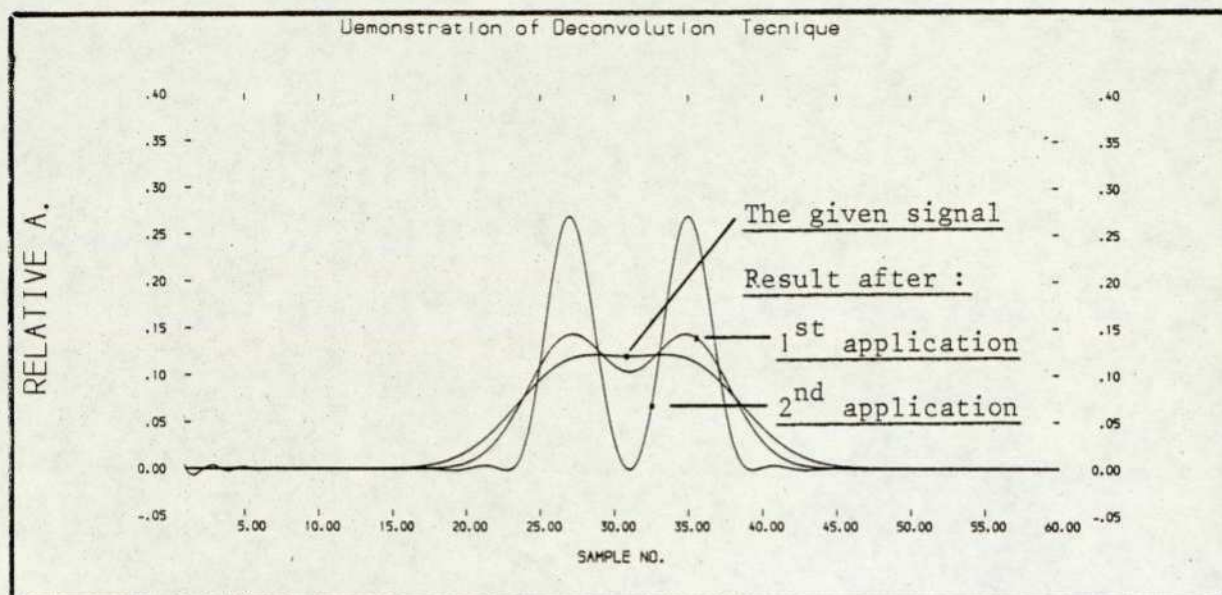


Fig. 5 Increasing the resolution of a given signal by applying a partial deconvolution formula (gaussian impulse response)

In simulating the forward transfer function (partial convolution formula), we expect better results, since the effect of the simulated transfer function would tend to reduce the signal frequency range rather than increasing it. Indeed, we show that for a triangular impulse response it is possible to find an accurate solution (see Sec.3.4 Ex.1) in some cases.

The last of the above results plays an important role in deriving the concentration estimation algorithm. In practice the signal resolution may be such that Beer - Lambert law may not be applied to this signal (see Sec.1.1). Consequently, a way to simulate a spectrum with a given resolution (given slit width) from the given absorption coefficients in a range of wavelengths is as follows (assuming two concentrations only) :

- (i) Given concentration (c_1, c_2) and absorption coefficients $(a_1(v_1), \dots, a_1(v_M))$ and $(a_2(v_1), \dots, a_2(v_M))$, Beer - Lambert law may be used in order to produce a high resolution spectrum.
- (ii) The implemented transfer function is then applied to the above signal to simulate an output from the spectrophotometer.

The above two steps will be included in the estimation loop (see Fig.7) where the simulated spectrum is compared with the given spectrum. According to some criteria a correction to the initial concentration vector (c_1, c_2) is made and the process is repeated until the comparison produces satisfactory results. This is the subject of the next two chapters.

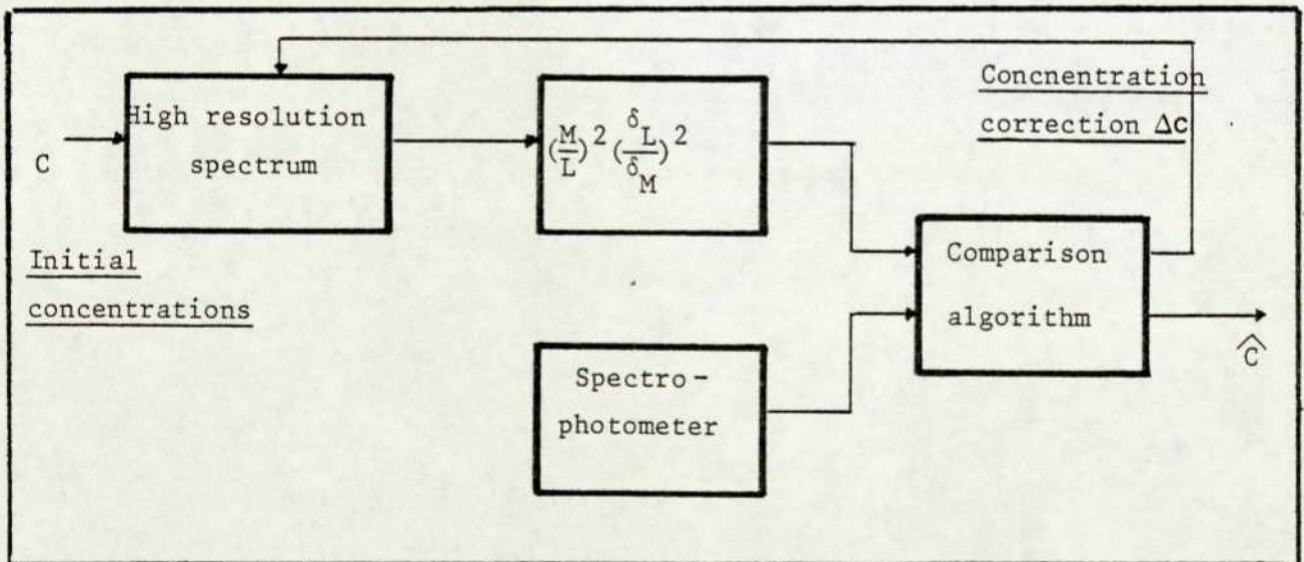


Fig. 6 The proposed concentration estimation algorithm

4. The Estimation Problem

4.1

Introduction

In this chapter the problem of estimating a single gas concentration in a mixture of several gases is investigated. The estimation process shown in Fig.1 is carried out along the following steps :

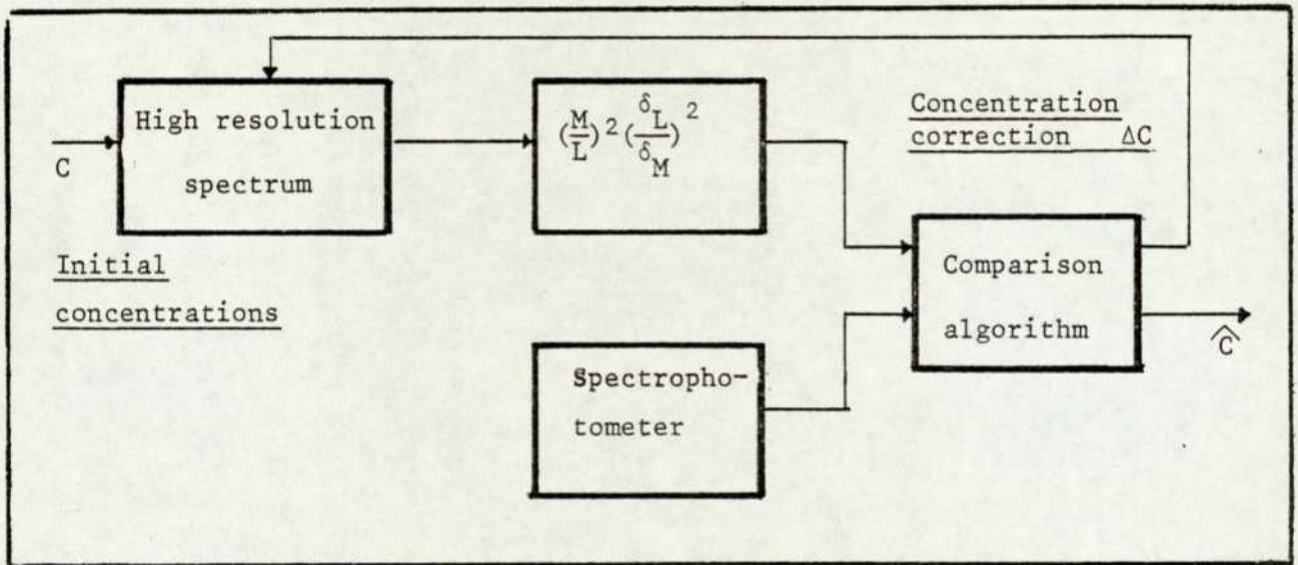


Fig. 1 A schematic diagram of the estimation process

(1) Simulation of a perfect resolution spectrum with some initial concentrations.

It is assumed that the spectrum measured is a spectrum of a mixture of gases with n compounds. The compound j in the mixture has a concentration c_j , and its absorption coefficient at wave-numbers (ν_1, ν_M) is given by :

$$\alpha_j = (\alpha_j(v_1), \dots, \alpha_j(v_M))$$

A spectrum of a gas mixture with n compounds and concentrations $(c_1, \dots, c_n)$ is defined by the absorption coefficients at the wave numbers $[v_1, \dots, v_M]$:

$$\alpha_1 = (\alpha_1(v_1), \dots, \alpha_1(v_M))$$

$$\alpha_n = (\alpha_n(v_1), \dots, \alpha_n(v_M))$$

The effective transmittance $x(v_i)$ at wave length v_i ($i=1, M$) of a mixture of n compounds is given by Beer - Lambert law :

$$x(v_i) = \exp\{-\sum_j c_j \alpha_j(v_i)\}$$

$$i = 1, M$$

We define the transmittance spectrum $x(v_i)$ as the perfect resolution spectrum (or perfect spectrum). We conclude, given the absorption coefficients vectors $\alpha_1, \dots, \alpha_n$, and a concentration vector $C = [c_1, \dots, c_n]$, the perfect spectrum is defined by the equation above.

(2) Simulation of the output of the spectrophotometer

It is assumed that the spectrophotometer has a known transfer function which may be defined as in Ch.2 (with a triangular or a gaussian impulse response) In such a case it has been shown that given the perfect resolution spectrum defined above, the spectrophotometer output, at a wave length v_i , may be expressed as a linear combination of the perfect resolution spectrum values around the wave number v_i :

$$f(v_i, C) = \sum_{n=-N}^N b_n x(v_i + nV)$$

where the $\{b_n\}$ are typical to each spectrophotometer and are derived in Ch.2 and Ch.3. V is the sampling period.

(3) Comparison of the simulated spectrum with the measured spectrum

(a) The result of the comparison is expressed by an error function (defined later).

(b) Upon the result of the comparison, a correction vector $[\Delta c_1, \Delta c_n]$ is formed and a new vector of estimates C' is found :

$$C' = (c'_1, \dots, c'_n) = (c_1, \dots, c_n) + (\Delta c_1, \dots, \Delta c_n)$$

(c) The process is repeated from step (1) until no significant change in either the error function or C' is obtained. The last C' calculated is the estimation process result - $\hat{C}$

In Sec.4.2 a mathematical model of the beam processing in the spectrophotometer is formulated. The three aspects of the process are :

- (i) The absorption process (Beer - Lambert Law) .
- (ii) The slit transfer function (the spectrophotometer transfer function derived in Ch.2 and 3)
- (iii) Additive gaussian noise.

This model enables us to carry out step (1) and step (2) of the estimation process described above (see Fig. 1), that is, to simulate a free noise spectrophotometer output with some initial concentrations C .

In Sec.4.3 the error function is defined. This error function is a function of the concentration vector C - the current estimate. The error function value reflects the distance of the vector C from the actual concentration vector C^0 , which we try to estimate.

Ideally, the error function should have two main features :

- (i) A single absolute minimum which occurs at C^0 - the actual concentrations of the compounds in the mixture.
- (ii) This minimum can be reached from any initial concentration C . That is, a criterion or a rule may be established such that it enables us to correct the current simulated concentration vector so that after repeating this process the actual concentration is reached.

Although (i) does not necessarily apply to the error function defined here, since the minimum (perhaps more than one) does not occur at C^0 , still, it is possible to show that the distance between $\hat{C}$, the estimation result, and C^0 , the actual concentration, is a function of the noise power. Consequently, for S/N ratio which is not too small, the estimate will be in the vicinity of C^0 .

Condition (ii) also may be not fulfilled. But, still, we shall show that in order that this will happen, two spectra in a given range of wavelength must have the same spectrum shape for two different concentration vectors. We, therefore, exclude this possibility as it is unlikely to happen in reality.

In Sec.4.4 a simple case of perfect resolution spectrum is first solved. That is, carrying out of step (2) in the estimation process (see above) becomes trivial (multiplication of the perfect resolution spectrum by unity). The formula for deriving the correction vector $(\Delta c_1, \dots, \Delta c_n)$ is found, and the rms error in the estimation is obtained.

Reducing the complexity of the problem even further to the linear case (which may be regarded as the estimation problem in the case of a very low concentration), enables us to identify different expressions and to generalize terms like S/N ratio signal power and correlation factor. The above generalization enables us to devise a tool for analysing different estimation problems and to predict the accuracy in other cases without carrying out the simulation process.

In Sec.4.5 we return to the general estimation problem. The assumption of a perfect resolution spectrophotometer is dropped, and a signal spectrum with an arbitrary resolution is assumed. It is shown that the S/N ratio, signal power and correlation factor may be further generalized so that they will play the same role as in the linear case in assisting us in analysing other cases of concentration estimation.

4.2 Modelling of the beam process in the spectrophotometer

The electrooptic process in the spectrophotometer may be divided, as in Fig. 2, to a series of subprocesses. First the absorption process of the ir energy, at every wave length takes place. This process is described by Beer - Lambert law.

$x(v)$, the output of this process is defined as the perfect resolution spectrum. The input density power, $x_{in}(v)$ is taken as unity. Secondly, the perfect spectrum is processed by the spectrophotometer transfer function. $f(v,C)$, the output at that stage is defined as the free-noise spectrophotometer output. Thirdly, the ir beam is converted, by the ir detector, to an electric signal. That is, in this stage the detector noise is added to the free noise spectrum. The output of this stage is the simulated spectrophotometer output.

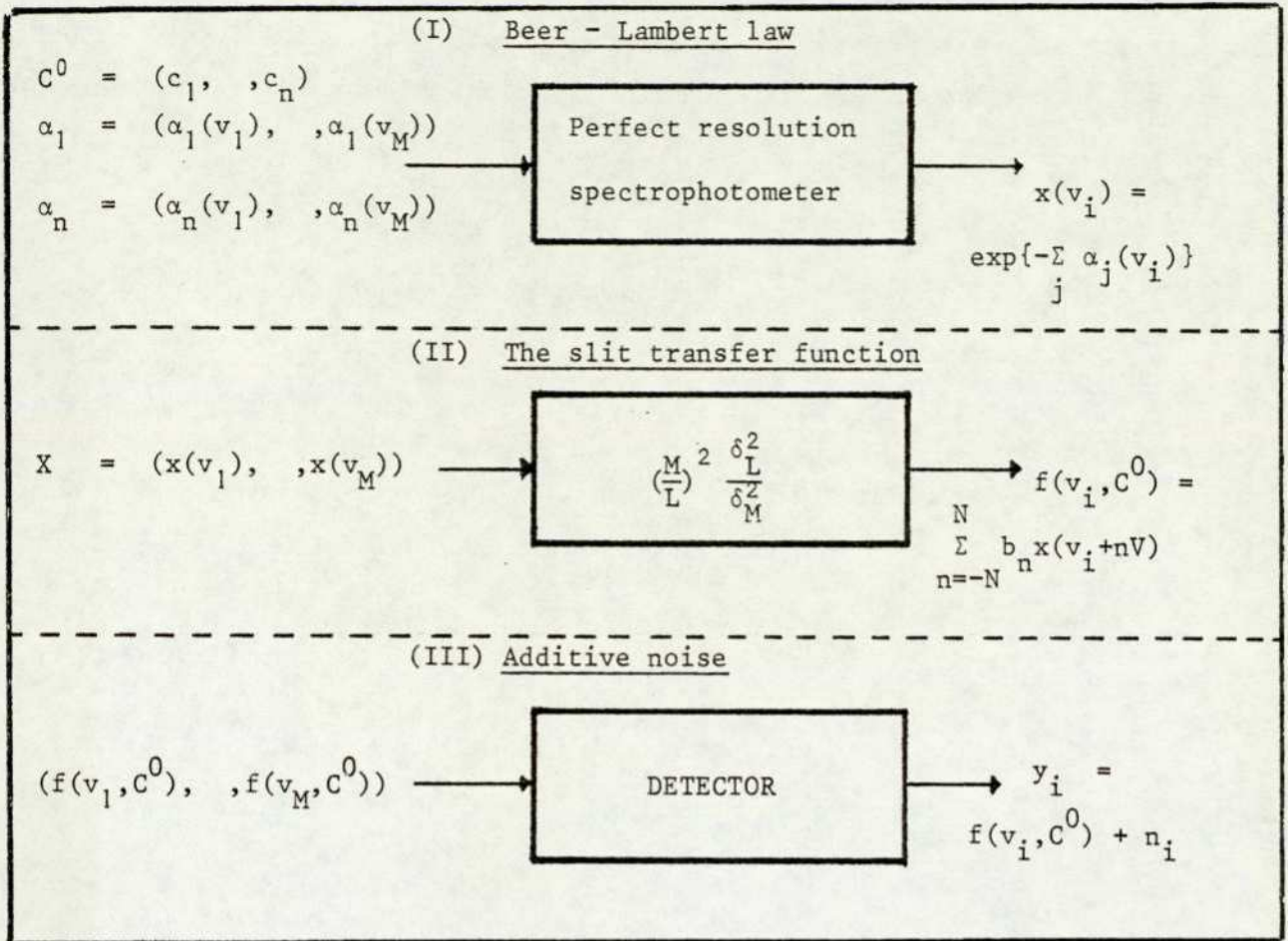


Fig. 2 Block diagram of the process in the spectrophotometer

$x(v)$ the perfect resolution spectrum at wave number v is given by (Beer - Lambert law) :

$$x(v) = \exp(-\sum_j c_j \alpha_j(v))$$

where the input intensity is assumed to be unity.

Given the absorption coefficients, $\alpha_j(v)$, of each compound j , $j = 1, n$ and at wave numbers v_i , $i = 1, \dots, M$, :

$$\alpha_j = (\alpha_j(v_1), \dots, \alpha_j(v_M))$$

$$j = 1, n$$

we have :

$$x(v_i) = \exp\left(-\sum_{j=1}^n c_j \alpha_j(v_i)\right) \quad (1.)$$

$$i = 1, M$$

$f(v_i, C)$ is given by applying the spectrophotometer transfer function (see Ch.3 Eq. 13)

$$f(v_i, C) = (M/L)^2 \frac{\delta_L^2}{\delta_M^2} \{x(v_i)\} \quad (2.)$$

Eq. 2 is expressed (Ch.3, Eq.18) by :

$$(M/L)^2 \frac{\delta_L^2}{\delta_M^2} = \sum_0^N a_n \cosh(nVD) \quad (3.)$$

where V is the sampling period (see Eq.21a, Ch.2) : Operating with 3. on $x(v)$, we have :

$$\begin{aligned} (M/L)^2 \frac{\delta_L^2}{\delta_M^2} \{x(v)\} &= \sum_0^N a_n \cosh(nVD) \{x(v)\} \\ &= x(v) + 1/2 \sum_1^N (x(v-nV) + x(v+nV)) a_n \end{aligned}$$

$$(M/L)^2 \frac{\delta_L^2}{\delta_M^2} \{x(v)\} = \sum_{n=-N}^N b_n x(v+nV) \quad (3a.)$$

$$b_0 = a_0$$

$$b_n = a_n/2$$

$$n = \pm 1, \pm 2, \dots, \pm N$$

Thus the effect of the instrument on the signal is expressed by a linear combination of the discrete input samples ($a_0, a_1, \dots, a_N$ are given in App. 3)

The effect of the detector on the output signal, beside multiplying it by a scaling factor (which may be taken as unity) is to add to it random noise (see 1.1.2) Assuming the output is sampled at the wave numbers ($v_1, v_2, \dots, v_M$)

$y(v_i)$ the spectrophotometer output at a wave number v_i is given by :

$$y(v_i) = \sum_{n=-N}^N b_n x(v_i+nV) + n_i \quad (4.)$$

$$i = 1, M$$

where :

$$x(v_i) = \exp\left(-\sum_j c_j j(v_i)\right)$$

The nature of the noise, added to the signal, is described in 1.1.2. It is assumed that the noise samples are uncorrelated and with constant power :

$$\overline{n_i n_j} = \begin{cases} 0 & i \neq j \\ \sigma^2_{n=N_0} & i = j \end{cases} \quad (5.)$$

Also the probability density function of the noise having the value n , $n_i < n < n_i + dn$ is given by :

$$p(n)dn = \frac{1}{\sqrt{2\pi N_0}} \exp\left(-\frac{1}{2}\left(\frac{n}{\sigma}\right)^2\right)dn \quad (5a.)$$

Finally, the set of equation (4.) is written in a form of :

$$y(v_i) = f(v_i, C^0) + n_i \quad (6.)$$

$$i = 1, \dots, M$$

where C^0 is the concentrations vector :

$$C^0 = (c_1^0, \dots, c_n^0)$$

and :

$$f(v_i, C^0) = \sum_{n=-N}^N b_n x(v_i + nV)$$

$$x(v_i) = \exp(-\sum_j c_j j(v_i))$$

c_i^0 , $i=1,n$ are the actual values of the concentrations at the time the measurment takes place. Thus the set of equations above describes a typical non-linear problem [20] where for a given set of measurements $y(v_1), y(v_2), \dots, y(v_M)$ we have to find $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ such that estimate $c_1^0, c_2^0, \dots, c_n^0$ with a minimum error in some sense.

The error function $E^2(C)$ we choose is :

$$E^2(C) = \sum_{i=1}^M ((f(v_i, C) - y(v_i))^2) \quad (7.)$$

It will be shown that this function has the features needed as defined in the introduction. That is, $\hat{C}$ that yields the minimum value of $E^2(C)$ is in the vicinity of C^0 . This $\hat{C}$ may be reached from from any initial concentration vector C . Furthermore, the same optimum estimate $\hat{C}$ is given by the maximum likelihood estimator. That is, the same error function $E^2(C)$ is derived by the MLE. In order to find the optimum estimates by a search method we expect the values of $E^2(C)$ to become smaller for values of $(c_1, \dots, c_n)$ that are closer to $(c_1^0, \dots, c_n^0)$. We also expect that the error function has a single minimum at the approximate of $(c_1^0, \dots, c_n^0)$. Such an error function is called a unimodal function [21]. It is, therefore, vital, first, to

investigate the error function.

4.3 Some features of the error function

In this subsection the error function defined in the last Section is investigated. It is shown that the error function can be described as the sum of two terms: the first the deterministic part is an ideal error function with a single minimum exactly at $C = C^0$; the second, the random part causes an error in the estimation results.

The gradient of the error function reflects the above aspect as well. The deterministic part has a magnitude (a norm) which is a function of the difference $(C - C^0)$, and becomes zero only for $C = C^0$. The gradient of the random part can have any direction (see Fig.3), and a magnitude proportional to the noise power. We assume that the process of correcting the current estimate may proceed as long as the gradient of the deterministic part is larger than the gradient of the random part of the error function (see Fig.3). This assumption leads us to the derivation of an expression which is a bound on the minimum value that the error function can reach.

Substituting (6.) into (7.), we have :

$$E^2(C) = \sum_i (f(v_i, C) - y(v_i))^2$$

$$E^2(C) = \sum_i (f(v_i, C) - f(v_i, C^0))^2 -$$

$$2 \sum_i (f(v_i, C) - f(v_i, C^0))n_i + \sum_i n_i^2 \quad (8.)$$

where :

$$C = (c_1, \dots, c_n)$$

$$f(v_i) = \sum_{n=-N}^N b_n x(v_i + nV)$$

$$x(v_i) = \exp\left(-\sum_{j=1}^n c_j a_j(v_i)\right)$$

$E^2(c_1, \dots, c_n)$ the error function has a random part and a deterministic part :

$$E^2(C) = E_d^2(C) + E_r(C)$$

$$E_d^2 = \sum_i (f(v_i, C) - f(v_i, C^0))^2$$

$$E_r^2(C) = -2 \sum_i (f(v_i, C) - f(v_i, C^0))n_i + \sum_i n_i^2$$

We begin by investigating its deterministic part. It is easy to verify that $E_d^2(C)$ has a zero exactly at :

$$C = (c_1^0, \dots, c_n^0)$$

$$E_d^2(C^0) = (f(v_i, C^0) - f(v_i, C^0))^2 = 0 \quad (8a.)$$

by expanding $E_d^2(C)$ at C^0 we shall show that the function $E_d^2(C)$ is a unimodal function around C^0 .

$$\begin{aligned}
 E_d^2(C^0 + \Delta C) &= \sum_i (f(v_i, C^0 + \Delta C) - f(v_i, C^0)) \\
 &\approx \sum_i \{f(v_i, C^0) + \sum_j \frac{\partial f(v_i, C^0)}{\partial c_j} \Delta c_j - f(v_i, C^0)\}^2 \\
 E_d^2(C^0 + \Delta C) &= \sum_i \left\{ \sum_j \frac{\partial f(v_i, C^0)}{\partial c_j} \Delta c_j \right\}^2
 \end{aligned}$$

Using the well known schwarz inequality :

$$(\sum_j a_j b_j) \leq \sqrt{\sum_i a_i^2} \sqrt{\sum_i b_i^2}$$

and substituting :

$$a_j = \frac{\partial f(v_i, C)}{\partial c_j}$$

$$b_j = \Delta c_j$$

we finally have :

$$E_d^2(C+\Delta C) < \sum_i \{ \sum_k \Delta c_k^2 \sum_j \left(\frac{\partial f(v_i, C^0)}{\partial c_j} \right)^2 \} \quad (8b.)$$

$$= r^2 \sum_{ji} \left(\frac{\partial f(v_i, C^0)}{\partial c_j} \right)^2$$

$$\Delta C = (\Delta c_1, \dots, \Delta c_n)$$

$$r^2 = \sum_j \Delta c_j^2$$

Thus the function around C^0 is a unimodal function. The error E_d^2 decreases the closer we are to the actual values - C^0 . This is not so for every C . It is possible that $|C_1 - C^0| < |C_2 - C^0|$, and yet $E_d^2(C_1) > E_d^2(C_2)$. But, still, we shall show that it is always possible to find ΔC such that :

$$E_d^2(C+\Delta C) < E_d^2(C)$$

This result will establish that we can use a search technique based on repeating local improvement, since from any given C it is possible to reach the point $E_d^2(C) = 0$.

The change in $E_d^2(C)$ along ΔC is given by :

$$\Delta E_d^2(C) = \nabla E_d^2 \Delta C$$

The right hand side term consists of the product of ΔC and ∇E - the gradient of E

$$\begin{aligned} \Delta E_d^2 &= \nabla \sum_i (f(v_i, C) - f(v_i, C^0))^2 \\ \Delta E_d^2 &= 2 \sum_i (f(v_i, C) - f(v_i, C^0)) \nabla f(v_i, C^0) \end{aligned} \quad (8c.)$$

where :

$$\nabla f(v_i, C) = \left(\frac{\partial f(v_i, C)}{\partial c_1}, \dots, \frac{\partial f(v_i, C)}{\partial c_n} \right)$$

The difference $E_d^2(C)$ has the form of :

$$\Delta E_d^2 = \Delta C \left(\sum_i A_i q_i \right) \quad (9.)$$

where :

$$A_i = 2(f(v_i, C) - f(v_i, C^0)) \quad (9a.)$$

$$q_i = \nabla f(v_i, C) = \left(\frac{\partial f(v_i, C)}{\partial c_1}, \dots, \frac{\partial f(v_i, C)}{\partial c_n} \right)$$

None of the vectors q_i are zero, since the components of each vector are a linear combination of exponential functions with positive coefficients, $\{b_n\}$, only (see Eq.3_a and App.5). A_i may be zero, positive or negative. But it could not be that all of the $\{A_i\}$, for some $C = [c_1, c_2, \dots, c_n]$ be zero. Since that means two different mixtures have the same spectrum output for two different concentration C and C^0 .

If for some ΔC $E_d^2(C)$ is less than zero, than we already have $C' = C + \Delta C$ that yields a smaller $E_d^2(C)$.

If $E_d^2(C) > 0$ it is enough to choose $\Delta C' = -\Delta C$

$$\Delta C' \sum_i A_i q_i = -\Delta C \sum_i A_i q_i < 0$$

If $\Delta E_d^2(C) = 0$ for some ΔC , it is enough to choose some other ΔC that yields $\Delta E_d^2(C) \neq 0$ which exists, since some A_i are not zero. Consequently, the search rule should be a rule that yields a path along where $E_d^2(C)$ decreases. This will bring us, eventually, to the point where $E_d^2(C) = 0$ for every C . At this point we have : $C = C^0$.

The random part of the error is :

$$E_r(C) = -2 \sum_i (f(v_i, C) - f(v_i, C^0)) n_i + \sum_i n_i^2 \quad (9b.)$$

We have shown that the deterministic part of the error function $E_d^2(C)$ has exactly one absolute minimum and that minimum may be reached from any C . This minimum would bring us exactly to $C = C^0$. We also have shown that $E_d^2(C)$ is unimodal around C^0 (see 8b.) with a magnitude that

decrease to zero.

The search is terminated when the gradient has a magnitude which is less than some ϵ and, therefore, it is impossible to reduce $E_d^2(C)$ for any C . We shall show that this condition leads to a minimum value of $E_d^2(C)$ that may be obtained.

As before, we write the gradient as the sum of the deterministic part and the random part :

$$E^2 = E_r + E_d^2$$

$$\nabla E_r = \left(\frac{\partial E_r(C)}{\partial c_1}, \dots, \frac{\partial E_r(C)}{\partial c_n} \right)$$

substituting (9b.) :

$$\nabla E_r = -2 \left\{ \frac{\partial}{\partial c_1} \sum_i f(v_i, C) n_i, \dots, \frac{\partial}{\partial c_n} \sum_i f(v_i, C) n_i \right\}$$

substituting (9a.), we have :

$$\Delta E_r = -2 \sum_i q_i n_i$$

and from (9.), we have :

$$\nabla E_d^2 = \sum_i A_i q_i$$

Finally, substituting the above into the equation of $\nabla E^2(C)$, we have :

$$\nabla E^2 = \nabla E_d^2 + \nabla E_r = \sum_i A_i q_i - 2 \sum_i q_i n_i$$

ΔE_r is a linear combination of random gaussian variables. Also, we note that the gradient value is changed mainly through $\{A_i\}$, that is the change is in the deterministic part, since $\{A_i\}$ consist of differences between two quantities (see 9a)

We suggest that the iteration process of finding the minimum $E^2(C)$ can be carried on as long as the deterministic part of the gradient has a larger magnitude than the random part has (see Fig.3).

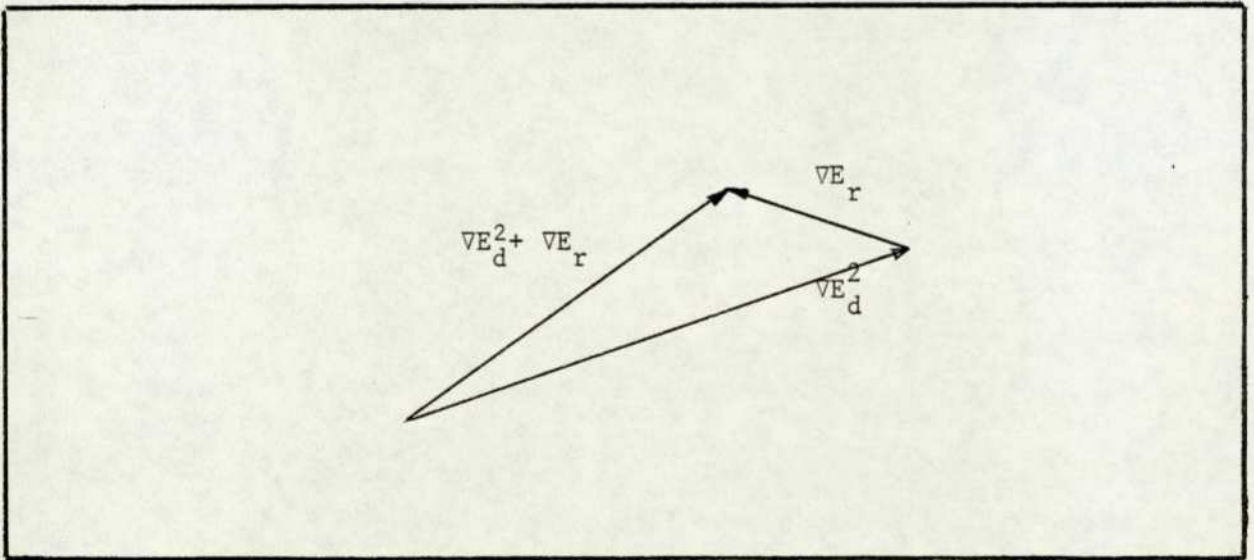


Fig. 3 The error in the gradient is smaller the smaller is the ratio $|\nabla E_r|/|\nabla E_d^2|$

$E_d^2(C)$ which is :

$$E_d^2(C) = \frac{1}{2} \sum_i A_i^2 = \sum_i (f(v_i, C) - f(v_i, C^0))$$

has a minimum value, above zero, that may be reached and is expressed by :

$$|\sum_i A_i q_i| > |2 \sum_i q_i n_i|$$

As an indication of the magnitude that the random part can have, we take its statistical mean value :

$$4 \left(\overline{\sum_i q_i n_i} \right) \left(\overline{\sum_i q_i n_i} \right) = 4 \sum_i q_i q_j \overline{n_i n_j}$$

the noise samples are uncorrelated and with a variance N_0 . We have :

$$\{4 \sum_{i,j} q_i q_j n_i n_j\}^{1/2} = 2N_0 \left(\sum_i q_i^2 \right)^{1/2}$$

Consequently, the values of $\{A_i\}$ are bounded by the expression :

$$\sum_i A_i q_i > 2N_0 \left(\sum_i q_i^2 \right)^{1/2}$$

The expression above may be developed further. Define $q_{\max}$ as :

$$|q_{\max}| > |q_i|$$

$$i = 1, M$$

we have :

$$\left| \sum_i A_i q_i \right| \leq \left| \sum_i A_i q_{\max} \right| = |q_{\max}| \left| \sum_i A_i \right|$$

define $\left| \left(\sum_i A_i \right)_{\min} \right|$ as :

$$\left| \left(\sum_i A_i \right)_{\min} \right| = \frac{2^{N_0^{1/2} \sum (q_i^2)^{1/2}}}{|q_{\max}|}$$

we have :

$$\left| \sum_i A_i \right| \geq \left| \left(\sum_i A_i \right)_{\min} \right| = 2^{N_0^{1/2} \frac{(\sum q_i^2)^{1/2}}{q_{\max}}} \quad (9c.)$$

where :

$$A_i = 2(f(v_i, C) - f(v_i, C^0))$$

We conclude, that although the search is terminated with some error, this error is proportional to the noise power. The lower the noise power the closer the estimate to $(c_1^0, c_2^0, \dots, c_n^0)$.

The estimation problem, when tackled from the statistical point of view leads to the same results.

Given a set of $c_1, c_2, \dots, c_n$, the probability density of having set of measurements $y_1, y_2, \dots, y_M$ is given by :

$$p(y/C) = \frac{1}{(2\pi N_0)^{1/2}} \exp\left\{-\frac{1}{2N_0} \sum_i (y_i - f(v_i, C))^2\right\} \quad (10.)$$

This is so, since from (6.) we have :

$$n_i = y_i - f(v_i, C)$$

where $\{n_i\}$ are uncorrelated gaussian variables (see 1.1.2).

Eq.(10.) is known as the Likelihood Function [20]. The vector C that maximizes (10.) is the maximum likelihood estimate. It is possible, instead, to search for a minimum of the exponential part of (10.)

$$\sum_i [y_i - f(v_i, C)]^2 \xrightarrow{C} \text{Min} \quad (11.)$$

Thus the Maximum Likelihood Estimator yields the same error function as in (7.). (11.) may be written also :

$$\sum_i (y_i - f(v_i, C))^2 = \sum_i y_i^2 - 2 \sum_i y_i f(v_i, C) + \sum_i f^2(v_i, C) \quad (12.)$$

For a set of measurements $y_1, y_2, \dots, y_M$, the first term in (12.) is constant. Consequently, it is enough to find a minimum to the following expression :

$$G(C) = \sum_i f^2(v_i, C) - 2 \sum_i y_i f(v_i, C) \xrightarrow{C} \text{Min.}$$

A necessary and sufficient condition for finding C is given by :

$$\Delta C \nabla G(C) = 0$$

where ΔC is arbitrary. Consequently, every component of $\text{grad}(G)$ should be equal to zero.

Which brings us to the following set of equations :

$$\nabla G(C) = 0$$

$$\sum_i (f(v_i, C) - y_i) \frac{\partial f(v_i, C)}{\partial c_j} = 0 \quad (13.)$$

$$j = 1, \dots, n$$

as it was shown, C that fulfil the above is such that :

$$\nabla \{E_d^2(C) + E_r(C)\} = 0$$

where the statistical mean of $|\nabla E_r(C)|$ is given by :

$$|\nabla E_r(C)| = 2N_0^{1/2} \sum_i (\langle \nabla f(v_i, C)^2 \rangle)^{1/2}$$

In the following subsections of this chapter a solution of the estimation problem is given to the case of a perturbation around the actual concentration C^0 . The result enables us to draw conclusions on the nature of the general case. Since a successfull search for $\hat{C}$

should be terminated in the approximate of C^0 .

4.4 The solution in the case of small perturbation and perfect resolution spectrum

In this subsection, the estimation problem in the case of a spectrophotometer output signal with perfect resolution (That is, the spectrum given by Beer - Lambert law, see Sec.4.2) is first solved. Therefore, the stage of converting the spectrum to a spectrum with a given resolution is not necessary. The formula for deriving the correction vector $(\Delta c_1, \dots, \Delta c_n)$ follows the solution given by Gauss and Lagender [21]. An expression for calculating the rms error is derived.

Reducing the complexity of the problem even further to the linear case (which may be regarded as the estimation problem in the case of a very low concentration) enables us to identify different expression and to generalize the S/N ratio, the correlation factor and the signal power. It is shown that these may be used in analysing other cases of estimation without carrying out the simulation process.

The optimum estimate $\hat{C}$ satisfies the following set of equations (see Eq.13):

Ch. 4

$$\sum_i (f(v_i, C) - y_i) \frac{\partial f(v_i, C)}{\partial c_j} = 0 \quad (14.)$$

$$j = 1, \dots, M$$

Where $f(v_i, C)$ is given by :

$$f(v_i, C) = \sum_n b_n x(v_i + nV) \quad (14a.)$$

$$x(v_i) = \exp(- \sum_j c_j \alpha_j(v_i))$$

and V is the sampling period.

In case of perfect resolution spectrophotometer, we have :

$$b_n = \begin{cases} 1 & n=0 \\ 0 & n \neq 0 \end{cases} \quad (15.)$$

That is, the effect of the spectrophotometer transfer function is negligible. This is the case where $1/(\text{slit width})$ is very small compared to the band width of the signal spectrum.

Substituting the above into (14a.), we have :

$$f(v_i, C) = x(v_i) = \exp(- \sum_j c_j \alpha_j(v_i)) \quad (15a.)$$

Eq.(14.) can also be written :

$$\frac{\partial}{\partial c_j} \sum_i (f(v_i, C) - y_i)^2 = 0 \quad (15b.)$$

The error function is defined (see Eq. 7) :

$$E^2(C) = \sum_i (f(v_i, C) - y_i)^2 \quad (15c.)$$

Substituting (15c.) into (15b.) we have :

$$\frac{\partial}{\partial c_j} E^2(C) = 0$$

$$j = 1, \dots, n$$

In vector notation, we have :

$$\nabla E^2 = 0 \quad (15d.)$$

The error function in (15c.) can be written in vector notation :

$$\begin{bmatrix} e_1, & \dots, e_M \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_M \end{bmatrix}$$

Which can also be written as :

$$E^2(C) = e'e$$

where a component e_i of e is given by :

$$e_i = f(v_i, C) - y_i$$

$$i = 1, \dots, M$$

Now, carrying out (15d.), we have :

$$\nabla E^2(C) = \nabla e_i^2 = 2 \sum e_i \nabla e_i \quad (16.)$$

$$2 \sum e_i \nabla e_i = 2e' \nabla e$$

where the gradient is taken over each component e_i of e . The definition of $\nabla(e_i)$ is :

$$\nabla e_i = \left(\frac{\partial e_i}{\partial c_1}, \dots, \frac{\partial e_i}{\partial c_n} \right)$$

$$\frac{\partial e_i}{\partial c_j} = \frac{\partial e(v_i, C)}{\partial c_j} = \frac{\partial}{\partial c_j} (f(v_i, C) - y_i) = \frac{\partial}{\partial c_j} f(v_i, C)$$

Substituting Eq.(15a.) and differentiating the above equation :

$$\nabla' e(v_i, C) = -(\alpha_1(v_i) f(v_i, C), \dots, \alpha_n(v_i) f(v_i, C)) \quad (16a.)$$

Consequently, Eq.(16.), also may be written :

$$E^2(C) = \begin{bmatrix} \nabla e_1, & \nabla e_M \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_M \end{bmatrix}^2$$

Substituting (16a.) in the above equation, we have :

$$\nabla E^2(C) = -2 \begin{bmatrix} \alpha_1(v_1)f(v_1,C), & ,\alpha_1(v_M)f(v_M,C) \\ \vdots & \vdots \\ \alpha_n(v_1)f(v_1,C), & ,\alpha_n(v_n)f(v_M,C) \end{bmatrix} \begin{bmatrix} \bar{e}_1 \\ \vdots \\ e_M \end{bmatrix} \quad (16c.)$$

The 2-dimensional matrix in (16b.) is known as the Jacobian of the set of function $(e_1, ,e_M)$ [21] :

$$\nabla e' = J = (\nabla e_1, ,\nabla e_M)$$

$$\nabla e' = J = - \begin{bmatrix} \alpha_1(v_1)f(v_1,C), & ,\alpha_1(v_M)f(v_M,C) \\ \vdots & \vdots \\ \alpha_n(v_1)f(v_1,C), & ,\alpha_n(v_M)f(v_M,C) \end{bmatrix} \quad (16d.)$$

Substituting (16d.) in (16c.) and in (15d.), we have :

$$\nabla E^2 = 2Je = 0 \quad (16e.)$$

As the set of equation (16e.) is non linear in C, an iteration technique is adopted. The technique is based on the Gaussian iteration for n variables [21]. In this technique, ΔC is approximated such that :

$$\nabla E^2(C + \Delta C) = 0$$

ΔC is substituted, the process is repeated till no correction is possible.

$$\nabla E^2(C + \Delta C) = 2J(C + \Delta C)e(C + \Delta C) = 0 \quad (17.)$$

Assuming :

$$J(C + \Delta C) = J(C)$$

In other words, the change in J , due to a shift by ΔC is neglected.

Also :

$$e(C + \Delta C) = e(C) + \nabla e \Delta C$$

That is, ΔC is small enough to allow first approximation. We have (see 16d.) :

$$J' = \nabla' e \quad (18.)$$

Substituting the above Eq. to (17.), we have :

$$E^2(C + \Delta C) = 2J(C)(e + J' \Delta C) = 0$$

$$J(C)e + JJ' \Delta C = 0$$

JJ' is $n \times n$ matrix that has inverse [21]. Consequently, the formula for ΔC is give by :

$$\Delta C = -(JJ')^{-1} J e \quad (19.)$$

Define :

$$JJ' = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix} \equiv R \quad (19a.)$$

$$R = \begin{bmatrix} \alpha_1(v_1)f(v_1, C), & \alpha_1(v_M)f(v_M, C) \\ \vdots & \vdots \\ \alpha_n(v_1)f(v_1, C), & \alpha_n(v_M)f(v_M, C) \end{bmatrix} \begin{bmatrix} \alpha_1(v_1)f(v_1, C), & \alpha_1(v_1)f(v_1, C) \\ \vdots & \vdots \\ \alpha_1(v_M)f(v_M, C), & \alpha_n(v_M)f(v_M, C) \end{bmatrix}$$

We have :

$$R_{kj} = \sum_i \alpha_k(v_i) \alpha_j(v_i) f^2(v_i, C) \quad (20.)$$

$$J e = \begin{bmatrix} \alpha_1(v_1)f(v_1, C), & \alpha_1(v_M)f(v_M, C) \\ \vdots & \vdots \\ \alpha_n(v_1)f(v_1, C), & \alpha_n(v_M)f(v_M, C) \end{bmatrix} \begin{bmatrix} f(v_1, C) - y_1 \\ \vdots \\ f(v_M, C) - y_M \end{bmatrix} \quad (21.)$$

Define :

$$J_e = \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yM} \end{bmatrix} \quad (21a)$$

$$R_{yj} = - \sum_i \alpha_j(v_i) f(v_i, C) (f(v_i, C) - y_i) \quad (21b)$$

Substituting (21a.) and (19a.) into (19.), we have :

$$\Delta C = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix} \quad (22.)$$

The above formula is a formula for finding the correction ΔC . In the following subsection a derivation of the estimation variance is obtained.

4.4.1 Derivation of the estimation variance

In the following subsection an expression for the estimation variance is derived. In Sec.4.4.2 interpretation of the expression found is given, and in Ch.5 the expression is compared to the rms error calculated in simulation of some estimations. The validation of the expression enables us to generalised the expression of the S/N ratio such that its magnitude indicates the rms error expected in the estimation process. The generalization is done by studying the error expression in the linear case. In this case, the error is also a function of the correlation factors and the S/N ratio. Substituting $C = C^0$, we have (see Eq. 6.) :

$$y_i = \exp\left(-\sum_j c_j \alpha_j(v_i)\right) + n_i \quad (22a.)$$

$$R_{yj} = -\sum_i \alpha_j(v_i) f(v_i, C) n_i$$

Substituting (22a.) into (22.), we have :

$$\Delta C = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} \sum_i \alpha_1(v_i) f(v_i, C) n_i \\ \vdots \\ \sum_i \alpha_n(v_i) f(v_i, C) n_i \end{bmatrix} \quad (23.)$$

Thus ΔC is a vector of random Gaussian variables, with a zero mean.

The variance in each estimate is given by :

$$\overline{\begin{bmatrix} \Delta c_1 \\ \vdots \\ \Delta c_n \end{bmatrix} \begin{bmatrix} \Delta c_1, & \Delta c_n \end{bmatrix}} = \begin{bmatrix} \overline{\Delta c_1^2}, & \overline{\Delta c_1 \Delta c_n} \\ \overline{\Delta c_n \Delta c_1}, & \overline{\Delta c_n^2} \end{bmatrix} \quad (23a.)$$

Substituting (23.) into (23a.), we have :

$$(\Delta C)(\Delta C)' = \begin{bmatrix} R_{11}, & ,R_{1n} \\ \vdots & \vdots \\ R_{n1}, & ,R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix} \begin{bmatrix} R_{y1}, & ,R_{yn} \end{bmatrix} \begin{bmatrix} R_{11}, & ,R_{1n} \\ \vdots & \vdots \\ R_{n1}, & ,R_{nn} \end{bmatrix}^{-1} \quad (24.)$$

From (22a.), we have :

$$R_{yi} R_{yj} = \sum_k \alpha_i(v_k) f(v_k, C) n_k \sum_m \alpha_j(v_m) f(v_m, C) n_m$$

$$\overline{R_{yi} R_{yj}} = \sum_{k,m} \alpha_i(v_k) f(v_k, C) \alpha_j(v_m) \overline{n_k n_m}$$

Taking the mean value of the above expressions, and bearing in mind that the noise components are uncorrelated (see 5.), we have

$$\overline{R_{yi} R_{yj}} = \sum_k \alpha_i(v_k) \alpha_j(v_k) f^2(v_k, C) \overline{n_k^2} \quad (25.)$$

We have (see 5.) :

$$\overline{n^2} = N_0$$

substituting the above into (25.), and with (20.), we have :

$$\overline{R_{yi}R_{yj}} = N_0 R_{ij} \quad (26.)$$

Finally, substituting into (24.) the expression above, we find the expression for the variance in each estimates :

$$\begin{bmatrix} \overline{\Delta c_1^2}, & \overline{\Delta c_1 \Delta c_n} \\ \vdots & \vdots \\ \overline{\Delta c_n \Delta c_1}, & \overline{\Delta c_n^2} \end{bmatrix} = N_0 R^{-1} R' R'^{-1}$$

The final expression for $\overline{\Delta c_1^2}$, $\overline{\Delta c_n^2}$ is :

$$\begin{bmatrix} \overline{\Delta c_1^2}, & \overline{\Delta c_1 \Delta c_n} \\ \vdots & \vdots \\ \overline{\Delta c_n \Delta c_1}, & \overline{\Delta c_n^2} \end{bmatrix} = N_0 R^{-1} \quad (27.)$$

where R_{ij} is given by :

$$R_{ij} = \sum_j \alpha_j(v_k) \alpha_j(v_k) f^2(v_k, C)$$

Thus we have shown that the variance of the fluctuation in C, around the solution is given by the equation above.

$\overline{(\Delta c_1^2, \Delta c_2^2, \dots, \Delta c_n^2)}$ may also be considered as the variance of the estimates vector $C = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$. The justification for this assumption is as follows :

In Sec. 4.2 it is shown that in a noiseless case a search technique when applied to a signal with unknown concentration(s) terminates exactly at $C = C^0$. In a case of a signal with additive noise, it is shown (see Eq.9a.) that the estimates fluctuate around C^0 with a rms value proportional to the noise power. In practice the S/N ratio is high enough such that the search for $\hat{C}$ is terminated in the approximate of C^0 such that the expression for $\overline{\Delta c_j^2}$, found above, approximates the rms error in the estimation.

In Ch.5 a comparison between the actual rms error and the theoretical rms error is given for several values of S/N ratio, which justifies our assumptions.

4.4.2 Interpretation of the expression of the variance

The following examples will help us to understand the meaning of the above expressions.

Example 1

Single concentration

$$C = c_1 = c$$

From Eq. 20 :

$$R_{11} = \sum_k \alpha^2(v_k) \exp(-2c\alpha(v_k))$$

From Eq. 21b. :

$$R_{11} = \sum_k \{ \alpha(v_k) \exp(-c\alpha(v_k)) \} \{ \exp(-c\alpha(v_k)) - y_k \}$$

$$\Delta c = \frac{\sum_i \{ \alpha(v_i) \exp(-c\alpha(v_i)) \} \{ \exp(-c\alpha(v_i)) - y_i \}}{\sum_k \alpha^2(v_k) \exp(-2c\alpha(v_k))}$$

Example 2

The linear case

Assuming :

$$f(v_i, c) = ca_i \quad (28.)$$

 $\nabla(f)$ is given by :

$$\nabla f = (a_1, a_2, \dots, a_M)$$

$$R_{y1} = \sum_i a_i (y_i - ca_i)$$

$$\Delta c = \frac{\sum_i a_i (y_i - a_i)}{\sum_i a_i^2} = -c \frac{\sum_i a_i^2}{\sum_i a_i^2} + \frac{\sum_i a_i y_i}{\sum_i a_i^2}$$

$$\Delta c = -c + \frac{\sum_i a_i y_i}{\sum_i a_i^2}$$

$$\hat{c} = \Delta c + c = \frac{\sum_i a_i y_i}{\sum_i a_i^2} \quad (28a.)$$

That is, in the linear case the estimate $\hat{c}$ is calculated directly, without iteration. Assuming noiseless case, we have :

$$y_i = c^0 a_i$$

$$i = 1, L$$

Substituting into (28a.), we have

$$\hat{c} = \frac{\sum_i a_i y_i}{\sum_i a_i^2} = c^0 \frac{\sum_i a_i^2}{\sum_i a_i^2} = c^0$$

Which verifies (28a.) : Indeed substituting (28.) into (13.), we have :

$$\sum_i (c a_i - y_i) a_i = 0$$

In vector notation :

$$(cA - y)A = 0 \quad (30.)$$

The equation above describes the orthogonality principle [19].

Substituting $y_i = a_i c + n_i$ in (28.), we have :

$$\begin{aligned} \hat{c} &= \frac{\sum_i (a_i^2 c^0 + a_i n_i)}{\sum_i a_i^2} \\ \hat{c} &= c^0 + \frac{\sum_i a_i n_i}{\sum_i a_i^2} \end{aligned}$$

Thus the estimator in the linear case is unbiased [20] and with a variance :

$$\overline{(\hat{c} - c^0)^2} = \frac{(\sum_i a_i n_i)^2}{(\sum_i a_i^2)^2}$$

Since the noise components are uncorrelated, we have (see Eq. 5) :

$$\overline{(\hat{c} - c^0)^2} = \frac{\sum_i a_i a_j \overline{n_i n_j}}{\sum_i a_i^2} = \frac{\sum_i a_i^2 \overline{n_i^2}}{\sum_i a_i^2}$$

$$\overline{n_i^2} = N_0$$

$$\overline{(\hat{c} - c^0)^2} = \frac{N_0}{\sum_i a_i^2} = \frac{1}{S^*/N_0} \quad (31.)$$

where the S/N is the Signal to Noise ratio.

Example 3 Linear case, 2 - concentrations

In case of $C = (c_1, c_2)$, we have :

$$y_i = a_i^1 c_1 + a_i^2 c_2 + n_i$$

$$i = 1, \dots, M$$

The variance in c_1 and c_2 is given by (see App.4) :

$$\overline{(c_i - c_i^0)^2} = \frac{1}{S_j^* / N_0 (1 - \rho^2)} \quad (32.)$$

$$i = 1, \dots, M$$

$$S_j^* = \sum_i (a_i^j)^2$$

$$\rho^2 = \frac{(\sum_i a_i^1 a_i^2)^2}{\sum_i (a_i^1)^2 \sum_i (a_i^2)^2}$$

We conclude, in the linear case the estimate variance is inversely proportional to the orthogonal part of the partial signal given by : $\{a_1^i, a_2^i, \dots, a_M^i\}$. This geometric interpretation is justified, since ρ is exactly the cosine between the two vectors :

$$A_1 = (a_1, a_2, \dots, a_M)$$

and

$$A_2 = (a_1, a_2, \dots, a_M)$$

see Fig.4

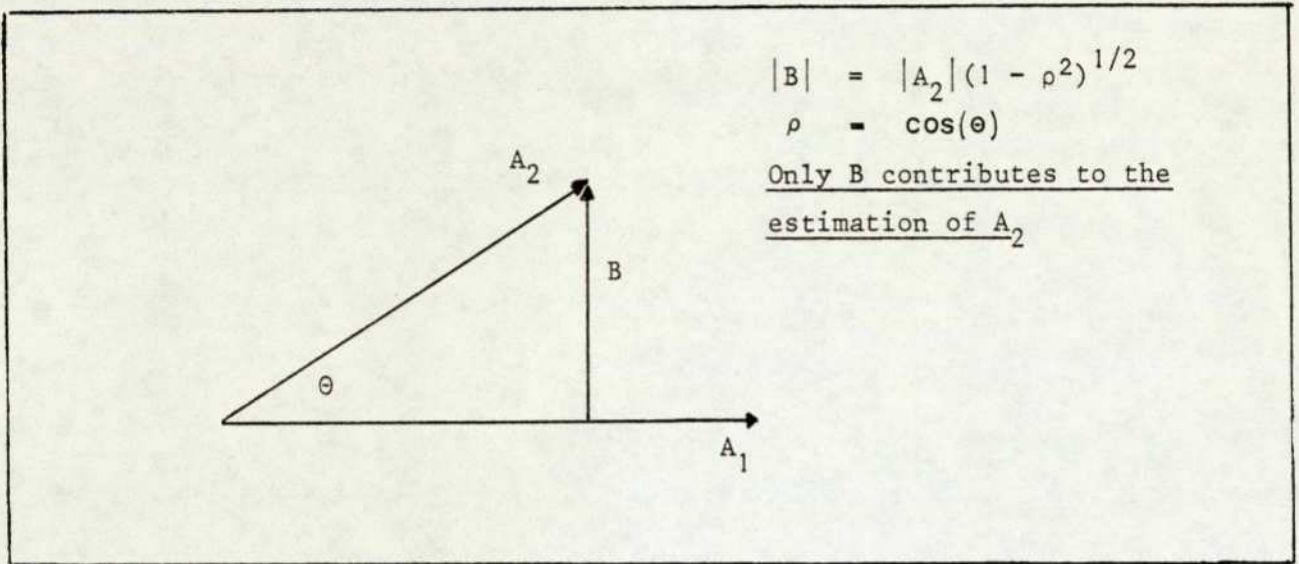


Fig. 4 Geometrical interpretation for the estimation problem in the linear case.

We have (see Fig. 4) :

$$1 - \rho^2 = 1 - \cos^2(\theta) = \sin^2(\theta)$$

$$S_2^*(1 - \rho^2) = \sum_i (a_i^2)^2 \sin^2(\theta) = \left\{ \left(\sum_i (a_i^2)^2 \right)^{1/2} \sin(\theta) \right\}^2$$

The last expression is the square of the component of A_2 that is orthogonal to A_1 .

In case of more than one concentration, the variance of the estimates $(c_1, c_2, \dots, c_n)$ is given by (see App.4) :

$$\begin{bmatrix} \overline{(\hat{c}_1 - c_1^0)^2} & \overline{(\hat{c}_1 - c_1^0)(\hat{c}_n - c_n^0)} \\ \overline{(\hat{c}_n - c_n^0)(\hat{c}_1 - c_1^0)} & \overline{(\hat{c}_n - c_n^0)^2} \end{bmatrix} = \begin{bmatrix} R_{11}^* & R_{1n}^* \\ R_{n1}^* & R_{nn}^* \end{bmatrix}^{-1} N_0 \quad (33.)$$

where :

$$R_{ij}^* = \sum_k a_k^i a_k^j$$

We have $R_{ij} = 0$ only if the set of $\{a_j^i\}$ consists of orthogonal family of vectors :

$$\sum_k a_k^i a_k^j = 0$$

$$i \neq j$$

This is the case where the angle between the two vectors is 90 degrees (see Fig.4).

4.4.3

The modified signal power

By comparing the expression of the variance in the linear case and the nonlinear case, it is possible to identify expression that is equivalent to signal power in the linear problem. The terms are similar in the sense that the error in the nonlinear problem depends on this term exactly as the error in the linear case depends on the S/N ratio. Rewriting (27.) the variance of the estimates in the nonlinear case :

$$\begin{bmatrix} \overline{\Delta c_1^2} & , & \overline{\Delta c_1 \Delta c_n} \\ \overline{\Delta c_1 \Delta c_n} & , & \overline{\Delta c_n^2} \end{bmatrix} = \begin{bmatrix} R_{11} & , & R_{1n} \\ R_{n1} & , & R_{nn} \end{bmatrix}^{-1} N_0$$

where :

$$R_{ij} = \sum_k \alpha_i(v_k) \alpha_j(v_k) f^2(v_k, C)$$

Comparing (33.) the variance in the linear case to the equation above the variance in the nonlinear case, we see that the two equations have the same form.

Consequently, R_{ij} above may be regarded as the modified correlation factor :

$$R_{ij} = \sum_k \{ \alpha_i(v_k) f(v_k, C) \} \{ \alpha_j(v_k) f(v_k, C) \}$$

In the same way, the modified signal is defined :

$$g_j(v_i) = \alpha_j(v_i)f(v_i, C) = \alpha_j(v_i)\exp(-\sum_k c_k \alpha_k(v_i)) \quad (34.)$$

Therefore, the variance in the case of single concentration (see Ex. 1) may be written :

$$\overline{\Delta c^2} = \frac{1}{S/N_0} = \frac{1}{1/N_0 \sum_i g^2(v_i)} \quad (35.)$$

$$S = \sum_i g^2(v_i)$$

$$N_0 = \overline{n_i^2} \quad i = 1, \dots, M$$

In case of two concentrations :

$$C = (c_1, c_2)$$

We have :

$$\overline{\Delta c_j^2} = \frac{N_0}{S_j(c_1, c_2)(1 - \rho^2(c_1, c_2))} \quad (36.)$$

S_j is the modified signal power :

$$S_j = \sum_i g_j^2(v_i) = \sum_i \alpha_j^2(v_i) f^2(v_i, C) \quad (36a.)$$

the modified correlation coefficient is given by :

$$\rho^2 = \frac{\left\{ \sum_i g_1(v_i) g_2(v_i) \right\}^2}{\sum_i g_1^2(v_i) \sum_i g_2^2(v_i)} \quad (37.)$$

Fig.4 gives a geometrical interpretation to the role the above equations play in the linear case. With contrast to the linear case, where the angle between the two vectors is constant (see Fig.4) and it is not a function of the concentrations. the angle between the two vectors in the nonlinear case is a function of the concentrations.

Some practical conclusions may be drawn from the above results. The modified signal power has a maximum value as a function of the cell length, consequently there is an optimum cell length that yields a minimum error.

S_j the modified power is given by :

$$S_j = \sum_i \alpha_j^2(v_i) \exp(-2 \sum_k c_k \alpha_k(v_i)) \quad (37a.)$$

$\alpha_j(v_i)$, the absorption coefficient is given by (see Ch.1, Eq.1a) :

$$\alpha_j(v_i) = \alpha_j^0(v_i)L \quad (38.)$$

where L is the cell length.

From Eq.(37a.) it is obvious that S_j the modified power has a maximum as a function of the cell length L . Consequently, there is a cell length that will yield a minimum error in estimation the concentration.

This phenomenon, which has practical application is discussed in more details in Ch.5, where some simulation shows this maximum.

4.5 The general estimation problem (imperfect resolution)

In the previous chapter, the non linear problem in the case of high resolution spectrophotometer has been treated. It has been shown that the estimation accuracy (the variance) and other features like the S/N (Signal/Noise) may be defined and play similar role as in the linear case. Moreover, in contrast to the linear case it has been shown that higher intensity (higher signal power) does not necessarily mean a better accuracy in the estimation results as it is in the linear case (see Sec. 4.4.3).

In this chapter, a further step is made in generalizing the estimation model. The assumption of high resolution spectrophotometer is dropped, we have a spectrum signal with arbitrary resolution. The general set of equations is (see Eq.6) :

$$y(v_i) = f(v_i, C) + n_i \quad (39.)$$

$$x(v_i) = \exp(-\sum_j c_j \alpha_j(v_i))$$

$$f(v_i, C) = \sum_{n=-N}^N b_n x(v_i + nV)$$

Again, the error function is given by :

$$E^2(C) = \sum_i (f(v_i, C) - y_i)^2$$

The solution to the general case developed here goes along the lines of argument developed in the last section. It is shown that the solution as well as the estimate variance can be brought to the same form as in the linear case. Thus engineering tools like the S/N ratio, signal power and correlation factor may be generalized so that they will play the same role as in the linear case.

The general equation for the correction ΔC in the current estimate C is given by (See Eq.19) :

$$C = -(JJ')^{-1} J e \quad (40.)$$

where:

$$e' = (e_1, \dots, e_M)$$

$$e_i = f(v_i, C) - y_i$$

$$J = \begin{bmatrix} \nabla' e_1, & \dots, \nabla' e_M \\ \vdots & \end{bmatrix}$$

$$\nabla e = \nabla(f(v_i, C) - y_i) = \nabla f(v_i, C)$$

Substituting Eq.(39.), we have :

$$\frac{\partial}{\partial c_j} f(v_i, C) = \frac{\partial}{\partial c_j} \sum_{n=-N}^N b_n x(v_i + nV)$$

again, using Eq.(39.) and substituting x , we have :

$$\frac{\partial}{\partial c_j} x(v_i + nV) = \frac{\partial}{\partial c_j} \exp(-\sum_k c_k \alpha_k (v_i + nV))$$

$$\frac{\partial}{\partial c_j} x(v_i + nV) = -\alpha_j (v_i + nV) x(v_i + nV)$$

$$\frac{\partial}{\partial c_j} \sum_{n=-N}^N b_n x(v_i + nV) = -\sum_{n=-N}^N b_n \alpha_j (v_i + nV) x(v_i + nV)$$

$$\nabla' e_i = \left\{ \sum_{n=-N}^N b_n \alpha_1 (v_i + nV) x(v_i + nV), \dots, \sum_{n=-N}^N b_n \alpha_M (v_i + nV) x(v_i + nV) \right\}$$

Define :

$$g_j(v_i) = \sum_{-N}^N b_n \alpha_j(v_i + nV) x(v_i + nV)$$

$$\nabla e_i' = (g_1(v_i), \dots, g_M(v_i))$$

As in the previous section :

$$J = (\nabla' e_1, \dots, \nabla' e_M)$$

in matrix form :

$$J = \begin{bmatrix} g_1(v_1), & \dots, g_1(v_M) \\ \vdots & \vdots \\ g_n(v_1), & \dots, g_n(v_M) \end{bmatrix}$$

We have :

$$JJ' = \begin{bmatrix} g_1(v_1), & \dots, g_1(v_M) \\ \vdots & \vdots \\ g_n(v_1), & \dots, g_n(v_M) \end{bmatrix} \begin{bmatrix} g_1(v_1), & \dots, g_n(v_1) \\ \vdots & \vdots \\ g_1(v_M), & \dots, g_n(v_M) \end{bmatrix}$$

$$JJ' = \begin{bmatrix} \sum_i g_1^2(v_i) & \dots, \sum_i g_1(v_i) g_n(v_i) \\ \vdots & \vdots \\ \sum_i g_n(v_i) g_1(v_i), & \dots, \sum_i g_n^2(v_i) \end{bmatrix}$$

$$JJ' = \begin{bmatrix} R_{11}, & , R_{1n} \\ R_{n1}, & , R_{nn} \end{bmatrix}$$

where :

$$R_{ij} = \sum_k g_i(v_k) g_j(v_k)$$

$$R_{ij} = \sum_m \left\{ \sum_k b_k^{\alpha_i} (v_m + kV) x(v_k + kV) \right\} \left\{ \sum_k b_k^{\alpha_j} (v_m + kV) x(v_m + kV) \right\} \quad (44.)$$

We also have :

$$J_e = \begin{bmatrix} g_1(v_1), & , g_1(v_M) \\ g_n(v_1), & , g_n(v_M) \end{bmatrix} \begin{bmatrix} \bar{f}(v_1, C) - y_1 \\ \bar{f}(v_M, C) - y_M \end{bmatrix}$$

$$J_e = \begin{bmatrix} g_1(v_1), & , g_1(v_M) \\ g_n(v_1), & , g_n(v_M) \end{bmatrix} \begin{bmatrix} R_{y1} \\ R_{yM} \end{bmatrix}$$

where :

$$R_{yj} = - \sum_i \alpha_j(v_i) f(v_i, C) (f(v_i, C) - y_i)$$

Finally, we have :

$$\Delta C = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix}$$

4.5.1 The modified signal power in the case of imperfect spectrum

The variance of the estimate $\hat{C} = (\hat{c}_1, \hat{c}_n)$ is given by (see 27.) :

$$\begin{bmatrix} \overline{\Delta c_1^2}, & \overline{\Delta c_1 \Delta c_n} \\ \vdots & \vdots \\ \overline{\Delta c_n \Delta c_1}, & \overline{\Delta c_n^2} \end{bmatrix} = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} N_0 \quad (48.)$$

where (see 44) :

$$R_{ij} = \sum_m \{ \sum_k b_k \alpha_i(v_m + kV) x(v_m + kV) \} \{ \sum_k \alpha_j(v_m + kV) x(v_m + kV) \}$$

It has been shown that the same expression, consists of correlation coefficients, is derived in the linear case.

Define the modified signal as :

$$g_j(v_i) = \sum_k b_k \alpha_j(v_i + kV) x(v_i + kV)$$

In the case of high resolution spectrophotometer, we have :

$$g_j^*(v_i) = \alpha_j(v_i) x(v_i)$$

consequently :

$$g_j(v_i) = \sum_{-N}^N b_k g_j^*(v_i + kV)$$

We conclude, by defining the G's as the modified signals, the expression of the error can be considered as a function of the correlation coefficients between the modified signals. As in the linear case, having a mixture with a single concentration :

$$\overline{\Delta c^2} = \frac{N_0}{\sum g^2(v_i)}$$

In case of two components, we have :

$$\overline{\Delta c_j^2} = \frac{N_0}{S_j(c_1, c_2)(1-\rho^2(c_1, c_2))} \quad (49.)$$

$$\rho^2 = \frac{(\sum_i g_1(v_i)g_2(v_i))^2}{\sum_i g_1^2(v_i) \sum_i g_2^2(v_i)}$$

$$j = 1, 2$$

4.6 The effect of the resolution on the estimation error

It is obvious that in the case of two compounds mixture, where both compounds have the same spectra shape (the shape of the modified signal) in a given range of wave length, it will not be possible to estimate the concentration of the one in the presence of the other by information based only in the given wave length range. More specifically, assuming :

$$g_2(v_i) = ag_1(v_i)$$

we have (see 49.) :

$$\rho^2 = \frac{\sum_i a^2 g_1^2(v_i) g_1(v_i)}{\sum_i g_1^2(v_i) \sum_i a^2 g_1^2(v_i)}$$

$$\rho^2 = \frac{a^2 \sum_i g_1^4(v_i)}{a^2 \sum_i g_1^4(v_i)} = 1$$

Substituting into (49.) we notice that the error Δc_j^2 increases to infinity.

On the other hand, assuming g_1 and g_2 have only two components respectively :

$$g^1 = (0, g_1(v_2))$$

$$g^2 = (g_2(v_1), 0)$$

In that case :

$$\rho^2 = \sum_i g_1(v_i) g_2(v_i) = 0$$

The rms error is given by :

$$\overline{\Delta c_j^2} = \frac{N_0}{S_j} = \frac{N_0}{\sum_i g_j^2(v_i)}$$

Having a spectrophotometer with a low resolution with, say, (b_0, b_1) non zero only, we have :

$$g_1'(v_1) = b_0 0 + b_1 g_1(v_2) = b_1 g_1(v_2)$$

$$g'_1(v_2) = b_1 0 + b_0 g_1(v_2) = b_0 g_1(v_2)$$

the modified signal power is given by :

$$S_1 = b_1^2 g_1^2(v_2) + b_0^2 g_1^2(v_2) = (b_0^2 + b_1^2) g_1^2(v_2)$$

Assuming the transfer function is normalized such that :

$$b_0^2 + b_1^2 = 1$$

we have :

$$S_1 = g_1^2(v_2)$$

(50.)

$$S_2 = g_2^2(v_1)$$

That is, the signal power is conserved. The modified signal of the second compound is given by :

$$g'_2(v_1) = b_0 g_2(v_1) + b_1 0 = b_0 g_2(v_1)$$

$$g'_2(v_2) = b_0 0 + b_1 g_2(v_1) = b_1 g_2(v_1)$$

the generalized correlation factor is given by :

$$\rho^2 = \frac{b_1^2 g_1^2(v_2) b_0^2 g_2^2(v_1) + b_0^2 g_1^2(v_2) b_1^2 g_2^2(v_1)}{g_1^2(v_2) g_2^2(v_1)}$$

$$\rho^2 = \frac{2b_0^2 b_1^2 g_1^2(v_2) g_2^2(v_1)}{g_1^2(v_2) g_2^2(v_1)} = 2b_0^2 b_1^2$$

Consequently, the error is given by :

$$\overline{\Delta c_i^2} = \frac{N_0}{s_i (1 - 2b_0^2 b_1^2)}$$

$$b_0^2 + b_1^2 = 1$$

Since the factor $1/(s_i)$ is the same as in the last example (see Eq.50.), and the factor $1/(1 - \rho^2)$ increases, the rms error in the estimate increases. That is, decreasing the resolution without increasing the signal power causes degradation of the estimation accuracy. In practice decreasing the resolution causes the increasing of the signal power and therefore, not necessarily the increasing of the error. In fact, there should be a resolution that yield a minimum error. This resolution is exactly the resolution where the normalized change in the modified power is balanced by the normalized change in $(1 - \rho^2)$ factor :

$$\frac{\Delta S_i}{S_i} = \frac{\Delta \rho^2}{\rho^2}$$

$$\frac{\Delta S_i}{S_i} = \frac{2\Delta\rho}{\rho}$$

4.7

Summary and conclusions

In this chapter, the information and conclusion we have reached in the last three chapters were gathered to give us a mathematical model of the beam process in the spectrophotometer. The set of equations of the electro optic process that yields, eventually, the spectrum of a mixture of gases as an output of the spectrophotometer (see Fig. 2) are :

- (i) The physical aspect is given by Beer - Lambert law.
- (ii) The optic process is represented by the slit transfer function.
- (iii) The electric process is represented by the random noise added to the signal.

The main part of the chapter was devoted to the problem of concentration estimation. The approach to this problem is described in Fig.1. The basic idea that underlies the technique is simulating a noiseless spectrum with a given initial concentration $C = (c_1, \dots, c_n)$. The actual spectrum and the simulated one are compared by some error function. Upon the comparison result a correction to the initial value C is made. This process is repeated until no significant correction term is obtained.

It has been shown that under our assumption (uncorrelated noise, constant power additive gaussian noise) the same error function (rms Error) is achieved by applying the Maximum Likelihood Estimator.

Since the problem is nonlinear in $(c_1, \dots, c_n)$, we had to show that a 'good' estimate may be obtained no matter what initial value C we start with. In the developing the equation for the correction ΔC , it has been shown that the same relation between the rms error and signal exists in the linear estimation problem. This led us to define the modified signal and the modified signal power. In the case of high resolution it has the form :

$$\text{Modified signal sample} = t \exp[-at]$$

$$\text{Modified signal power sample} = t^2 \exp[-2at]$$

where t is a function of the cell length and the wavelength.

It has been shown that by using the definition above for the signal power, the expression of the variance has the same form as in the linear case.

One of the immediate conclusion was that there is a cell length that yields maximum modified signal power and therefore minimum estimation error. Our main conclusion is that a search technique may be applied in order to estimate the concentration(s) in a mixture. This is the subject of the next chapter.

5. Simulation and Results

5.1

Introduction

This chapter gives an account of the computer simulation of the signal processing techniques developed here. It was carried out on a (CDC 7600) computer with data obtained from instrumental tests. The aims of the simulation were :

- (1) To establish, using the theoretical results derived in the previous chapters, the techniques for the estimation of the concentration of a compound in a mixture of gases, to test their accuracy and other features.
- (2) To test and verify the theoretical results derived in the previous chapters.

Fig.1 describes the general model of simulation that was followed.

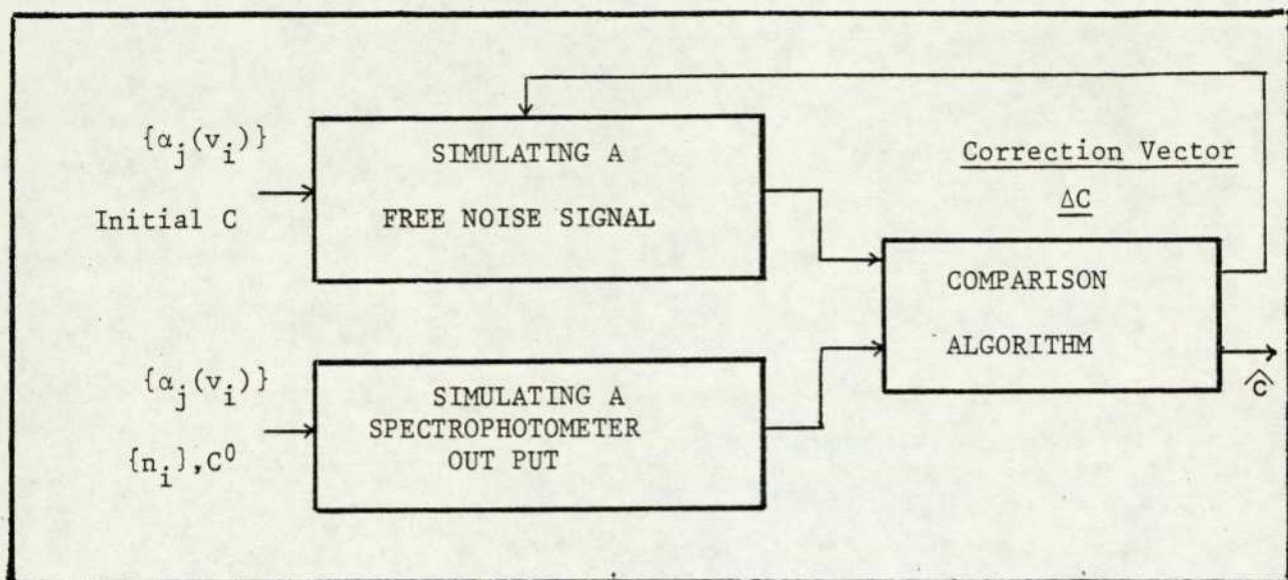


Fig. 1

A general simulation model

The simulation was based on two fundamental steps :

- (1) Simulation of a signal with a given resolution noise level and concentration(s). This is the signal that is assumed to be the spectrum with the unknown concentration(s).
- (2) Application of a search algorithm aimed to find the concentration(s) in the above signal. The search is terminated when no significant change occurs either in the estimation or in the error function.

The simulations have been carried out on a range of S/N ratios, signal resolutions and concentrations. Part of the work involved recording a bank of signals and creating a package of programs to enable access and manipulate the recorded signals.

The chapter is divided into three parts :

- (1) The theoretical background for the two algorithms is stated. The algorithms are described in block diagrams.
- (2) Results of several tests are given in a graph form, and interpretation of the results is discussed.
- (3) A brief description of the work involved in creating the bank of signals and the program package is given.

5.2

The estimation algorithms

The two estimation algorithms that have been developed are based on the Maximum Likelihood estimator (MLE).

(1) The spectrum with the unknown concentration(s) is given before the search for the concentration(s) starts using the MLE. In this case, the search technique is such that the concentration vector, C , which is found minimizes the error function (See Eq.1.). It was assumed that the signal measured is an output of a high resolution spectrophotometer.

(2) A sequential technique is based on an output of high resolution spectrophotometer. In this case, by applying a log transformation to the signal, the estimation equation is approximated to a linear equation. Consequently, the estimation of the concentration may be calculated immediately as the signal samples enter the estimation processor and without any iteration. The result is that the amount of storage required for running the estimation algorithm is much less than the amount required in the former algorithm. Only a few accumulative

vectors, with dimension equal to the number of compounds, are stored, whereas in the previous case the whole signal record has to be stored.

Below, in a brief description, the estimation problem and assumptions that underlie the estimation process are repeated.

In measuring the concentration of a single compound in a mixture difficulties arise, firstly, because of an additive noise that is generated at the detector that converts the infra red beam, at each wave length, to an electric signal, Secondly, there is interference of overlapping spectra of other compounds mixed with the desired spectrum, and thirdly, the shape of the signal spectrum is a distorted image of the true spectrum. The basic distortion is due to the finite beam width (in wave length) that represents the absorption at a wave length. In other words, the given absorption at a wavelength is always the result of some averaging (convolution) over local range (slit width). See Fig.2,Ch.4 (block diagram of the spectrophotometer).

The following assumptions were made :

1. The spectrum signal from the spectrophotometer is given as a discrete vector Y :

$$Y = (y_1, y_2, \dots, y_M)$$

where y_i the vector component is the measured spectrum (the output spectrum) at a wave number v_i .

(2) The measured spectrum is described mathematically :

$$y_i = f(v_i, C^0) + n_i$$

$$C^0 = (c_1^0, c_2^0, \dots, c_n^0)$$

That is the output consists of the sum of two components, a random noise sample n_i at wave number v_i , and the absorption signal given by :

$$x(v_i, C) = \exp(-\sum_j \alpha_j c_j \alpha_j(v_i))$$

$$f(v_i, C) = \sum_{-N}^N b_n x(v_i + nV)$$

$$i = 1, \dots, M$$

Since a high resolution spectrophotometer is assumed, we have :

$$f(v_i, C) = x(v_i, C)$$

c_j is the concentration of the compound j , $\alpha_j(v_i)$ is the absorption coefficient at a wave number (v_i) .

(3) The noise components $\{n_i\}$ are uncorrelated with a constant power :

$$\overline{n_i} = 0$$

$$\overline{n_i n_j} = \begin{cases} 0 & i \neq j \\ \sigma^2 = N_0 & i = j \end{cases}$$

5.2.1 The maximum likelihood estimation algorithm

Rewriting (Eq.11, Ch.4) :

$$E^2(C) = \sum_i (f(v_i, C) - y_i)^2 \xrightarrow{C} \text{Min.}$$

In Sec.4.3 it has been shown that a search method may be applied to the error function above in order to find such a C that yields a minimum to the error function above. A set of equations that gives the correction ΔC was developed (see Ch.4 Eq.22) :

$$\begin{bmatrix} \Delta C_1 \\ \vdots \\ \Delta C_n \end{bmatrix} = \begin{bmatrix} R_{11} & , R_{1n} \\ \vdots & \vdots \\ R_{n1} & , R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix} \quad (1.)$$

where :

$$R_{kj} = \sum_i \alpha_k(v_i) \alpha_j(v_i) f^2(v_i, C)$$

$$R_{yj} = - \sum_i (y_i - f(v_i, C)) \alpha_j(v_i) f(v_i, C)$$

In the case of 2-concentrations in the mixture, we have :

$$\begin{bmatrix} \Delta c_1 \\ \Delta c_2 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ R_{y2} \end{bmatrix} \quad (1a.)$$

The error in the estimation is given by :

$$\begin{bmatrix} \overline{\Delta c_1^2} \\ \overline{\Delta c_2^2} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}^{-1} N_0$$

Consequently :

$$\overline{\Delta c_1^2} = \frac{N_0 R_{22}}{R_{11} R_{22} - R_{12}^2}$$

$$\overline{\Delta c_1^2} = \frac{N_0}{R_{11} (1 - \rho^2)}$$

$$\overline{\Delta c_2^2} = \frac{N_0}{R_{22} (1 - \rho^2)}$$

As it was shown before, R_{11}, R_{22} , play the role of the signal power, and ρ plays the role of the normalized correlation factor (see Ch.4, Ex.3).

A new C' is calculated by :

$$c'_1 = c_1 + \Delta c_1$$

$$c'_2 = c_2 + \Delta c_2$$

The new vector C' is substituted in the above equation. the process is repeated (see Fig.1) till the improvement is lower than some given tolerance. The result are stored, and another trial is repeated, with a new vector Y , where Y comprises the same vector $[f(v_1, C^0), f(v_2, C^0), \dots, f(v_M, C^0)]$ but with fresh noise components.

The rms error of ten trials is calculated and is compared with the theoretical error given above. Figure 2 describes, in a block diagram, this process. Choosing sufficiently large number of trials is important in order to guarantee that the mean value of the rms error calculated on the whole ensemble (10 in our case) is close enough to the theoretical value (which is the limit of the mean over an infinite number of trials, see App.6).

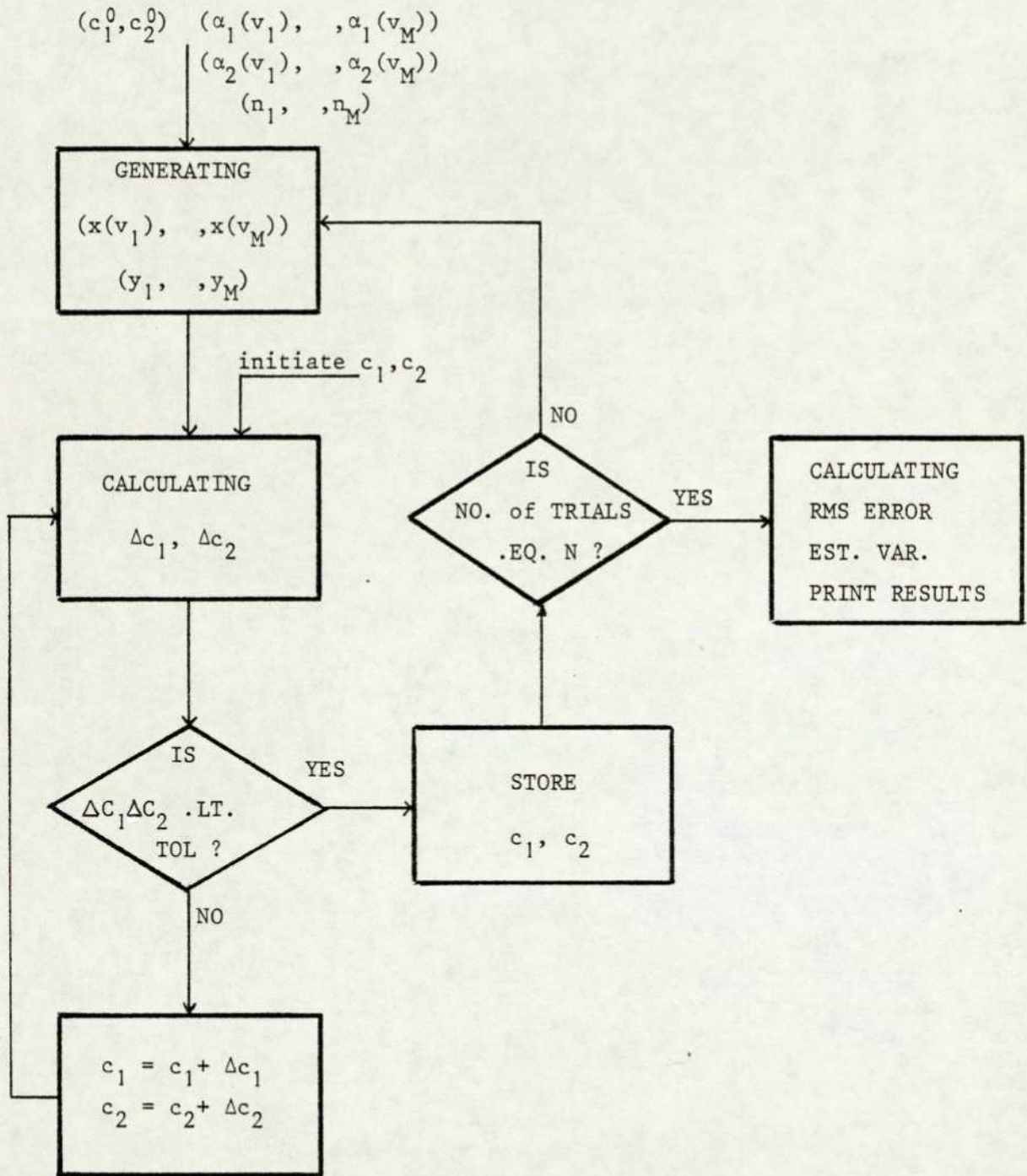


Fig. 2 Block diagram of the simulated MLE algorithm

5.2.2 The sequential estimation algorithm
(for high resolution signal)

In case of high resolution spectrophotometer the set of equations that describe the output are given by (see Sec.5.2) :

$$y_i = f(v_i, C) + n_i \quad (2.)$$

$$f(v_i, C) = b_0 x(v_i) = \exp(- \sum_j c_j \alpha_j(v_i)) \quad (2a.)$$

By taking the logarithm of each sample the set of equations is reduced to a set of linear equations :

$$\log(y_i) = \log\{f(v_i, C) + n_i\}$$

$$\log\{f(v_i, C) + n_i\} = \log\{f(v_i, C)\} + \log\{1 + \frac{n_i}{f(v_i, C)}\}$$

Substituting (2a.), we have :

$$\log(y_i) = - \sum_j c_j \alpha_j(v_i) + \log(1 + \frac{n_i}{f(v_i, C)})$$

Assume that the S/N ratio is high enough so that :

$$\frac{n_i}{f(v_i, C)} \approx \frac{n_i}{f(v_i, C) + n_i} = \frac{n_i}{y_i}$$

Further, $\log(1 + \frac{n}{y})$ can be expressed as a Taylor series :

$$\log\left(1 + \frac{n_i}{y_i}\right) = -\frac{n_i}{y_i} + \frac{1}{2}\left(\frac{n_i}{y_i}\right)^2 + \dots$$

Neglecting $(n_i/y_i)^2$, higher order terms and substituting to (2a.), we have :

$$\log(y_i) \approx -\sum_j c_j \alpha_j(v_i) - \frac{n_i}{y_i} \quad (3.)$$

$$i = 1, \dots, M$$

$$y_i \sum_j c_j \alpha_j(v_i) + y_i \log(y_i) = -n_i \quad (3a.)$$

The above set of equations is linear in $(c_1, \dots, c_n)$.

The likelihood function of the above case is given by (See Ch.4, Eq.10) :

$$p(y_1, \dots, y_M/c_1, \dots, c_n) = p(n_1, \dots, n_M/c_1, \dots, c_n)$$

$(n_1, \dots, n_M)$ are a vector of a gaussian random variables. Substituting Eq.(3a), we have :

$$\begin{aligned} p(y_1, \dots, y_M/c_1, \dots, c_n) &= \\ &= (1/2\pi N_0)^{M/2} \exp\{-1/2N_0 \sum_i y_i^2 \{\log(y_i) + \sum_j c_j \alpha_j(v_i)\}^2\} \end{aligned}$$

Ch.5

The maximum likelihood estimate is given by C that maximizes the expression above, or :

$$\sum_i y_i \{ \log(y_i) + \sum_j c_j \alpha_j(v_i) \}^2 \xrightarrow{C} \text{Min.}$$

$$\frac{\partial}{\partial c_k} \{ \sum_i y_i^2 \{ \log(y_i) + \sum_j c_j \alpha_j(v_i) \}^2 \} = 0$$

$$k = 1, \dots, n$$

which also may be written in a vector operator form :

$$\Delta E^2 = \nabla(e'e) = 0$$

(4.)

$$2Je = 0$$

where the vector e is given by :

$$e' = ((y_1 \log(y_1) + y_1 \sum_j c_j \alpha_j(v_1)), \dots, (y_M \log(y_M) + y_M \sum_j c_j \alpha_j(v_M)))$$

The correction vector ΔC is given by (see Ch.4, Eq.19,19a) :

$$Je = JJ' \Delta C = 0 \quad (5.)$$

$$J = \nabla' e =$$

$$= \nabla' ((y_1 \log(y_1) + y_1 \sum_j c_j \alpha_j(v_1)), (y_M \log(y_M) + y_M \sum_j c_j \alpha_j(v_M)))$$

$$J = \begin{bmatrix} y_1 \alpha_1(v_1), & y_M \alpha_1(v_M) \\ \vdots & \vdots \\ y_1 \alpha_n(v_1), & y_M \alpha_n(v_M) \end{bmatrix} \quad (6.)$$

Since this estimation problem is a linear in $[c_1, \dots, c_n]$ (see Eq.3a), an explicit expression can be found for $[\hat{c}_1, \dots, \hat{c}_n]$. Eq. (4.) is used to find the solution for $(\hat{c}_1, \dots, \hat{c}_n)$. The vector e can be written :

$$e' = e'_1 + e'_2$$

$$e'_1 = (y_1 \log(y_1), \dots, y_M \log(y_M))$$

$$e'_2 = (y_1 \sum_j c_j \alpha_j(v_1), \dots, y_M \sum_j c_j \alpha_j(v_M))$$

$$Je = J(e_1 + e_2) = Je_1 + Je_2$$

developing the expression Je_2 further more, we have :

$$J \begin{bmatrix} y_1 \sum_j c_j \alpha_j(v_1) \\ \vdots \\ y_M \sum_j c_j \alpha_j(v_M) \end{bmatrix} = J \begin{bmatrix} y_1 \alpha_1(v_1), & y_1 \alpha_n(v_1) \\ \vdots & \vdots \\ y_M \alpha_1(v_M), & y_M \alpha_n(v_M) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Comparing the two dimension matrix above to Eq.(6), we recognize it as J' . We therefore have :

$$J e + J J' \hat{C} = 0$$

Comparing the equation above to the formula of ΔC in the nonlinear case (see Eq.5), we notice that they both have a similar form. The explicit expression for $\hat{C}$ is given by :

$$\hat{C} = -(J J')^{-1} J e_1 \quad (6a.)$$

As in the nonlinear case we define :

$$(J J')^{-1} J e_1 = \begin{bmatrix} R_{11} & R_{1n} \\ \vdots & \vdots \\ R_{n1} & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix} \quad (7.)$$

$$R_{ij} = \sum_k y_k \alpha_i(v_k) \alpha_j(v_k)$$

$$R_{yj} = \sum_k y_k \alpha_j(v_k) \log(y_k)$$

In the same way the variance in each estimate is given by (see Ch.4, Eq.27) :

$$\begin{bmatrix} \overline{\Delta c^2_1} \\ \overline{\Delta c^2_n} \end{bmatrix} = \begin{bmatrix} \overline{R_{11}}, & \overline{R_{1n}} \\ \overline{R_{n1}}, & \overline{R_{nn}} \end{bmatrix}^{-1} N_0 \quad (8.)$$

The following example verifies the above derivation.

Example 1

Single concentration

Using Eq.6 and 7, we find :

$$\hat{c} = \frac{\sum_k y_k \alpha^2(v_k) \log(y_k)}{\sum_k y_k^2 \alpha^2(v_k)}$$

The above result can be tested by assuming a noiseless case :

$$y_i = \exp(-c \alpha(v_i))$$

we have :

$$\hat{c} = \frac{c^0 \sum_k y_k^2 \alpha^2(v_k)}{\sum_k y_k^2 \alpha^2(v_k)} = c^0$$

which verifies the above formula. The estimate variance is given by (see Eq.8) :

$$\overline{\Delta c^2} = \frac{N_0}{\sum_k y_k^2 \alpha^2(v_k)} \quad (9.)$$

This result may be derived directly by substituting into the equation of $\hat{c}$:

$$\begin{aligned}\log(y_i) &= \log\{\exp(-c^0 \alpha(v_i)) + n_i\} \\ &= c^0 \alpha(v_i) - \frac{n_i}{y_i} + \frac{1}{2} \frac{n_i^2}{y_i^2} - \dots\end{aligned}$$

we have :

$$\begin{aligned}c &= \frac{c^0 \sum_k y_k \alpha^2(v_k) + \sum_k y_k^2 \alpha(v_k) n_k / y_k}{\sum_k y_k^2 \alpha^2(v_k)} \\ c &= c^0 + \frac{\sum_k y_k^2 \alpha(v_k) n_k}{\sum_k y_k^2 \alpha^2(v_k) y_k} \\ (c - c^0) &= \frac{\sum_k y_k \alpha(v_k) n_k}{\sum_k y_k^2 \alpha^2(v_k)}\end{aligned}$$

we note that the mean value of $(c - c^0)$ is not zero but of order of :

$$B = -\frac{1}{2} \frac{\sum_i y_i^2 \alpha(v_i) (n_i / y_i)^2}{\sum_i y_i^2 \alpha^2(v_i)} = -\frac{N_0}{2} \frac{\sum_i \alpha(v_i)}{\sum_i y_i^2 \alpha^2(v_i)}$$

which is the mean of the second term which was neglected in the expansion of $\log(y_i)$ above. That is the estimator has a non zero bias with an order of magnitude of the S/N ratio and therefore may be

neglected in reality.

The estimator variance is given by :

$$\begin{aligned} (\hat{c} - c^0)^2 &\approx \frac{\left\{ \sum_i y_i \alpha(v_i) n_i \right\}^2}{\left\{ \sum_i y_i^2 \alpha^2(v_i) \right\}^2} \\ \overline{(\hat{c} - c^0)^2} &\cong \frac{\sum_i y_i y_j \alpha(v_i) \alpha(v_j) \overline{n_i n_j}}{\left\{ \sum_i y_i^2 \alpha^2(v_i) \right\}^2} = \frac{\sum_i y_i^2 \alpha^2(v_i) N_0}{\left\{ \sum_i y_i^2 \alpha^2(v_i) \right\}^2} \\ \overline{(\hat{c} - c^0)^2} &\cong \frac{N_0}{\sum_i y_i^2 \alpha^2(v_i)} \end{aligned}$$

which again verifies the general result (see Eq.9).

A sequential approach may be adopted easily. the formula for $[\hat{c}_1, \hat{c}_n]$ is given by (see Eq.6a,7) :

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_n \end{bmatrix} = \begin{bmatrix} R_{11}, & R_{1n} \\ \vdots & \vdots \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ R_{yn} \end{bmatrix}$$

$$R_{ij} = \sum_k y_k^2 \alpha_i(v_k) \alpha_j(v_k)$$

$$R_{yj} = \sum_k y_k^2 \alpha_j(v_k) \log(y_k)$$

define :

$$ACR_{ijm} = \sum_{k=1}^m y_k^2 \alpha_i(v_k) \alpha_j(v_k)$$

$$ACR_{yjm} = \sum_{k=1}^m y_k^2 \alpha_{i(v_k)} \log(y_k)$$

we have :

$$ACR_{ijm} = ACR_{ijm-1} + y_m^2 \alpha_{i(v_m)} \alpha_{j(v_m)}$$

$$ACR_{yjm} = ACR_{yjm-1} + y_m^2 \alpha_{i(v_m)} \log(y_m)$$

Thus in applying the above equations, the signal vector $[y_1, \dots, y_M]$ is not stored, since every component y_i is immediately processed. In case of two compounds (two concentrations) there are only five variables to be stored : ACR_{11} , ACR_{12} , ACR_{22} , ACR_{y1} , and ACR_{y2} . During the estimation process the estimates c_1 and c_2 may be calculated at any instance by applying Eq.6a :

$$\begin{bmatrix} \hat{c}_{1m} \\ \hat{c}_{2m} \end{bmatrix} = \begin{bmatrix} ACR_{11m} & ACR_{12m} \\ ACR_{12m} & ACR_{22m} \end{bmatrix}^{-1} \begin{bmatrix} ACR_{y1m} \\ ACR_{y2m} \end{bmatrix} \quad (9a.)$$

Fig. 3 describes this algorithm as it was simulated on the computer.

In the following subsection the results of several simulations are presented and some of the conclusions are discussed.

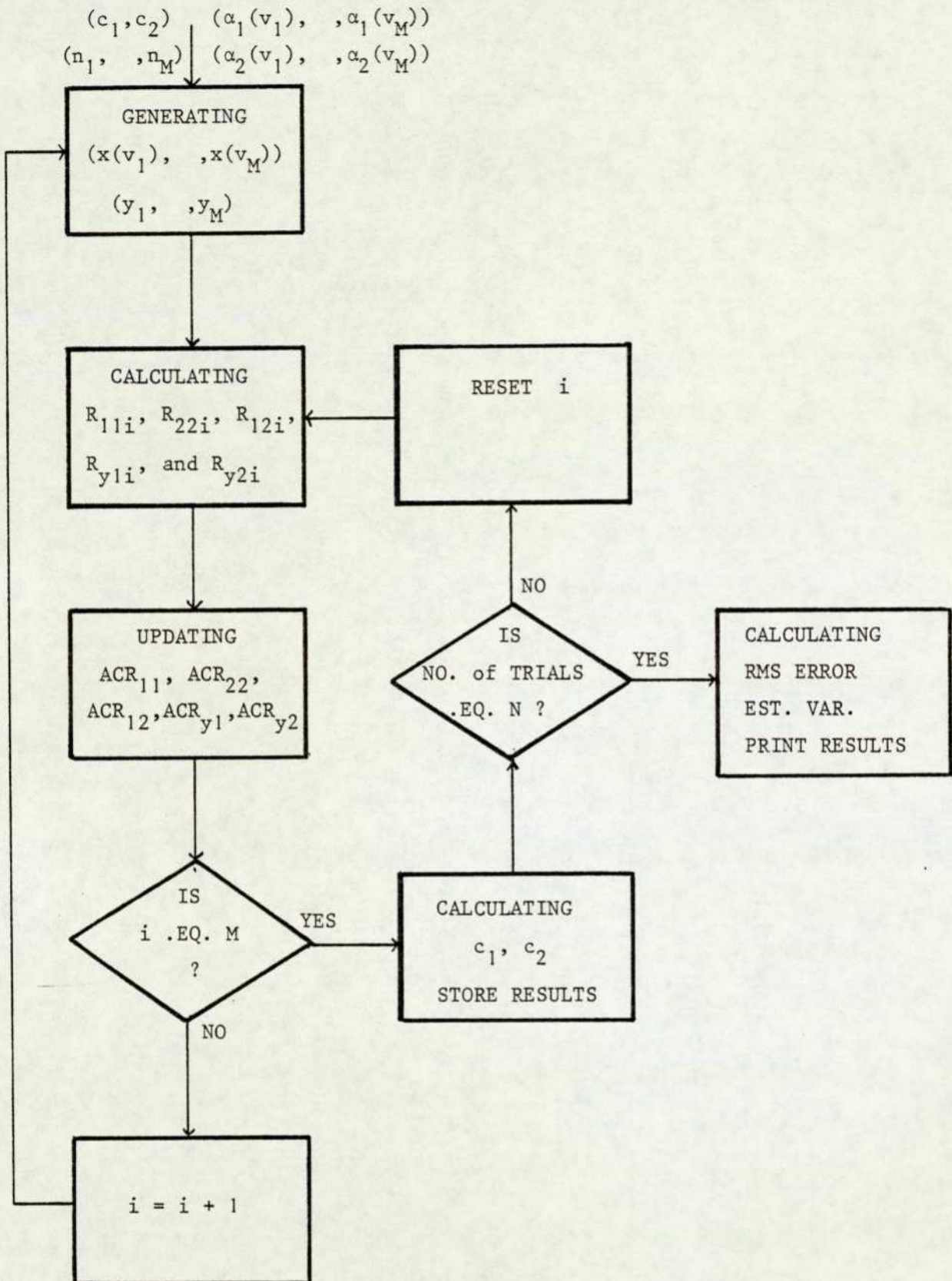


Fig. 3 Block diagram of the simulated SE algorithm

5.3 Simulation and results

5.3.1 Introduction

Simulation have been carried out with both techniques, the MLE (Maximum Likelihood Estimator) and the SE (Sequential estimator).

The tests were set to find out :

- 1) A validation of the theoretical assumptions and the general expressions that were derived, e.g. the estimation variance.
- 2) A validation of the ability of the algorithms developed in estimating one compound in the presence of another. Also, the accuracy obtainable in using the estimation algorithms, developed here, in different levels of noise.
- 3) Demonstrating the application of the general expression of the estimation variance in predicting the behaviour of the error in the estimation

Estimation of one gas (CH₄) in the presence of the other (H₂S) has been carried out on signals with S/N of 40-70db and concentrations of 2-200 ppm. The concentration of the interfering gas (H₂S) was 1000 ppm.

Each estimation was repeated 10 times, the rms error was calculated and plotted as a function of the concentration.

In another set of simulations the theoretical error (estimator variance) was compared to the rms error above. The close

agreement of the two enabled us to use the theoretical expression in predicting the error as a function of the cell length.

The reference spectra used were transmittance spectra of :

- (i) A spectrum of CH₄ (Methane) : $[f_1(v_1), \dots, f_1(v_M)]$,
- (ii) A spectrum of H₂S : $[f_2(v_1), \dots, f_2(v_M)]$.

Both in the wave number range of 1100-1400 1/cm. the spectrophotometer cell length was 10 cm. A brief description of the work that was involved in recording, digitizing and storing the signals above is given in Sec.5.4.

The digitized spectra were, first, normalized to cover the range of :

$$0. < f_j(v_i) < 1.$$

$$i = 1, \dots, M$$

$$j = 1, 2$$

The absorption coefficients :

$$(\alpha_1(v_1), \dots, \alpha_1(v_M))$$

$$(\alpha_2(v_1), \dots, \alpha_2(v_M))$$

were found by (see Eq.1,1a,Ch.1) :

$$\alpha_j(v_i) = -\log\{f_j(v_i)\}$$

$$\alpha_j(v_i) = \alpha_j^0(v_i)L$$

$$i = 1, \dots, M$$

$$j = 1, 2$$

Varying the cell length was done by varying L in units of 10cm.

The transmittance spectrum at wave number v_i and a mixture with concentrations c_1^0 and c_2^0 was simulated by (see Eq. 14a., Ch4):

$$f(v_i, c_1^0, c_2^0) = f_1(v_i) f_2(v_i)$$

$$f(v_i, c_1^0, c_2^0) = \exp(-L(c_1^0 \alpha_1(v_i) + c_2^0 \alpha_2(v_i)))$$

$$i = 1, \dots, M$$

The spectrophotometer output was simulated by (assuming high resolution spectrum) :

$$y_i = f(v_i, c_1^0, c_2^0) + n_i$$

$$i = 1, \dots, M$$

where n_i was an output of a gaussian random number generator, with a variance of :

$$N_0 = \sigma^2 = 10^{-(1/10(S/N))}$$

5.3.2 Accuracy of the estimation as a function of the S/N ratio

These simulations validate the correctness of the two algorithms and of some general expressions that describe the rms error (estimator variance).

Given a simulated output of a spectrophotometer :

$$(Y_1, \dots, Y_M)$$

a search was initiated (by applying the MLE algorithm) which terminated with the results $\hat{c}_1, \hat{c}_2$. The trial was repeated 10 times, each time with a new vector $(Y_1, \dots, Y_M)$ where c_1^0, c_2^0 , the simulated concentration, kept constant and the noise components were replaced. The rms error is given by :

$$\text{RMS Error of } c_i = \left\{ (1/9) \sum_j^{10} (\hat{c}_i^j - \overline{\hat{c}_i})^2 \right\}$$

$$\overline{\hat{c}_i} = 1/10 \sum_i^{10} (\hat{c}_i^j)$$

The rms error was compared to the estimator variance, which is given by (see Eq.1b and 10) :

$$\overline{\Delta c_j^2} = \frac{N_0}{S_j(1 - \rho^2)} = \frac{10^{-1/10(S/N)}}{S_j(1 - \rho^2)}$$

where :

$$i = 1, \dots, M$$

$$j = 1, 2$$

$$S_j = \sum_i y_i \alpha_j(v_i)$$

$$\rho^2 = \frac{\sum_i y_i^2 \alpha_1(v_i) \alpha_2(v_i)}{S_1 S_2}$$

Fig.4 describes the rms error in simulation the MLE in a different levels of S/N ratio (see also Fig.2) and in presence of 1000 PPM of H₂S. Table 1 gives some of the error results :

2. PPM	S/N [db]	40.	50.	60.	70.
	Error [%]	60.	18.	5.	1.8

20. PPM	S/N [db]	40.	50.	60.	70.
	Error [%]	4.	1.2	.3	.08

Table 1 Error in estimation 2. and 20. PPM of CH₄ in the
presence of 1000. PPM of H₂S

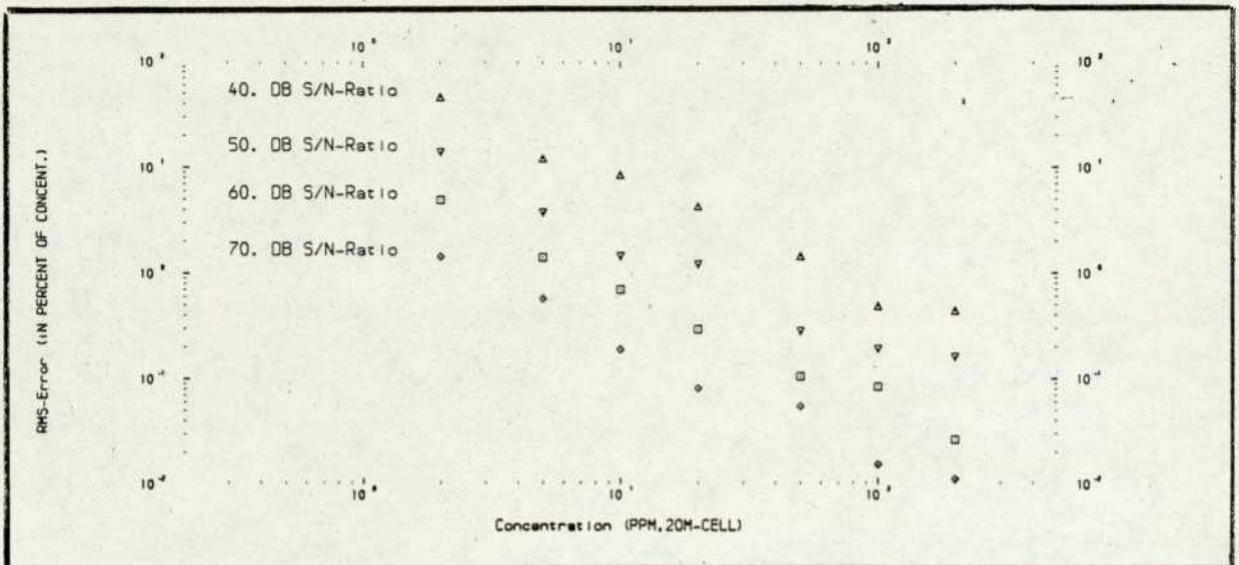


Fig. 4 RMS error in estimating the concentration of CH₄ in the presence of H₂S (1000.PPM) using the MLE technique

Fig. 5 show the rms error in applying the SE algorithm to the same signal. Table 2 compares the rms error, in the estimation, between the two algorithms.

S/N [db]	40.	50.	60.	2. PPM
MLE	60.	18.	5.	
SE	50.	15.	3.	
MLE	4.	1.2	.3	20 PPM
SE	3.	.8	.22	

Table 2 Comparison of the error, in estimation 2. and 20.
PPM of H4, between the MLE and SE algorithms

The above results show that the SE (Sequential Estimator) has slightly better performance than the MLE (Maximum Likelihood Estimator).

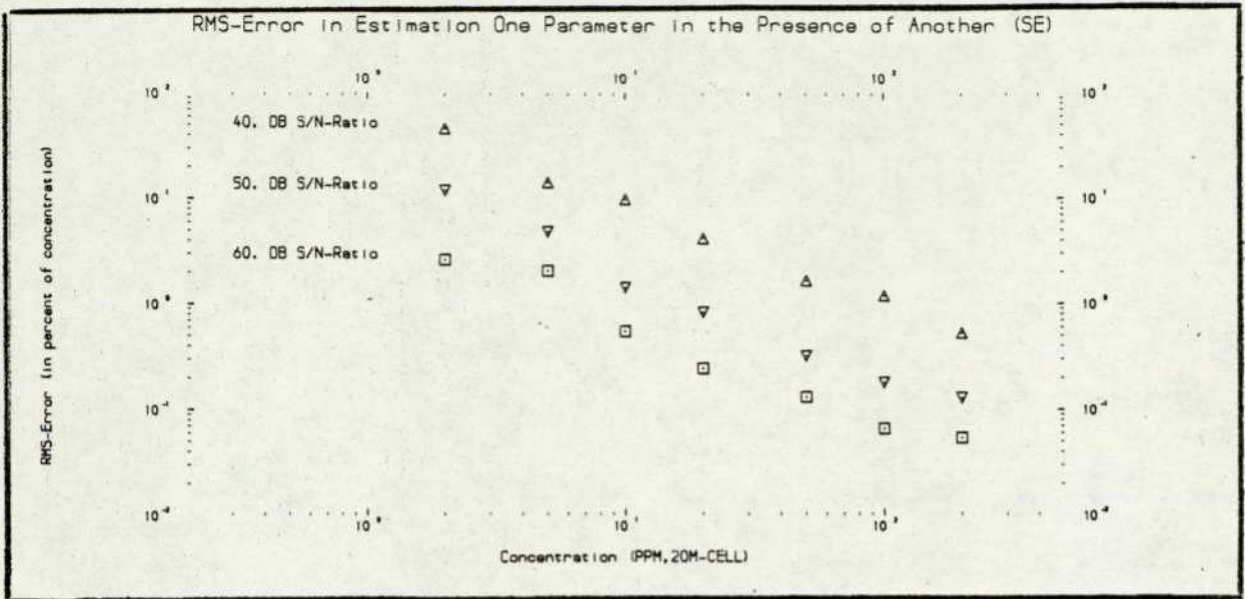


Fig. 5 RMS error in estimating the concentration of CH₄ in the presence of H₂S (1000.PPM) using the SE technique

Fig.6 and Fig.7 compare the estimator variance with the rms error. the rms error is the error over 10 trials (see Eq.10a), where the expression for the variance (see Eq.11) is the result of theoretical considerations (see Sec.4.4 Eq.36). The close vicinity of the two justify our theoretical assumptions. Consequently, our definition of the modified signal power and modified correlation factor (see Sec.4.4.Eq.36a,37) is proved to be correct in the sense that this two terms play the same role in the nonlinear estimation problems as the regular signal power and the correlation factor play in the linear estimation problems. Some application of the theoretical expressions of the error is shown in the next subsection.

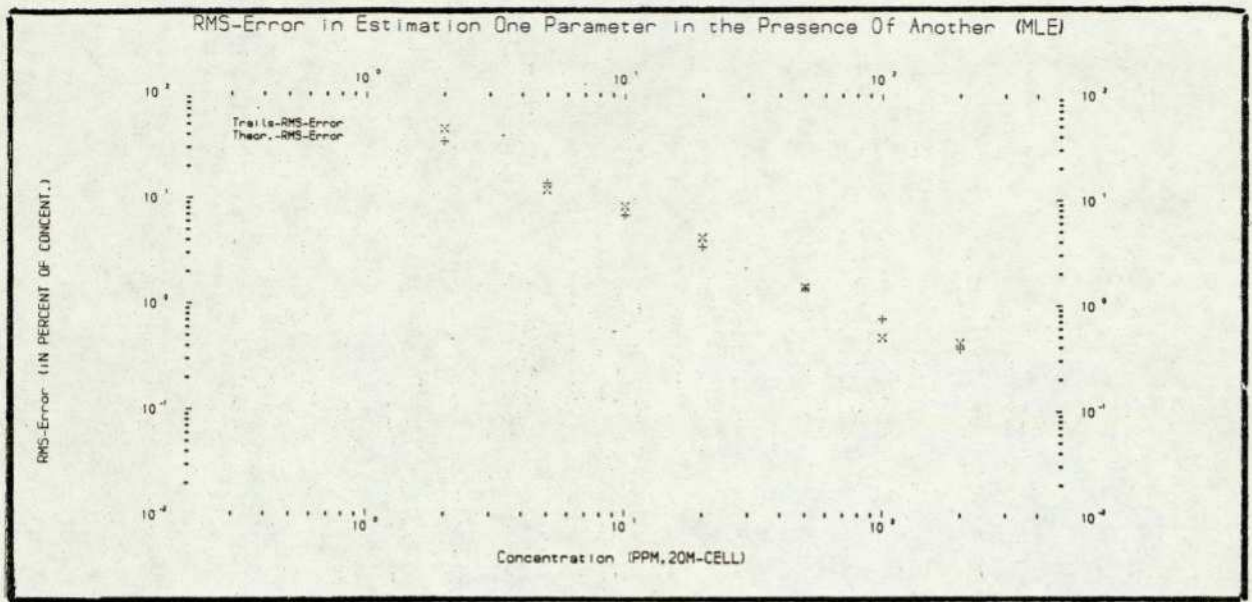


Fig. 6 Comparison of the rms error and the estimation variance in applying the MLE

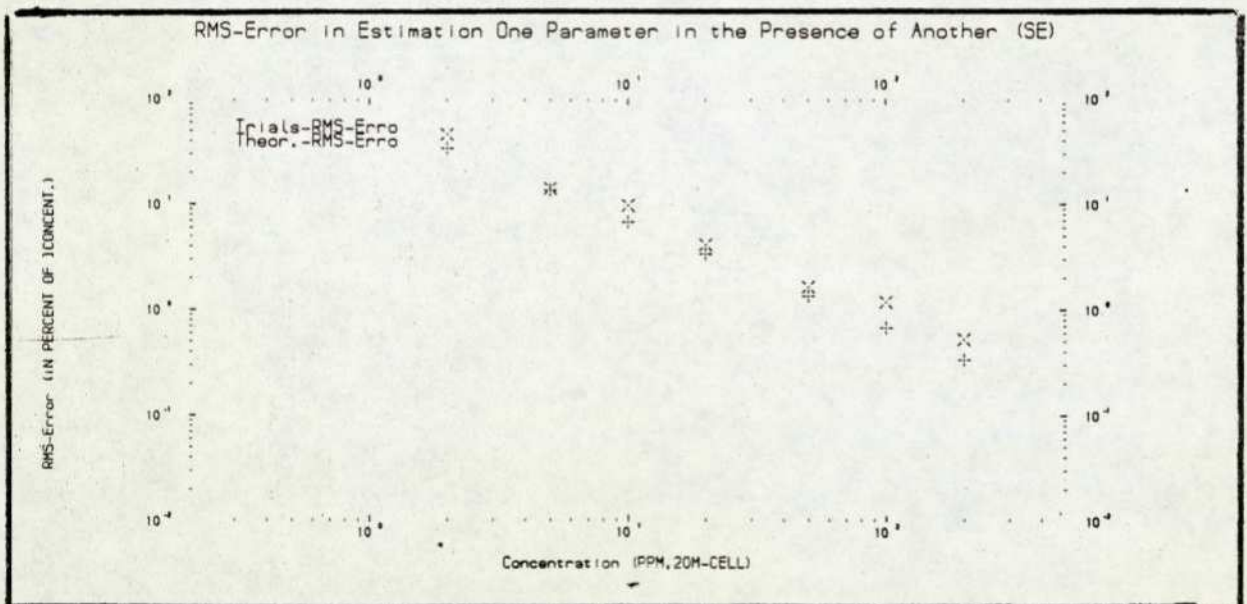


Fig. 7 Comparison of the rms error and the estimation variance in applying the SE

5.3.3 The estimator variance as a function of the cell length

As the cell length becomes larger the spectrum absorption lines (in a gas mixture with a constant pressure and concentrations) become stronger. This is obvious, since that means larger number of molecules that absorb the ir beam energy. Thus it seems that description of the estimation error as a function of number of molecules (PPM-m units), in the absorption cell, will reveal some more information on the effect of the cell length on the accuracy of the estimation.

Intuitively, one expects a better estimation the stronger the absorption lines are. Does it mean that the longer the cell length the better the estimation ? Surely, for a very long cell length this can not be true, since most of ir energy will be absorbed by the molecules in the cell with the result of a very weak output signal from the detector (see Ch.1 Fig.) and consequently poor S/N ratio in the output and large error in the estimation.

Fig.8 reveals that indeed this is the case. The estimation variance described as a function of the number of the molecules (PPM m) in the cell has a minimum. Increasing the cell length (and the number of the molecules) beyond that point will result with larger error in the estimation.

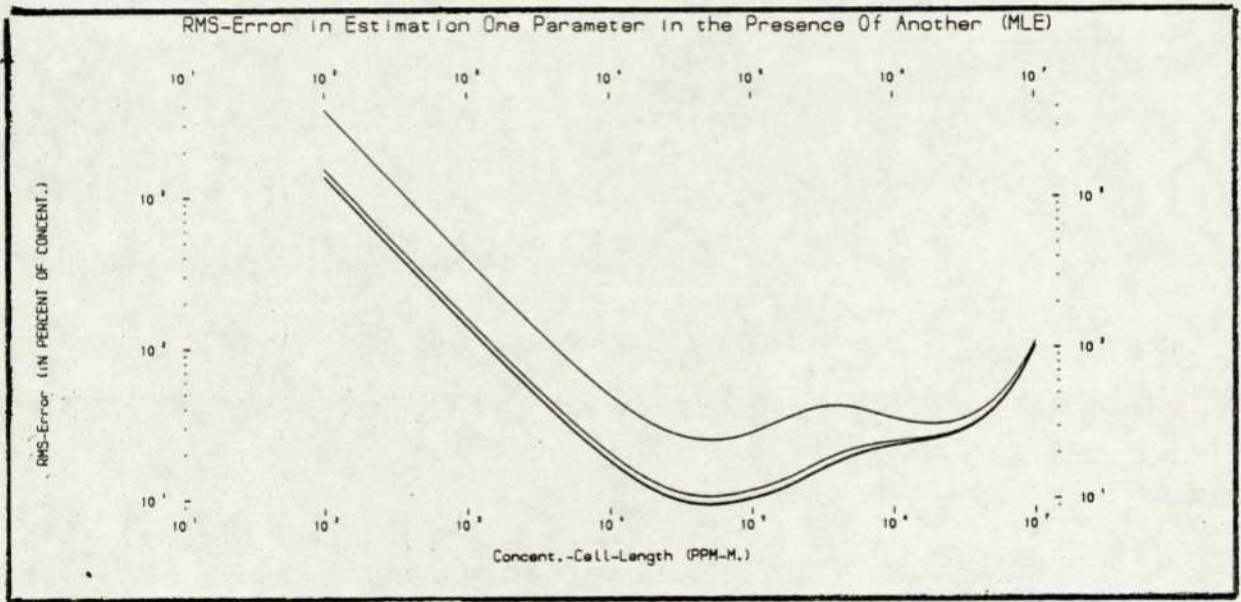


Fig. 8 The error graph (estimation of CH₄ in the presence of H₂S) indicates the location of the optimum cell length

Inspecting the estimator variance expression (see Eq.11) :

$$\overline{\Delta c_j^2} = \frac{N_0}{S_j (1 - \rho^2)}$$

$$S_j = \sum_i y_i^2 \alpha_j^2(v_i)$$

$$\rho^2 = \frac{\sum y_i a_1(v_i) a_2(v_i)}{S_1 S_2}$$

$$j = 1, 2$$

Where :

$$y_i = \exp(-L(c_1 \alpha_1^0(v_i) + c_2 \alpha_2^0(v_i))) + n_i$$

$$a_j(v_i) = L a_j^0(v_i)$$

and L is the cell length.

we notice that the modified signal power as a function of the cell length has a maximum as expected. This is so since each term, in S_j , as a function of the cell length- L , has the form :

$$L A_i \exp[-L B_i]$$

It is well known that expression like the above has a maximum at :

$$L_{\text{opt.}}(i) = 1/B_i \quad i = 1, \dots, M$$

where B_i in our case is the absorption coefficient per a unit cell length. For cell length- L large enough and small enough we have :

$$L_1 \gg 1/B_i$$

$$L_2 \ll 1/B_i$$

$$i = 1, \dots, M$$

respectively.

The optimum cell length is therefore :

$$L_2 \leq L_{opt.} \leq L_1$$

This subsection establishes the theoretical assumptions that were made in developing the two estimation algorithms, validates the general expression for the estimation rms error and opens the way for the practical use of the above algorithms. The next subsection describes briefly the work involved in recording digitizing and storing the bank of spectra, and the package of programs that was created for accessing and manipulating the signals.

5.4 Creating the bank of signals and the supporting package of programs

The bank of signals and the package of programs were created to allow off-line simulations, on a large main frame computer (CDC 7600), with a variety of cases like different S/N ratio's, resolutions and concentrations.

A Pye Unicam spectrophotometer (Model SP 4000) was used, and its signal output was recorded. The following spectra were recorded and processed :

- (a) The energy profile of the spectrophotometer in the range of 400-4000 $1/\text{cm}$,
- (b) Noise output of the spectrophotometer at different slit widths ,
- (c) Transmittance spectrum of 100% CH_4 , range 1100-1500 $1/\text{cm}$ cell,
- (d) Transmittance spectrum of 100% CH_4 , range 2600-3300 $1/\text{cm}$, 10cm cell,
- (e) Transmittance spectrum of H_2S , range of 1100-1500 $1/\text{cm}$, 10cm cell,
- (f) Transmittance spectrum of a mixture of H_2S and CH_4 (10% and 90% respectively) 10cm cell.

Creating the bank of signals has been carried out in three stages :

- 1) The spectrum of several gases was recorded on one channel of analogue tape recorder. On a second parallel channel a start pulse, initiated by the spectrophotometer, was recorded and thus enabled us to identify the beginning of each recorded. The recording process took place at the SIRA Institute (see Fig.9)

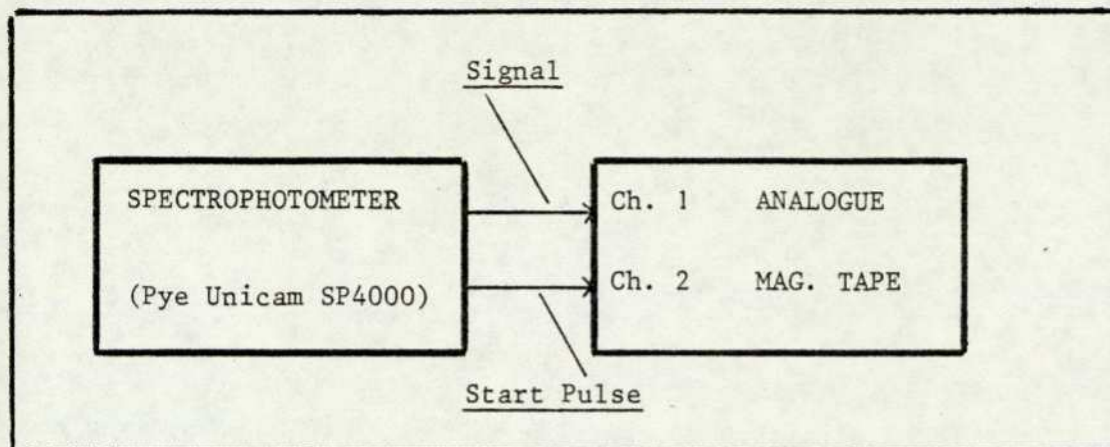


Fig. 9. Recording the spectrum on an analogue tape recorder

(7 "/sec, B.W.: DC-1.25kHz)

2) In the second stage the tape was replayed, the signal output was sampled and stored on a digital tape. In this process Channel 2 was sampled (the Start Pulse Channel), until a start pulse was detected by the computer. From that instant the A/D started to sample Channel 1 (the Signal Channel) and the samples were stored in the computer memory. This process was stopped after 17,408 words (7 minutes, 40 samples/sec) and the file was transferred, in records of 256 words (512 8-bit words), to a digital tape. This stage took place at the Instrument Systems Centre of the Department of Systems Science (see Fig's 10 and 11).

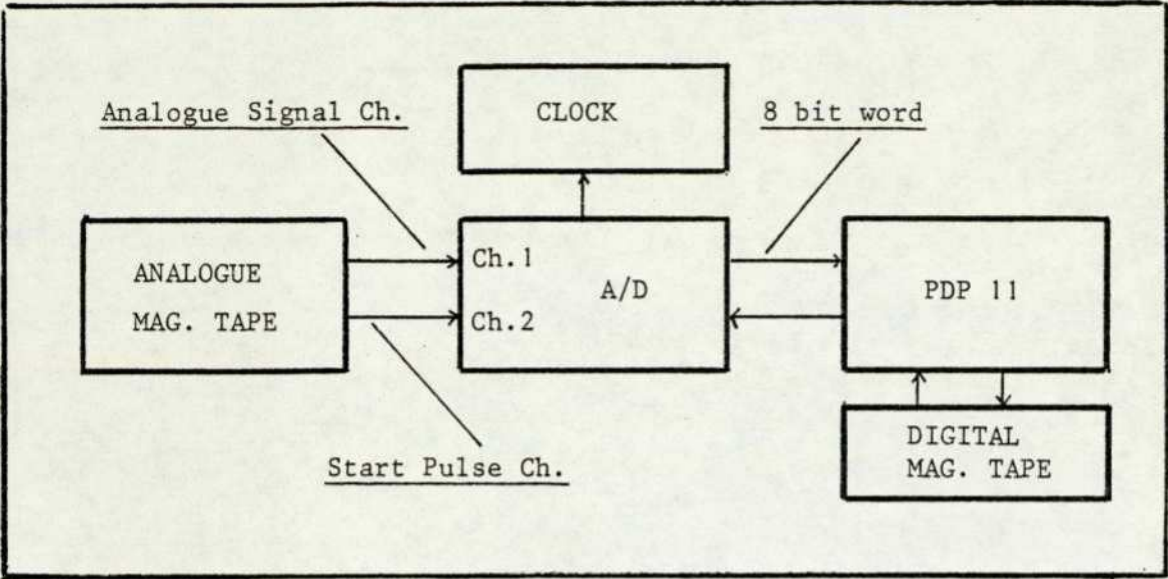


Fig. 10 The set up for transferring the analogue signal to a digital tape

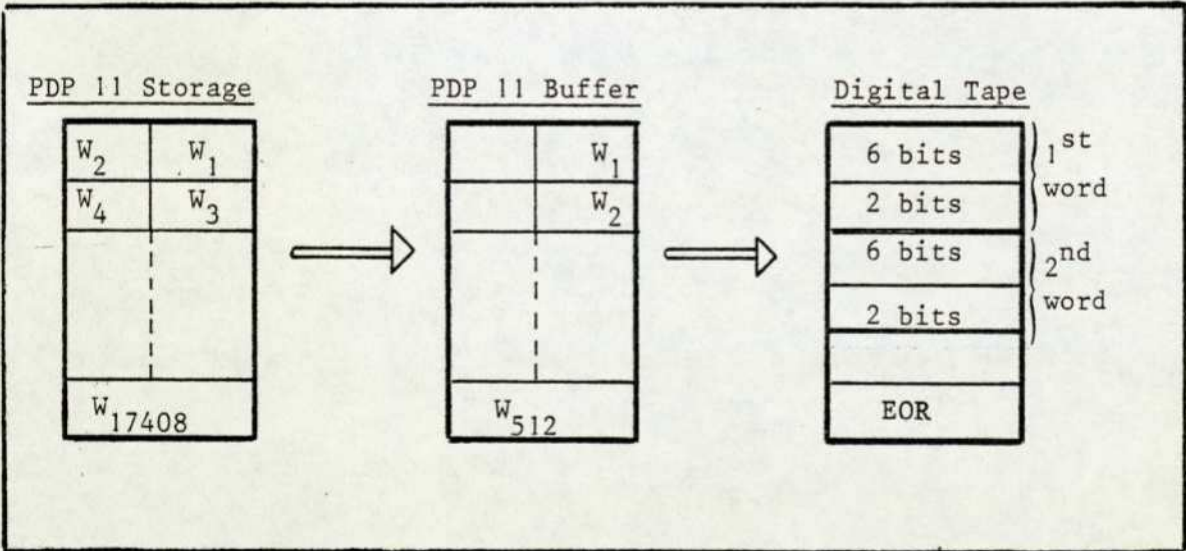


Fig. 11 Schematic diagram of the sampled signal being transferred from the PDP 11 to a digital tape

3) At this stage the digital tape was converted to a standard ULCC tape and a package of programs, to manipulate the stored signal, was created.

Fig's 12 and 13 are plots of the signals immediately after being stored on ULCC disk and 'unpacked'. These two signal were used in the estimation process. The plots confirm that the elaborated process of creating the bank of signals ended successfully.

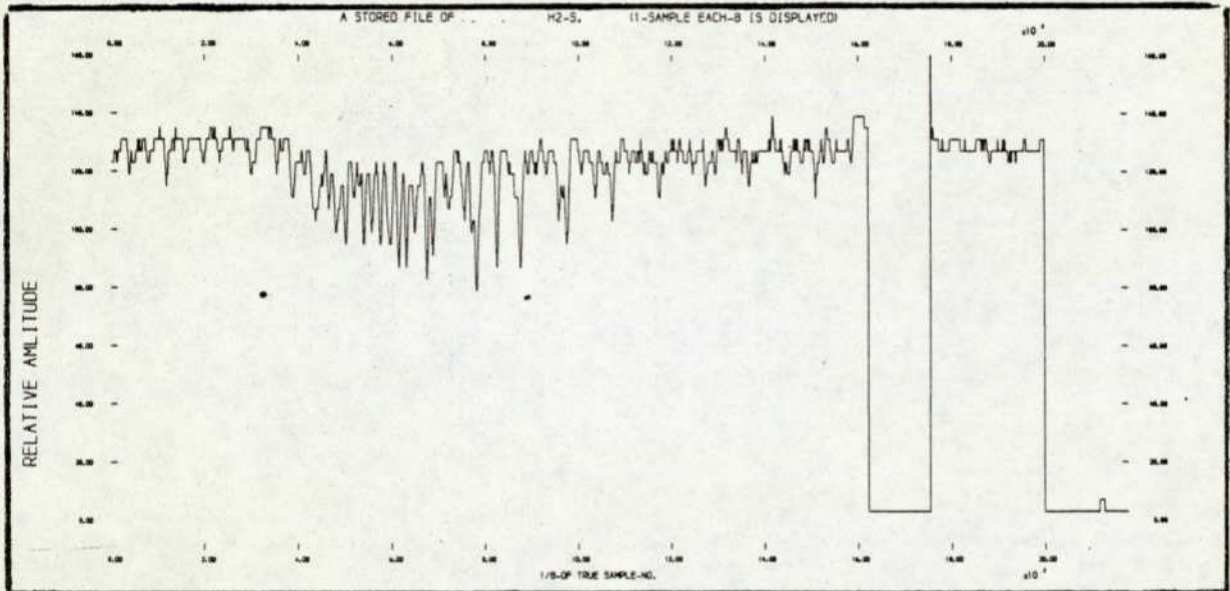


Fig. 12 A plot of the stored signal after unpacking it

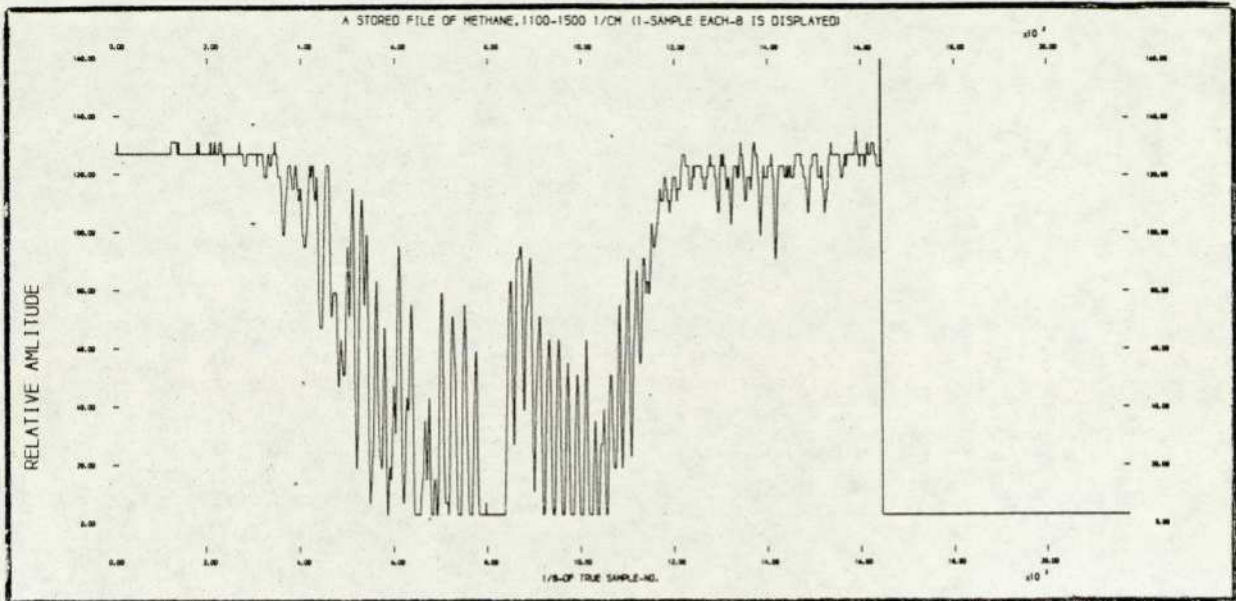


Fig. 13 A plot of the stored signal after unpacking it

the package performs the following functions :

UNPACK

Unpacked the sampled signal from 5 words of 12-bits each to an 8-bits word per computer word.

GETSLIT

Creates a discrete filter with a triangular shape response of unit area and slit width as a parameter.

CONVOLV

The routine convolves samples of stored record with a triangular slit filter.

ATTEN, GETMIX, GETNOIS, PLOT

This set of routines simulates different mixtures with different concentrations, adds non-correlated gaussian noise samples (zero mean, given variance) to samples of a spectrum and plots the various stages.

Fig 14 and 15 show application of the routines GETSLIT and CONVOLV, where a discrete signal record was convolved with 2. and 14.1/cm slit width transfer function.

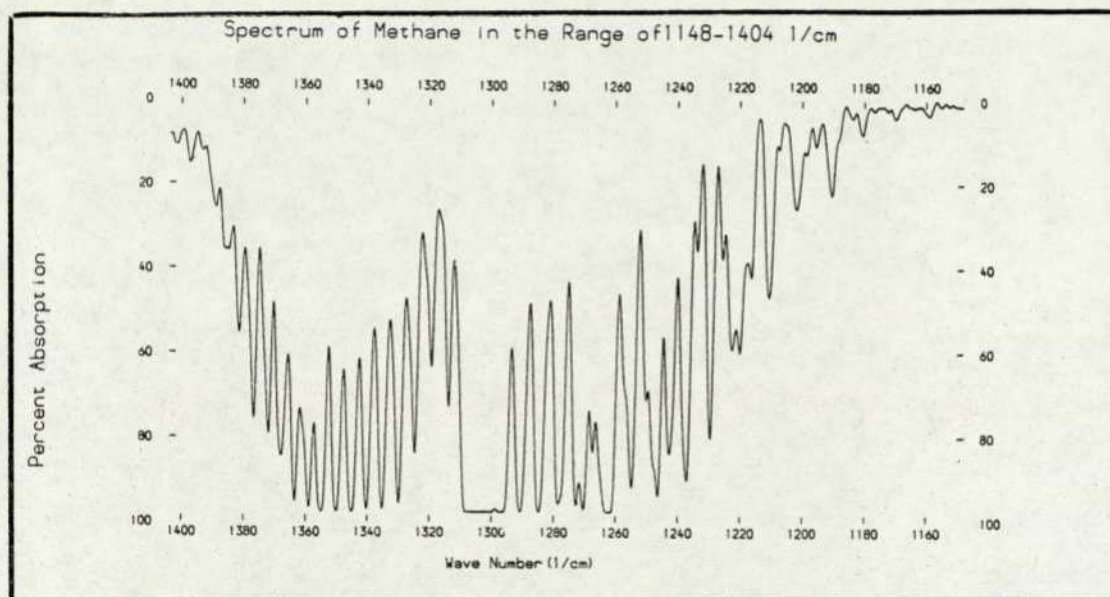


Fig.14 The result output of applying GETSLIT and CONVOLV to a stored signal (convolving with 2. 1/cm triangular slit)

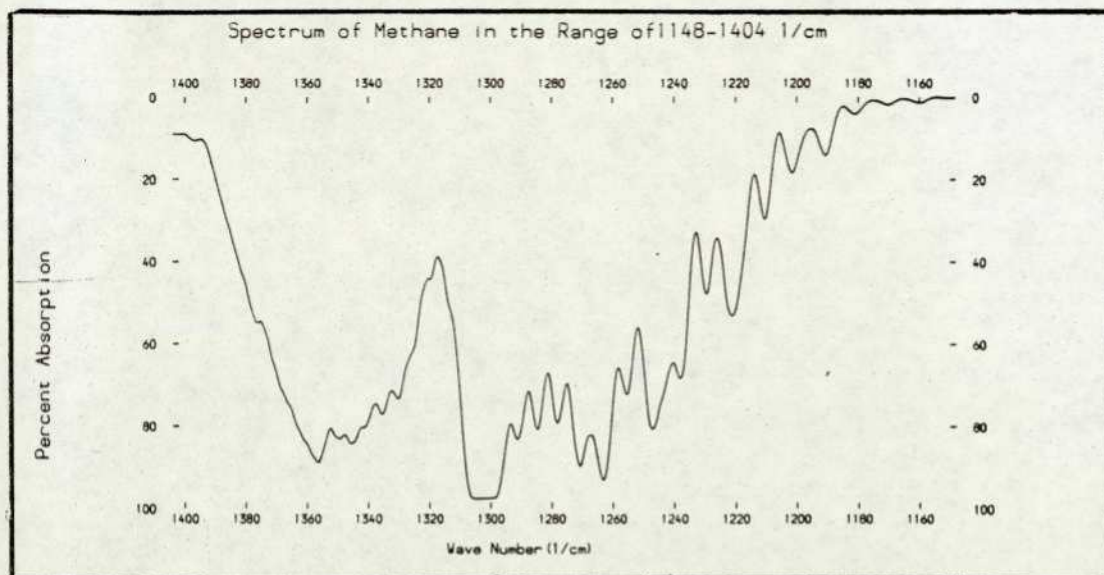


Fig. 15 The result output of applying GETSLIT and CONVOLV to a stored signal (convolving with 14. 1/cm triangular slit)

Fig.16 demonstrates the effect of adding gaussian random noise signal (using GETNOIS). Fig.17 shows the application of the routine ATTEN in simulating different concentrations.

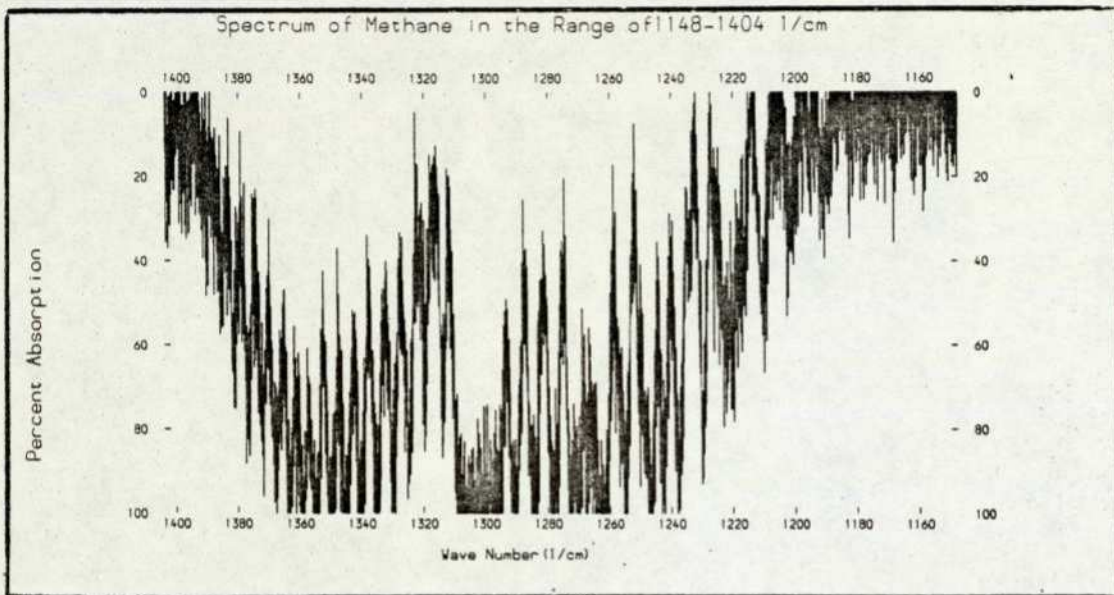


Fig. 16 Adding gaussian noise to a stored spectrum
(using GETNOIS)

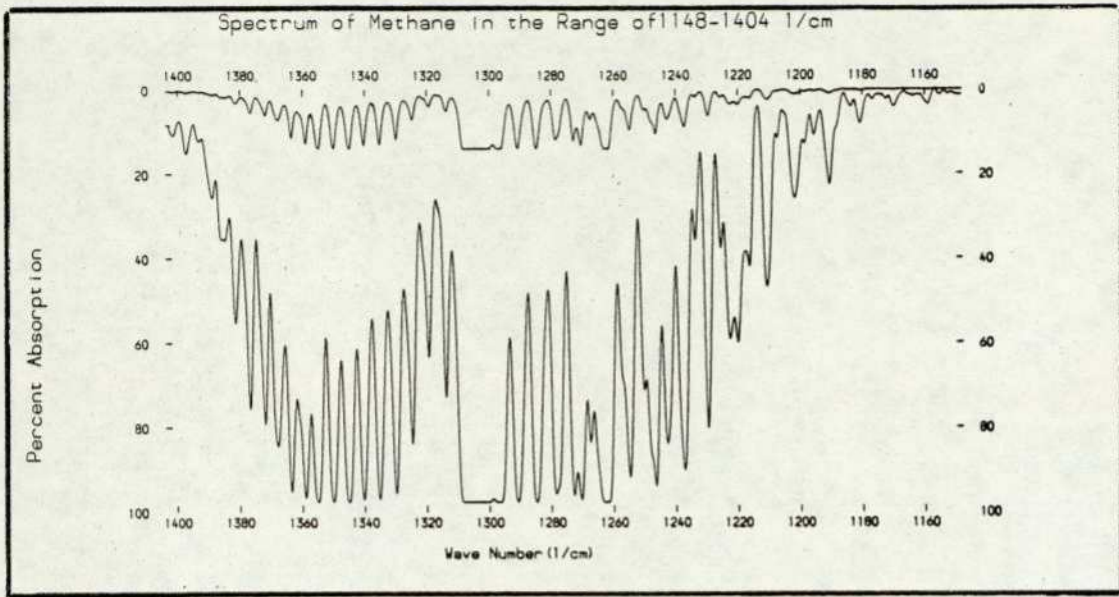


Fig. 17 Simulating spectra with different concentrations
(using ATTEN)

5.5

Summary and Conclusion

This chapter is a report of the simulation of two algorithms that were developed- the MLE (Maximum Likelihood Estimator) and the SE (Sequential Estimator). One set of simulations was aimed to confirm the value of the techniques developed and to find a figure of merit to indicate their potential ability to estimate the concentration of one gas in the presence of another (or several) at different noise levels.

The second aim was to validate some of the specific theoretical results developed :

(i) The theoretical expressions of the estimator variance as an

expression of the measured rms error.

(ii) The existence of an optimum cell length that yields minimum rms error in estimation of a concentration.

The above results establish the value of the expressions derived in analysing and predicting the performance in estimation concentration in general.

In part of the chapter (Sec.5.3) a brief report is given of the work involved in creating the bank of signals and the package of programs that were used in the above simulations.

6. Conclusions and recommendations

6.1 Summary and conclusions

The objectives of the research reported in this thesis were :

(1) To develop a discrete mathematical model to describe the electro optic process by which the beam is processed in the spectrophotometer. This includes three subprocesses :

- a) the absorption process (Beer - Lambert law),
- b) the process defined by the slit transfer function,
- c) The effect of the detector .

2) To develop a technique that estimates the concentration of a single compound in a mixture. The model developed in (1) was used in order to develop a search technique that finds the estimates of the concentration(s).

A substantial part of the work was devoted to implementation of the spectrophotometer slit transfer function (forward and inverse). This implementation was derived in the form of convolution/deconvolution formulas which were tested for their accuracy as a function of the signal frequency.

The objectives above were fulfilled in the sense that the model developed was used in constructing, validating and testing the estimation technique.

The research has opened the door to new questions and perhaps the way for new answers. It is our hope to solve some of the unanswered questions by applying the same approach we have adopted here, namely, developing and extending the theoretical basis, and testing the validity of the theoretical arguments by applying it to practical problems.

The development of the mathematical model in (1), was carried out along the following lines :

- a) Given a transfer function with an impulse response, in a polynomial form or a gaussian shaped form, a differential equation that describes the transfer function was found.
- b) A discrete form that represents the differential equation above was derived and implemented in the case of gaussian impulse response and triangular response systems.
- c) The technique developed in (a) and (b) was tested by applying it to the spectrophotometer inverse transfer function, and thus deriving a deconvolution formula. The limited accuracy of the formulas derived led us to propose the concept of partial deconvolution where the resolution of a signal is increased by a controlled amount which is measured by the slit width. In other words, applying the partial deconvolution formula to a signal which is an output of a spectrophotometer with a slit width-M, produces a signal which is an output of a spectrophotometer with a slit width-L. This concept was implemented for gaussian and triangular impulse response transfer functions.

In the case of the triangular impulse response transfer function it was found that the error in deconvolving higher frequency than $1/2(\text{slit width})$ increases substantially. A theoretical investigation shows that the range of existence of the implementation of the deconvolution (convolution) transfer function is defined by the location of the first pole of the transfer function. In the case of a triangular transfer function this pole is at (see Sec.3.4.2) :

$$\omega_p = \pi / \text{slit width}$$

Since the gaussian impulse response transfer function (and the inverse) does not have any poles in the finite complex plane, we conclude that theoretically, at least, it is possible to deconvolve any frequency required.

An approach that appears quite frequently in the literature is to transfer the signal to the frequency domain (using FFT) where the deconvolution operation is merely multiplying the signal by $(1/T(j\omega))$ and $T(j\omega)$ is the system transfer function transformed to the frequency domain (one should bear in mind that in our case the inverse transfer function has infinite poles on the $j\omega$ -axis, since $T(j\omega) = \sin^2(\frac{\omega M}{2}) / (\frac{\omega M}{2})^2$). It is of practical importance to test the ability of this approach to resolve a higher frequency than $1/2M$.

Other techniques that have solved the problem of limited resolution are [9] [10] [22] :

- a) interferometer spectroscopy,
- b) modulation techniques

In constructing the estimation algorithms two approaches have successfully been tested on perfect resolution signals, the MLE (Maximum Likelihood Estimator) and the SE (Sequential Estimator). One of them, the MLE, can easily be extended in order to be used with imperfect resolution signals. Both of the algorithms have been tested in estimation concentration of a single compound in the presence of another in practical S/N ratios (40-70db)

In formulating the MLE the gaussian search technique was used. According to this technique, the estimate concentration - $\hat{C}$ is the C that yields a minimum error where the error function is the sum of squares of the differences between the sampled signal and the simulated signal.

Part of the theoretical work was outlining an approach that shows that indeed the error function has a single minimum that may be reached from any initial C. As a result a bound on the minimum error that may be achieved is derived.

It seems that this approach may be generalized to some class of error functions that consists of the sum of squares. This generalization has not yet been made.

In assuming uncorrelated noise components with constant power the theoretical expression for the estimation error (estimation variance)

was found. The validity of the general expression was tested by comparing it to the rms error calculated over 10 trials on different S/N ratios and concentrations. Again, it seems that the theoretical expression of the estimator variance is general enough to describe the error in a large class of error functions.

The validation of the theoretical expression of the estimation variance enables us to use it in predicting the error in the estimation rather than performing elaborate simulations. Thus to create an engineering tool to assist in solving problems of optimization in development and design. As an example it is shown how the cell length may be optimized in order to achieve a minimum error in estimating the concentration. The theoretical expression may play an important role in predicting the error

The contributions of this work are summarized as follows :

(a) Two signal processing techniques for estimating the concentration of a single gas in a mixture were derived, implemented and tested. Both techniques can deal with a mixture of several gas compounds that have overlapped spectra. One of the techniques (SE), based on a high resolution spectrum signal, does not require storage for the spectrum signal and gives fast results. The second technique can accept a spectrum signal with any resolution, and by applying an iteration technique gives the estimated concentration of each compound as a result.

(b) A discrete mathematical model that incorporates the three main aspects of the spectrophotometer was created : the absorption process (Beer - Lambert law), the slit transfer function and the detector noise. Also, a block diagram is given which shows how the dual beam technique (See Ch.1) can be simulated on a computer.

(c) A technique of simulating an analogue process (continuous process) by a discrete process was developed. It has been shown that the time domain transfer function (time domain operator) which incorporates analogue operators (like D the differentiator) and discrete operators (like the central difference) can be derived. The time domain transfer function is then implemented (or approximated) by a series of discrete operators.

By using this approach, the time domain transfer function is found for finite symmetric impulse response systems. It is implemented discretely for the case of a triangular and a gaussian impulse response transfer function. Also, a discrete deconvolution formula (the inverse time domain transfer function) in the above case is implemented.

The partial deconvolution concept was developed and implemented as a mean of controlling the slit width of the transfer function (gaussian and triangular).

(d) An expression of the variance in estimating a vector of concentrations, $[c_1, c_2, \dots, c_n]$ was found. this expression is tested against the rms error, calculated over a set of trials, and found to be

in good agreement. It seems that this expression can be generalized so that it may be applied to some class of nonlinear estimation problems. It was pointed out that the expression found is similar to the variance in the case of a linear estimation problem. Consequently, it was suggested that engineering tools like the S/N ratio and the correlation factor that are in use in the linear estimation problem can be generalized so that they can be applied to nonlinear problems as well.

In investigating the error function, an approach was outlined aiming to prove the unimodality of an error function which consists of the sum of squares. As a result of this approach a bound on the error function that can be obtained is derived.

6.2 Recommendation for further research

Further research work should be carried out along the following lines :

(1) Extending and developing further the package of supporting programs
It is proposed to transfer the package from main frame computer (CDC, 7600) to the interactive system in the Instrumentation Centre at the Systems Science department, The City University. The package of programs can be divided into three sets :

(i) Utility programs

this set will include programs like PACK which packs a given signal for storing on a disk in an efficient way and UNPACK for displaying the signal.

(ii) Simulation programs

This includes programs like GETNOIS and ATTEN for simulating certain situations of noise levels, concentrations etc.

(iii) Signal processing techniques

This set includes programs that perform convolution/deconvolution, parameter estimation etc.

The three sets of programs are designed to serve three tasks :

(i) Extending the bank of signals by storing more signals and creating facilities for retrieving any signal stored efficiently.

(ii) Simulating tests in order to provide the tool for validation of further theoretical work.

(iii) Providing the facilities for testing practical problems in parameter estimation, signal filtering etc.

(2) Some of the theoretical questions that still open are :

(i) Application of Fast Fourier Transform (FFT) to the estimation problem:

In many cases the ir spectrum signal has an audio signal spectrum with most of its energy concentrated in a few

harmonics. Consequently it may be that in some practical cases it is more efficient to include the forward/backward FFT in the estimation loop (see Introduction Ch.4) rather than using the implementation of the slit transfer function in the time domain.

(ii) Application of FFT in simulating analogue processes :

By simulation of an analogue process we mean performing a discrete process like a discrete convolution so that the resulting discrete output is equal (in appropriate time intervals) to the output of the given analogue process (see Ch.2 Introduction). FFT is in common use nowadays in performing discrete linear processes. Yet, very little has been said about the requirements that this technique should fulfil in order to be suitable to simulate an analogue process. The importance of the discrete simulation of analogue processes lies in the fact that the signal source (spectrophotometer) is analogue, whereas the tool for processing it is a discrete tool (computer).

(iii) Further investigation into the deconvolution technique that has been developed :

As was said before, the deconvolution formula developed is valid for a range of frequencies of $1/2(\text{slit width})$. It is an interesting question with practical aspects as well whether it is possible to extend this range beyond its current limit, and, at the same time, preserve its feature as

an analogue simulator.

(iv) New modulation techniques in solving the resolution problem :

Modulation techniques like Hadamard transformation [10] have been used successfully in increasing the resolution of the spectrophotometer. While the signal processing that this technique requires is straight forward, an alteration in the spectrophotometer structure is needed. It is still an open question whether there is a technique that requires less alteration in the spectrophotometer structure and puts more emphasis on the signal processing needed for the benefit of, say, a software means of controlling the resolution.

6.3 Practical application in spectrophotometric instrumentation

The contributions of the work for practical application were :

(1) Developing and testing (by simulation) a fast deconvolution technique that requires a small memory storage (the underlying formula that was tested had 11 parameters and requires about 34 locations. See Sec 3.4 and 3.5).

The technique can be applied to a signal sample entering a u-processor unit rather than waiting for the whole ensemble of samples and enables a trade-off between the resolution and the noise level of the signal (partial deconvolution formula). It was proposed to implement this technique on a u-processor, in the first stage, as an auxiliary unit and later as an integral part of the spectrophotometer, which thus provides a software means for controlling the slit width.

(2) Creating two concentration estimation algorithms that may be applied to a spectrum embedded in noise, which consists of several compounds with overlapped spectra. Since the estimation algorithms include a built-in filtering mechanism, no additional filtering is needed. The limitation in the scanning speed is, therefore, imposed not by the filter time constant, but by either the ir detector or the amount of computation needed between two samples that enter the u-processor unit.

6.3.1 Some considerations in implementing the sequential estimator algorithm

Below, the basic considerations that have to be taken into account in implementing the SE algorithm on a u-processor are discussed. Assuming a case of a mixture with two gas compounds with absorption coefficients :

$$A = (\alpha_1(v_1), \alpha_1(v_2), \dots, \alpha_1(v_M))$$

$$B = (\alpha_2(v_1), \alpha_2(v_2), \dots, \alpha_2(v_M))$$

The two stages in the estimation are :

(a) Creating two reference vectors

In order to store the reference vector A, a cell with 100% of compound A has to be installed in the spectrophotometer (see Fig.1). The wave number mark initiates the process of storing the sampled spectrophotometer output, $[\alpha(v_M), \dots, \alpha(v_M)]$ in the u-processor memory. This process is stopped after storing M samples. The process is repeated with a cell with 100% of compound B (see Fig.2).

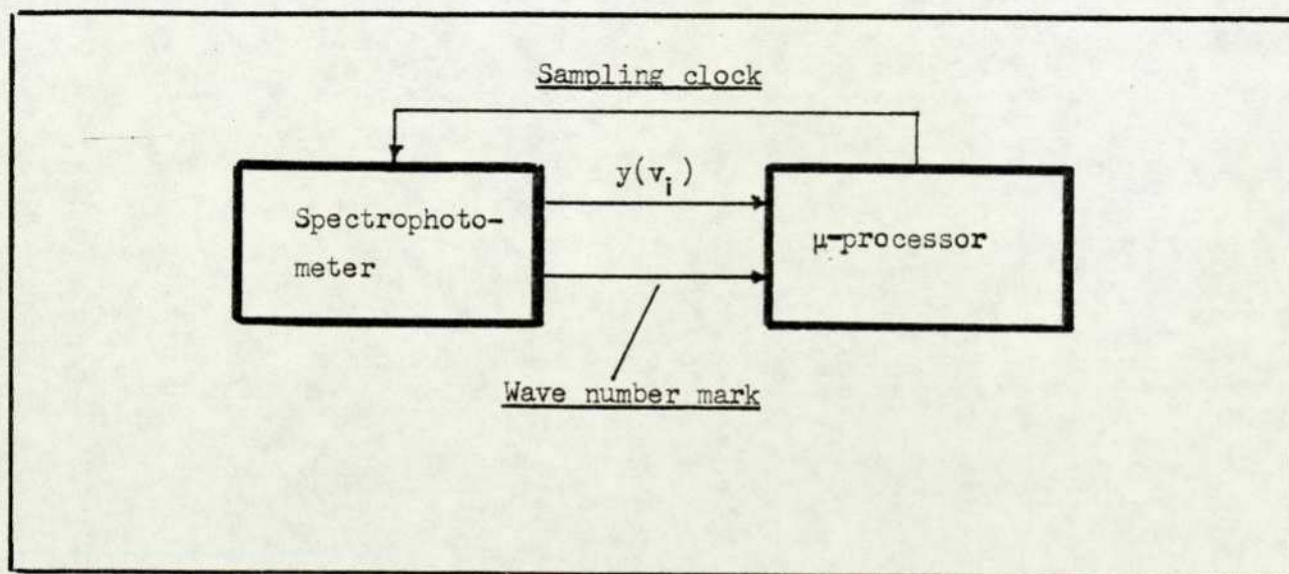


Fig. 1 Connecting a u-processor unit to the spectrophotometer for performing concentration estimation

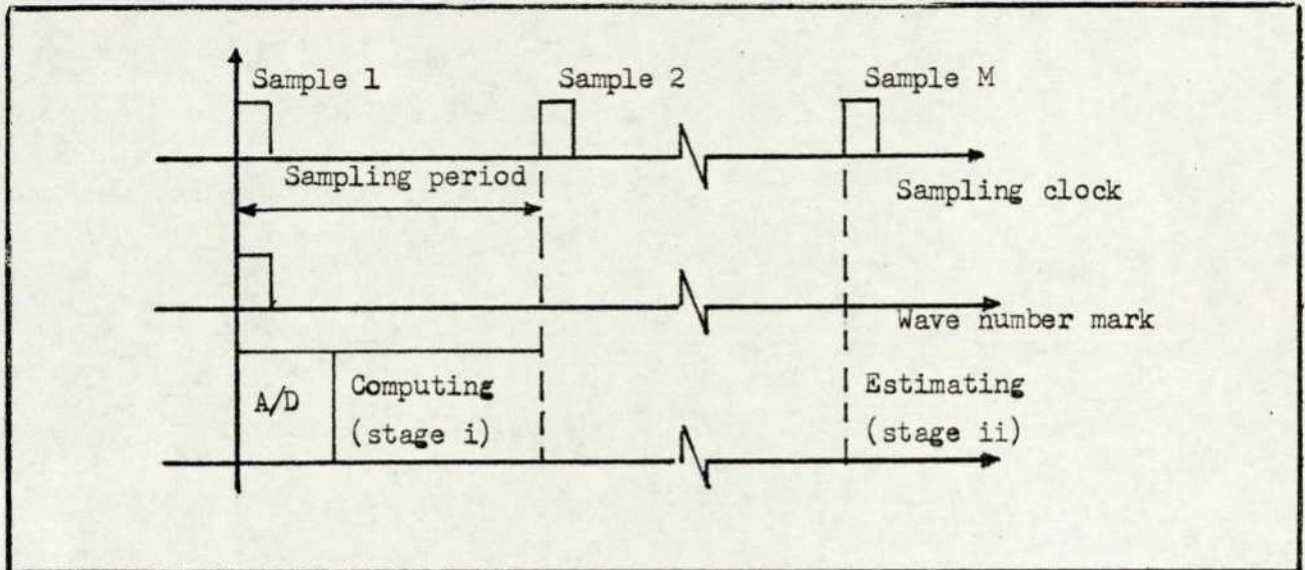


Fig. 2 Timing of the inter-connection between the u-processor and the spectrophotometer

(b) Estimation

(i) In this stage, a cell with the unknown concentrations is installed in the spectrophotometer and the sampled output, $y(v_i)$ $i=1, \dots, M$, are entered the u-processor unit. This time the wave number mark initiates the following computations on every $y(v_i)$ and $\log(y_i)$ (see Sec.5.2.2) :

	Operation	Approximation of time duration
(1) $y^2(v_i)$	Mult.	50 usec
(2) $\log(y(v_i))$	Log.	260. usec
(3) $\alpha_1(v_i)\alpha_2(v_i)$	Mult.	50. usec
(4) $DR_{12} = y^2\alpha_1(v_i)\alpha_2(v_i)$	"	" "
(5) $\alpha_1^2(v_i)$	"	" "
(6) $DR_{11} = y^2\alpha_1^2(v_i)$	"	" "
(7) $\alpha_2^2(v_i)$	"	" "
(8) $DR_{22} = y^2\alpha_2^2(v_i)$	"	" "
(9) $DR_{y1} = y^2(v_i)\alpha_1(v_i)\log[y(v_i)]$		" "
(10) $y^2(v_i)\log(v_i)$	"	" "
(11) $DR_{y2} = y^2(v_i)\alpha_2(v_i)\log[y(v_i)]$		" "
(12) $ACR_{11} = ACR_{11} + DR_{11}$	Addition	20 "
(13) $ACR_{12} = ACR_{12} + DR_{12}$	"	" "
(14) $ACR_{22} = ACR_{22} + DR_{22}$	"	" "
(15) $ACR_{y1} = ACR_{y1} + DR_{y1}$	"	" "
(16) $ACR_{y2} = ACR_{y2} + DR_{y2}$	"	" "
(17)	Comparison	2usec.

It is suggested that a multiplication facility be implemented in hardware in order to achieve this sort of time response. The log operation may be implemented by a piece wise polynomial of a second order which will describe the logarithm function in the range of

interest. Such an implementation will require about the time duration of four multiplications and three additions ($a_0 + a_1x + a_2x^2$), which is only about 260 usec. The total time of the above operations is, therefore, 722usec. Adding the time needed for the A/D operation it is, roughly, less than 1 msec.

(ii)

When the number of samples reaches M the calculation of the concentrations is carried out :

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} ACR_{11} & ACR_{12} \\ ACR_{21} & ACR_{22} \end{bmatrix}^{-1} \begin{bmatrix} ACR_{y1} \\ ACR_{y2} \end{bmatrix}$$

A scanning time over a range of 3600 [l/cm] (Py Unicam 4000) can be as little as 4 min. The scanning speed is :

$$\text{Scan. Speed} = 3600/120 = 30 \text{ [l/cm/sec]}$$

which shows that in the case of two compounds the theoretical sampling period may be as low as .03 [l/cm], which is far beyond the sampling rate required in practice while computation of the necessary variables can still be performed.

REFERENCES

1. Steele D. Theory of Vibrational Spectroscopy.
Saunders Company, 1971.
2. Brown E.B. Modern Optics (ch. 9).
Chapman and Hall, 1966.
3. Mattson J.S. Computers in Chemistry and
Mark H.B. Instrumentation, Vol.7.
Macdonald H.C. Mark Deker, 1977.
4. Seshadri K.S. The shapes and intensities of ir
Jones R.N. absorption bands - a review.
Spectrochimica Acta, 19, p.1013, 1963.
5. Hudson R.D. Infra red Systems Engineering.
John Wiley, 1969.
6. Brown E.B. Ref. 2, ch. 7.
7. Davenport W.B. An Introduction to the Theory of
Root W. Random Signals and Noise.
McGraw Hill, 1958.
8. Papoulis A. Probability, Random Variables, and
Stochastic Processes.
McGraw Hill, 1965.
9. Barringer A.R. Progress in the remote sensing of
Schuck J.P. vapours for air pollution: geologic
and oceanographic applications.
Measurement & Control, 11, July 1978.

10. Nelson E.D. Hadamard spectroscopy.
Fredman M.L. Journal of the Optical Society of America, 60 (12), 1970.
11. Pye Unicam SP 4000 manual, 1974.
12. Schunck G. Non dispersive ir gas analysers for industrial processes and the protection of the environment.
Measurement & Control, 11, July 1978.
13. Moffat A.J. The application of optical correlation techniques to the remote sensing of SO₂ plumes using sky light.
Millan M.M. Atmospheric Environment, 5, p.677, 1971.
14. Papoulis A. Signal Analysis.
McGraw Hill, 1977.
15. Papoulis A. The Fourier Integral and its Application.
McGraw Hill, 1962.
16. Hill I.R. Rayleigh (Strutt) method of making an approximate correction to spectra for finite slit widths.
Steele D. The Journal of the Chemical Society, Faraday Transactions II, 70, 1974.
17. Abramowitz M. Handbook of Mathematical Functions.
Stegun I.A. Dover, 1965.

18. Churchill R.V. Complex Variables and Applications.
Brown J.W. McGraw Hill, 1974.
Verhey R.F.
19. Acton F.S. Numerical Methods that Work.
Harper, 1970.
20. Van Trees H.L. Detection, Estimation and Modulation
Theory. Part 1.
John Wiley, 1968.
21. Wilde D.J. Foundations of Optimization.
Beightler C.S. Prentice Hall, 1967.
22. Griffiths P.R. Chemical Infra Red and Fourier
Transformation.
John Wiley, 1975.

APPENDIX 1

We would like to approximate the operator $\frac{(MD)^2}{\delta_M^2}$ as a series of operators:

$$\frac{(MD)^2}{\delta_M^2} \approx \sum_{n=0}^N a_n \cosh(nD)$$

We have (see Equation 7, Chapter 3):

$$\frac{(MD)^2}{\delta_M^2} = \frac{1}{2} \frac{(MD)^2}{\cosh(MD) - 1}$$

Consequently, the following equation has to be fulfilled:

$$a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND) \approx \frac{M^2 D^2}{2\{\cosh(MD) - 1\}}$$

$$\{a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND)\} \{\cosh(MD) - 1\} \approx \frac{1}{2} M^2 D^2$$

Developing the above to a series in D^{2j} we have:

$$\begin{aligned} & a_0 \left\{ 1 + \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + \dots \right\} + \\ & a_1 \left\{ 1 + \frac{D^2}{2!} + \frac{D^4}{4!} + \dots + \frac{D^{2j}}{(2j)!} + \dots + \frac{D^{2N}}{(2N)!} \right\} \left\{ \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + \dots \right\} + \\ & a_2 \left\{ 1 + \frac{(2D)^2}{2!} + \frac{(2D)^4}{4!} + \dots + \frac{(2D)^{2j}}{(2j)!} + \dots + \frac{(2D)^{2N}}{(2N)!} \right\} \left\{ \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + \dots \right\} + \\ & \vdots \\ & a_k \left\{ 1 + \frac{(kD)^2}{2!} + \frac{(kD)^4}{4!} + \dots + \frac{(kD)^{2j}}{(2j)!} + \dots + \frac{(kD)^{2N}}{(2N)!} \right\} \left\{ \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + \dots \right\} + \\ & \vdots \\ & a_N \left\{ 1 + \frac{(ND)^2}{2!} + \frac{(ND)^4}{4!} + \dots + \frac{(ND)^{2j}}{(2j)!} + \dots + \frac{(ND)^{2N}}{(2N)!} \right\} \left\{ \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + \dots \right\} + \end{aligned}$$

APPENDIX 1

$$\begin{aligned}
 a_0 & \left\{ \frac{(MD)^2}{2!} + \frac{(MD)^4}{4!} + - - - - - + \frac{(MD)^{2j}}{(2j)!} + - - + \right\} + \\
 a_1 & \left\{ \frac{(MD)^2}{2!} + \left(\frac{(MD)^4}{4!} + \frac{M^2 D^4}{(2!)^2} \right) + \left(\frac{(MD)^6}{6!} + \frac{M^4 D^6}{4! 2!} + \frac{M^2 D^6}{2! 4!} \right) + - - + \left(\frac{(MD)^{2j}}{(2j)!} + \right. \right. \\
 & \left. \left. \frac{M^{2j-2} D^{2j}}{(2j-2)! 2!} + \frac{M^2 D^{2j}}{2! (2j-2)!} \right) + - - + \right\} + \\
 & \left. \left. \left. \begin{array}{c} | \\ | \\ | \\ | \\ | \end{array} \right. \right. \right. \\
 a_k & \left\{ \frac{(MD)^2}{2!} + \left(\frac{(MD)^4}{4!} + \frac{k^2 M^2 D^4}{(2!)^2} \right) + \left(\frac{(MD)^6}{6!} + \frac{M^4 D^6 k^2}{4! 2!} + \frac{M^2 D^6 k^4}{2! 4!} \right) + - - \right. \\
 & \left. + \left(\frac{(MD)^{2j}}{(2j)!} + \frac{M^{2j-2} D^{2j} k^2}{(2j-2)! 2!} + - - + \frac{M^2 D^{2j} k^{2j-2}}{2! (2j-2)!} \right) + - - + \right\} + \\
 a_N & \left\{ \frac{(MD)^2}{2!} + \left(\frac{(MD)^4}{4!} + \frac{N^2 M^2 D^4}{(2!)^2} \right) + \left(\frac{(MD)^6}{6!} + \frac{M^4 D^6 N^2}{4! 2!} + \frac{M^2 D^6 N^4}{2! 4!} \right) + - - \right. \\
 & \left. + \left(\frac{(MD)^{2j}}{(2j)!} + \frac{M^{2j-2} D^{2j} N^2}{(2j-2)! 2!} + - - + \frac{M^2 D^{2j} N^{2j-2}}{2! (2j-2)!} \right) + - - + \right\} +
 \end{aligned}$$

The general expression is:

$ \sum_{k=0}^N \sum_{\ell=0}^{j-1} \frac{D^{2j} M^{2(j-\ell)} k^{2\ell}}{2^{2(j-\ell)} (2\ell)!} a_k = \frac{(MD)^2}{2} $
$ \begin{array}{ll} j = 1 - - - N + 1 & k^{2\ell} = \begin{cases} 1 & k = 0, \ell = 0 \\ 0 & k = 0, \ell \neq 0 \end{cases} \\ k = 0 - - - N & \end{array} $

$$\frac{M^2 D^2}{\delta_M^2} \approx$$

$$a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + - - - a_N \cosh(ND)$$

APPENDIX 1

The above equation incorporates a set of linear equations in $a_0, a_1, \dots, a_N$. The expression found for $(MD)^2/\delta_M^2$ will be accurate for a signal with zero derivatives of order above N.

Generalising the results

Instead of doing complete deconvolution, we may perform the consequence of deconvolution and convolution:

$$\left(\frac{1}{M^2} \frac{(MD)^2}{\delta_M^2} \right) \left(L^2 \frac{\delta_L^2}{(LD)^2} \right) = \frac{\delta_L^2}{\delta_M^2}$$

$$x(t) = y(t) \frac{\delta_L^2}{\delta_M^2}$$

Deconvolution: $L < M$

Convolution : $L > M$

$$\delta_M^2 = \{\cosh(MD) - 1\} 1/2$$

$$\delta_L^2 = \{\cosh(LD) - 1\} 1/2$$

$$\frac{\delta_L^2}{\delta_M^2} = \frac{\cosh(LD) - 1}{\cosh(MD) - 1} = a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND)$$

$$\cosh(LD) - 1 = \{\cosh(MD) - 1\} \{a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND)\}$$

APPENDIX 1

Thus we shall have the same set of equations as in the last case but the right-hand side is now:

$$\cosh(LD) - 1 = \frac{(LD)^2}{2!} + \frac{(LD)^4}{4!} + \dots + \frac{(LD)^{2N}}{(2N)!}$$

Therefore we have:

$$\sum_{k=0}^N \sum_{\ell=0}^{j-1} \frac{D^{2j} M^{2(j-\ell)} k^{2\ell}}{(2(j-\ell))! (2\ell)!} = \frac{(LD)^2}{2!} + \frac{(LD)^4}{4!} + \dots + \frac{(LD)^{2(N+1)}}{(2(N+1))!}$$

$$j = 1 \dots N + 1 \quad k^{2\ell} = \begin{cases} 1 & k = 0, \ell = 0 \\ 0 & k = 0, \ell \neq 0 \end{cases}$$

$$k = 0 \dots N$$

$$\frac{\delta_L^2}{\delta_M^2} = \frac{\cosh(LD) - 1}{\cosh(MD) - 1} =$$

$$a_0 + a_1 \cosh(D) + a_2 \cosh(2D) + \dots + a_N \cosh(ND)$$

APPENDIX 2

We would like to approximate the operator $\exp\{-\frac{1}{2}(DL)^2\}$ by a series of operators:

$$\exp\{-\frac{1}{2}(DL)\} = \sum_{n=0}^N a_n \cosh(nD)$$

Again, the approach is to develop the two sides of the equation to a power series in D:

$$\exp\{-\frac{1}{2}(DL)^2\} = 1 - \frac{(DL)^2}{2} \frac{1}{1!} + \frac{(DL)^4}{2} \frac{1}{2!} + \dots + (LD)^{2j} \frac{(-1)^j}{(2j)!} + \dots$$

$$\cosh(nD) = 1 + \frac{(nD)^2}{2!} + \frac{(nD)^4}{4!} + \dots + \frac{(nD)^{2j}}{(2j)!} + \dots$$

Collecting terms with the same power, we have the following set of equations:

$$a_0 + a_1 + \dots + a_n + \dots + a_N = 1$$

$$0 + \frac{D^2}{2!} a_1 + \dots + \frac{(nD)^2}{2!} a_n + \dots + \frac{(ND)^2}{2!} a_N = -\frac{(LD)^2}{2}$$

$$0 + \frac{D^4}{4!} a_1 + \dots + \frac{(nD)^4}{4!} a_n + \dots + \frac{(ND)^4}{4!} a_N = \frac{(LD)^4}{2^2 2!}$$

$$0 + \frac{D^{2j}}{(2j)!} a_1 + \dots + \frac{(nD)^{2j}}{(2j)!} a_n + \dots + \frac{(ND)^{2j}}{(2j)!} a_N = \frac{(LD)^{2j} (-1)^j}{2^j j!}$$

APPENDIX 2

Thus the set of equations may be written:

$$\sum_j \frac{a_n (nD)^{2j}}{(2j)!} = \frac{(LD)^{2j} (-1)^j}{2^j j!}$$
$$\begin{array}{l} j = 0, - , N \\ n = 0, - , N \end{array} \quad (nD)^{2j} = \begin{cases} 1 & j = 0, n = 0 \\ 0 & j \neq 0, n = 0 \end{cases}$$

$$\exp\{-\frac{1}{2}(LD)^2\} \cong \sum_{n=0}^N a_n \cosh(nD)$$

Thus the implementation above is accurate for signals with zero derivative of order greater than N.

Appendix 3

Tables for deconvolution formulas in the case of gaussian and triangular impulse response systems

Table 1 gives the coefficients $\{a_n\}$ in expanding the inverse transfer function of a system with a triangular response :

$$\frac{(MD)^2}{\delta_M^2} = \sum_{n=0}^{10} a_n \cosh(nD)$$

where M is the slit width.

The expansion is given for values of $M = 1, 2, 3, 4, 5$ and 6 . M is measured in sampling period units (M=1 means the slit width is equal to the sampling period)

M	1	2	3	4	5	6
a_0	1.342e+00	5.511e+00	1.561e+02	8.146e+03	3.094e+05	7.599e+06
a_1	-4.323e-01	-6.962e+00	-2.720e+02	-1.465e+04	-5.595e+05	-1.377e+07
a_2	1.183e-01	3.569e+00	1.862e+02	1.065e+04	4.130e+05	1.023e+07
a_3	-3.624e-02	-1.546e+00	1.018e+02	-6.248e+03	-2.478e+05	-6.195e+06
a_4	1.869e-02	5.629e-01	4.462e+01	2.941e+03	1.198e+05	3.301e+06
a_5	-2.805e-03	-1.693e-01	-1.555e+01	-1.097e+03	-4.600e+04	-1.180e+06
a_6	6.174e-04	4.088e-02	4.227e+00	3.177e+02	1.372e+04	3.575e+05
a_7	-1.077e-04	-7.614e-03	-8.663e-01	-6.896e+01	-3.067e+03	-8.120e+04
a_8	1.382e-05	1.025e-03	1.261e-01	1.057e+01	4.837e+02	1.302e+04
a_9	-1.154e-06	-8.875e-05	-1.164e-02	-1.022e+00	-4.805e+01	-1.315e+03
a_{10}	4.686e-08	3.706e-06	5.126e-04	4.688e-02	2.262e+00	6.293e+01

Table 1

Table 2 gives the coefficients $\{a_i\}$ of the expansion of the following operator to a series of $\cosh(nD)$:

$$\left(\frac{M}{L}\right)^2 \frac{\delta_L^2}{\delta_M^2} = \sum_{n=0}^{10} a_n \cosh(nD)$$

where M is the given slit width (in sampling period units), and L is the simulated slit width. The table was calculated for :

$$M = 6$$

and

$$L = 1, 2, 3, 4 \text{ and } 5$$

L	1	2	3	4	5
a_0	7.007e+06	5.443e+06	3.448e+06	1.643e+06	4.609e+05
a_1	1.269e+07	-9.864e+06	6.249e+06	-2.977e+06	-8.351e+05
a_2	9.430e+06	7.325e+06	4.640e+06	2.210e+06	6.198e+05
a_3	-5.711e+06	-4.436e+06	-2.809e+06	-1.377e+06	-3.749e+05
a_4	2.794e+06	2.170e+06	1.374e+06	6.541e+05	1.831e+05
a_5	-1.088e+06	-8.449e+05	-5.349e+05	-2.545e+05	-7.121e+04
a_6	3.295e+05	2.558e+05	1.619e+05	7.700e+04	2.152e+04
a_7	-7.485e+04	-5.810e+04	-3.676e+04	-1.747e+04	-4.880e+03
a_8	1.200e+04	9.315e+03	5.891e+03	2.799e+03	7.810e+02
a_9	-1.212e+03	-9.405e+02	-5.946e+02	-2.823e+02	-7.872e+01
a_{10}	5.800e+01	4.500e+01	2.844e+01	1.350e+01	3.760e+00

Table 2

Table 3 gives the coefficients of the expansion of the inverse transfer function with a gaussian impulse response to a series of $\cosh(nD)$:

$$\exp\left(-\frac{1}{2}(\sigma D)^2\right) = \sum_{n=0}^{10} a_n \cosh(nD)$$

The table was calculated for :

$$\sigma^2 = 1, 3, 6, 9, 1/3, 1/6, \text{ and } 1/9$$

σ is measured in sampling period units.

App. 3

σ^2	1	3	6	9
a_0	8.425e+00	6.730e+02	6.081e+04	1.387e+06
a_1	-1.144e+01	-1.178e+03	-1.091e+05	-2.504e+06
a_2	5.781e+00	8.032e+02	7.889e+04	1.837e+06
a_3	-2.410e+00	-4.327e+02	-4.584e+04	-1.091e+06
a_4	8.438e-01	1.856e+02	2.129e+04	5.205e+05
a_5	-2.542e-01	-6.310e+01	-7.820e+03	-1.967e+05
a_6	5.761e-02	1.672e+01	2.225e+03	5.767e+04
a_7	-1.049e-02	-3.346e+00	-4.744e+02	-1.265e+04
a_8	1.388e-03	4.764e-01	7.142e+01	1.958e+03
a_9	-1.185e-04	-4.308e-02	-6.782e+00	-1.908e+02
a_{10}	4.894e-06	1.862e-03	3.059e-01	8.815e+00

Table 3.1

σ^2	1/3	1/6	1/9
a_0	1.821e+00	1.322e+00	1.199e+00
a_1	-1.081e+00	-4.019e-01	-2.434e-01
a_2	3.445e-01	1.028e-01	5.685e-02
a_3	-1.113e-01	-3.050e-02	-1.639e-02
a_4	3.364e-02	8.856e-03	4.702e-03
a_5	-8.930e-03	-2.302e-03	-1.215e-03
a_6	1.978e-03	5.039e-04	2.651e-04
a_7	-3.465e-04	-8.758e-05	-4.598e-05
a_8	4.459e-05	1.121e-5	5.879e-06
a_9	-3.731e-06	-9.349e-07	-4.897e-07
a_{10}	-1.517e-07	3.793e-08	1.985e-08

Table 3.2

Table 4 gives the first fifty coefficients $\{a_i\}$ of the expansion of the inverse transfer function with a gaussian impulse response to a series of $\cos(\omega nT)$:

$$\exp\left(\frac{1}{2} (\sigma D)^2\right) = \sum_{n=0}^{10} a_n \cosh(\omega nT)$$

where :

$$\omega = \frac{\pi}{T}$$

$$\sigma = T_s$$

The table shows that the absolute value of the coefficients decrease to zero very slowly. Consequently, large number of coefficients would be required to simulate the inverse transfer function (see Ch.2, Introduction)

1.040e+01	-2.669e+01	1.944e+01	-1.393e+01
1.014e+01	-7.563e+00	5.793e+00	-4.545e+00
3.644e+00	-2.977e+00	2.472e+00	-2.082e+00
1.776e+00	-1.531e+00	1.333e+00	-1.170e+00
1.035e+00	-9.217e-01	8.258e-01	-7.441e-01
6.737e-01	-6.128e-01	5.598e-01	-5.133e-01
4.723e-01	-4.360e-01	4.037e-01	-3.749e-01
3.490e-01	-3.257e-01	3.047e-01	-2.856e-01
2.682e-01	-2.524e-01	2.379e-01	-2.247e-01
2.125e-01	-2.012e-01	1.909e-01	-1.813e-01
1.724e-01	-1.642e-01	1.565e-01	-1.494e-01
1.427e-01	-1.365e-01	1.306e-01	-1.252e-01
1.200e-01	-1.152e-01		

Table 4

Appendix 4

The estimation problem in the linear case

Given a discrete signal consisting of :

$$Y = \sum_{j=1}^n c_j X_j + N \quad (1.)$$

where Y is the measurement vector :

$$Y = (y_1, \dots, y_M)$$

the vectors X_j $j=1, \dots, n$:

$$X_j = (x_j(v_1), \dots, x_j(v_M))$$

are the partial signals, and c_j $j=1, \dots, n$ are the unknown concentrations. The vector N :

$$N = (n_1, \dots, n_M)$$

is a vector of gaussian variables, where each of the variable is described by :

$$\overline{n_i} = 0$$

$$\overline{n_i n_j} = \begin{cases} 0 & i \neq j \\ \sigma^2 = N_0 & i = j \end{cases}$$

We would like to find estimator

$$C = (\hat{c}_1, \dots, \hat{c}_n)$$

given the vector Y and $X_1, \dots, X_n$. The MLE estimation rule is given by [20] :

$$\frac{\partial}{\partial c_i} (Y - \sum_j c_j X_j) \Big|_{c_i = \hat{c}_i} = 0$$

Performing the differentiation above we have the following set of equations :

$$2(Y - \sum_j \hat{c}_j X_j) X_i = 0 \quad (1a.)$$

$$i = 1, \dots, n$$

where by $X_i X_j$ we mean :

$$X_i X_j = \sum_{k=1}^n x_j(v_k) x_i(v_k)$$

Open bracket in Eq.1_a we have the following set of equations :

$$\sum_j \hat{c}_j X_j X_i = Y X_i$$

$$i = 1, \dots, n$$

In a matrix form :

$$\begin{bmatrix} X_1^2 & , X_1 X_n \\ \vdots & \vdots \\ X_n X_1 & , X_n^2 \end{bmatrix} \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_n \end{bmatrix} = \begin{bmatrix} YX_1 \\ \vdots \\ YX_n \end{bmatrix} \quad (1b.)$$

If the partial signal vectors $\{X_i\}$ are linear independent the $n \times n$ matrix of $X_i X_j$ has an inverse matrix and therefore :

$$\begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_n \end{bmatrix} = \begin{bmatrix} X_1^2 & , X_1 X_n \\ \vdots & \vdots \\ X_n X_1 & , X_n^2 \end{bmatrix}^{-1} \begin{bmatrix} YX_1 \\ \vdots \\ YX_n \end{bmatrix} \quad (2.)$$

Define R_{ij} and R_{yj}

$$R_{ij} = X_i X_j = \sum_{k=1}^n x_i(v_k) x_j(v_k)$$

$$R_{yj} = YX_j = \sum_{i=1}^n y_i x_j(v_i)$$

we have :

$$\begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_n \end{bmatrix} = \begin{bmatrix} R_{11} & , R_{1n} \\ \vdots & \vdots \\ R_{n1} & , R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{y1} \\ \vdots \\ R_{yn} \end{bmatrix} \quad (3.)$$

Eq.(3.) gives the Maximum Likelihood Estimation (MLE) for $C=c_1, \dots, c_n$ in the linear case defined in Eq.1.

In the following section the variance of the estimates is derived. Substituting into (1b.) the following equation and using the definitions above for R_{ij} and R_{yj} :

$$Y = \sum_j c_j^0 X_j + N$$

We have :

$$\begin{bmatrix} R_{11}, & , R_{1n} \\ \vdots & \vdots \\ R_{n1}, & , R_{nn} \end{bmatrix} \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_n \end{bmatrix} = \begin{bmatrix} R_{11}, & , R_{1n} \\ \vdots & \vdots \\ R_{n1}, & , R_{nn} \end{bmatrix} \begin{bmatrix} c_1^0 \\ \vdots \\ c_n^0 \end{bmatrix} + \begin{bmatrix} x_1(v_1), & , x_1(v_M) \\ \vdots & \vdots \\ x_n(v_1), & , x_n(v_M) \end{bmatrix} \begin{bmatrix} n_1 \\ \vdots \\ n_M \end{bmatrix}$$

Define the matrix of $\{R_{ij}\}$ as R we have :

$$R(\hat{C} - C^0) = \begin{bmatrix} x_1(v_1), & , x_1(v_M) \\ \vdots & \vdots \\ x_n(v_1), & , x_n(v_M) \end{bmatrix} \begin{bmatrix} n_1 \\ \vdots \\ n_M \end{bmatrix}$$

where $\hat{C}$ is the vector of the estimates and C^0 is the vector of the actual concentrations. Therefore the vector of differences $(\hat{C} - C^0)$ is given by :

$$(\hat{C} - C^0) = \begin{bmatrix} \hat{c}_1 - c_1^0 \\ \vdots \\ \hat{c}_n - c_n^0 \end{bmatrix} R^{-1} \begin{bmatrix} x_1(v_1), & , x_1(v_M) \\ \vdots & \vdots \\ x_n(v_1), & , x_n(v_M) \end{bmatrix} \begin{bmatrix} n_1 \\ \vdots \\ n_M \end{bmatrix}$$

In order to find the variance we perform the product :

$$(\hat{C} - C^0)(\hat{C} - C^0)',$$

which is equal to :

$$(\hat{C} - C^0)(\hat{C} - C^0)' = R^{-1} \begin{bmatrix} x_1(v_1), & x_1(v_M) \\ \vdots & \vdots \\ x_n(v_1), & x_n(v_M) \end{bmatrix} \begin{bmatrix} n_1 \\ \vdots \\ n_M \end{bmatrix} \begin{bmatrix} n_1, & n_M \end{bmatrix} \begin{bmatrix} x_1(v_1), & x_n(v_1) \\ \vdots & \vdots \\ x_1(v_M), & x_n(v_M) \end{bmatrix} (R^{-1})'$$

R is symmetric matrix and so is R^{-1} . Taking the mean value of both sides of the above equation and bearing in mind that the noise componenets are noncorrelated with constant power, we have :

$$\begin{aligned} \overline{(\hat{C} - C^0)(\hat{C} - C^0)'} &= R^{-1} \begin{bmatrix} x_1(v_1), & x_1(v_M) \\ \vdots & \vdots \\ x_n(v_1), & x_n(v_M) \end{bmatrix} \begin{bmatrix} N_0, & 0 \\ 0, & N_0 \end{bmatrix} \begin{bmatrix} x_1(v_1), & x_n(v_1) \\ \vdots & \vdots \\ x_1(v_M), & x_n(v_M) \end{bmatrix} (R^{-1})' \\ &= R^{-1} \begin{bmatrix} x_1^2 & , & x_1 x_n \\ x_n x_1, & x_n^2 \end{bmatrix} (R^{-1})' N_0 \end{aligned}$$

By our definition the matrix with $\{X_i X_j\}$ is the correlation matrix R.

We therefore have :

$$\begin{bmatrix} \overline{(\hat{c}_1 - c_1^0)^2} & , & \overline{(\hat{c}_1 - c_1^0)(\hat{c}_n - c_n^0)} \\ \overline{(\hat{c}_n - c_n^0)(\hat{c}_1 - c_1^0)} & , & \overline{(\hat{c}_n - c_n^0)^2} \end{bmatrix} = \begin{bmatrix} R_{11}, & R_{1n} \\ R_{n1}, & R_{nn} \end{bmatrix}^{-1} N_0 \quad (4.)$$

Example 1

Single Concentration

$$C = c$$

Using Eq.3, we have :

$$\hat{c} = \frac{R_{y1}}{R_{11}} = \frac{\sum y_i x_i}{\sum x_i^2}$$

The estimation variance is given by :

$$\overline{(\hat{c} - c^0)^2} = \frac{N_0}{R_{11}} = \frac{N_0}{\sum x_i^2}$$

The result is verified by direct calculations.

Example 2

Two concentrations

$$C = (c_1, c_2)$$

using equation 3., we have :

$$\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \frac{1}{R_{22}R_{11} - R_{12}^2} \begin{bmatrix} R_{22} & -R_{12} \\ -R_{21} & R_{11} \end{bmatrix} \begin{bmatrix} R_{y1} \\ R_{y2} \end{bmatrix}$$

The estimation variance is given by (see Eq.4.) :

$$\overline{(c_1 - c_1^0)^2} = \frac{N_0 R_{22}}{R_{22}R_{11} - R_{12}^2}$$

$$\overline{(\hat{c}_1 - c_1^0)^2} = \frac{N_0}{R_{11} \left(1 - \frac{R_{12}^2}{R_{22}R_{11}}\right)}$$

Define :

$$R_{ii} = S_i$$

$$\rho^2 = \frac{R_{12}^2}{R_{11}R_{22}}$$

we, finally, have :

$$\overline{(\hat{c}_1 - c_1^0)^2} = \frac{N_0}{S_1(1 - \rho^2)}$$

$$\overline{(\hat{c}_2 - c_2^0)^2} = \frac{N_0}{S_2(1 - \rho^2)}$$

Appendix 5

The formula for increasing the slit width

Given a discrete output signal of a spectrophotometer with a unit slit width ($M=1$) the output of a spectrophotometer with a slit width $L=2^n$ is given by (see Ch.3, Eq.15c) :

$$H_{2^n} = \frac{1}{2^{2n}} \frac{\delta_{2^n}^2}{\delta_1^2} = \frac{1}{2^n} \cosh(D)+1) (\cosh(2D)+1), \dots, \cosh(2^{n-1}D)+1) \quad (1.)$$

Eq.1 can be written also in a form :

$$\begin{aligned} H_{2^n} &= \frac{1}{2} H_{2^{n-1}} (\cosh(2^{n-1}D)+1) \\ H_{2^n} &= \frac{1}{2} H_{2^{n-1}} + \frac{1}{2} H_{2^{n-1}} \cosh(2^{n-1}D) \end{aligned} \quad (2.)$$

By using Eq.2 we find the coefficients $\{a_i\}$ that describe Eq.1 as a sum of operators :

$$H_{2^n} = \sum_{j=0}^N a_j \delta_{2^j}$$

Starting with $n=0$, we have (see Ex.1, Ch-3) :

$$H = 1$$

Taking $n=1$ and applying Eq.(2.), we have :

$$H_2 = 1/2 + (1/2)\cosh(D) = 1/2 + (1/4)\delta_2$$

Table 1. gives the coefficients for sum cases and the general case of increasing the slit width.

	M = 1	M = 2	M = 4	M = 8	M = 2 ⁿ
a ₀	1	2/4	4/16	8/64	2 ⁿ /2 ²ⁿ
a ₁		1/4	3/16	7/64	2 ⁿ⁻¹ /2 ²ⁿ
a ₂			2/16	6/64	
a ₃			1/16	5/64	
a ₄				4/64	
a ₅				3/64	
a ₆				2/64	
a ₇				1/64	
a _{2ⁿ-1}					1

Table 1 Coefficients of the operator that multiply the slit width by M

Appendix 6

Interval of confidence

Having estimation results x_1, x_N , each one with a variance σ^2 ,
the result :

$$\bar{x} = \frac{\sum x_i}{N}$$

has a variance σ'^2 of :

$$\sigma'^2 = \frac{\sigma^2}{N}$$

Using Tschebychev inequality [8], we have :

$$\Pr\{|\bar{x} - \bar{x}| < \epsilon\} > 1 - \frac{\sigma'^2}{\epsilon^2} = 1 - \frac{\sigma^2}{N\epsilon^2}$$

where $\bar{x}$ is the statistical mean of x .

Taking :

$$N = 10$$

$$\epsilon = \sigma$$

we have:

$$\Pr\{|\bar{x} - \bar{x}| < \epsilon\} > 1 - \frac{\sigma^2}{10\sigma^2} = .9$$

Thus by taking the average over ten trials we guarantee that 90 % of the results will be within a distance σ from the statistical mean of the estimate.