

# BIOLOGICAL MEMBRANES

Theory of transport, potentials and electric impulses

OVE STEN-KNUDSEN

*Professor Emeritus of Biophysics  
University of Copenhagen*



**CAMBRIDGE**  
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK  
40 West 20th Street, New York, NY 10011-4211, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa  
<http://www.cambridge.org>

© Cambridge University Press 2002

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Times 10/13 pt    *System* L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [TB]

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

*Library of Congress Cataloguing in Publication data*

Sten-Knudsen, Ove.

Biological membranes : mass transfer, membrane potentials, and electrical impulses /  
by Ove Sten-Knudsen.

p.      cm.

Includes bibliographical references and index.

ISBN 0 521 81018 3 (hb)

1. Neural conduction – Mathematical models.    2. Action potentials  
(Electrophysiology) – Mathematical models.    3. Mass transfer.    4. Biophysics.    I. Title.

QP363 .S775 2002

571.6'4 – dc21 2001052488

ISBN 0 521 81018 3 hardback

# Contents

<i>Foreword</i>	<i>page</i> xix
<i>Preface</i>	xxi
1 Mathematical prelude	1
1.1 Introduction	1
1.2 Basic concepts of differential calculus	2
1.2.1 Limits	2
1.2.2 Functions	2
1.2.3 The derivative	3
1.2.3.1 A few derived functions	6
1.2.4 Approximate value of the increment $\Delta y$	6
1.2.5 Differential	8
1.2.5.1 The chain rule	9
1.2.5.2 The derivative of the inverse function	10
1.3 Basic concepts of integral calculus	11
1.3.1 Definite and indefinite integral	12
1.3.2 The fundamental law	14
1.3.3 Evaluation of a definite integral	15
1.3.4 The mean value theorem	17
1.4 The natural logarithm	18
1.4.1 Definition of the natural logarithm	18
1.4.2 Elementary properties of the logarithm	19
1.4.2.1 Logarithm of a product	20
1.4.2.2 Logarithm of a quotient	21
1.4.2.3 Logarithm of an exponential	22
1.5 The exponential function	22
1.5.1 Definition of the exponential function	22
1.5.2 Derivative and integral	25

1.6	Taylor's theorem	26
1.6.1	Taylor and Maclaurin series	26
1.6.1.1	Expansion of a polynomial	26
1.6.1.2	Expansion of an arbitrary function	27
1.6.1.3	The binomial series	28
1.6.2	Series of the logarithmic and exponential functions	30
1.6.2.1	The logarithm	30
1.6.2.2	The exponential function	31
1.6.3	Approximate expressions of functions	32
1.6.4	Evaluation of an undetermined expression $0/0$	33
1.7	Basic techniques of integration	34
1.7.1	The method of substitution	34
1.7.2	Partial integration	36
1.8	Functions of several variables	38
1.8.1	Geometrical representation	40
1.8.2	Partial derivatives	40
1.8.3	Total differential	43
1.8.4	The chain rule once more	45
1.9	Some ordinary differential equations	49
1.9.1	Four first-order differential equations	50
1.9.1.1	The equation $y' + \alpha y = 0$	51
1.9.1.2	The equation $y' + \alpha y = K$	52
1.9.1.3	The equation $y' + \alpha y = Q(x)$	53
1.9.1.4	The equation $y' + P(x)y = Q(x)$	53
1.9.2	Two second-order differential equations	55
1.9.2.1	The equation $y'' + \kappa^2 y = 0$	55
1.9.2.2	The equation $y'' - \kappa^2 y = 0$	56
1.10	A note on partial differential equations	57
2	Migration and diffusion	62
2.1	Introduction	62
2.2	Flux	65
2.3	Types of passive transport	65
2.4	Migration	66
2.4.1	Friction coefficient and mobility	66
2.4.2	Migration flux	68
2.5	Diffusion	70
2.5.1	Phenomenological description	70
2.5.2	Diffusion flux (Fick's law)	73
2.5.2.1	The diffusion coefficient	74
2.5.2.2	A simple application of Fick's law	75

2.5.3	The diffusion equation	76
2.5.3.1	Diffusion with mass conservation	77
2.5.3.2	Diffusion with concurrent mass production	79
2.5.3.3	Classification of diffusion processes	79
2.5.4	Stationary diffusion processes	80
2.5.4.1	One-dimensional diffusion	81
(i)	Diffusion through a plate	81
(ii)	Diffusion through two adjoining, different media	84
(iii)	Unstirred layers	87
(iv)	Plate covered on one side by a membrane of permeability $P$	89
(v)	Diffusion with mass consumption	91
2.5.4.2	Diffusion in a cylinder with radial symmetry	99
(i)	The diffusion equation	100
(ii)	Diffusion through a cylindrical shell	103
(iii)	Diffusion in a cylinder with mass consumption	105
(iv)	Diffusion from a cylinder into the surrounding medium with mass consumption (Krogh's cylinder)	108
2.5.4.3	Diffusion with radial symmetry in a sphere	111
(i)	The diffusion equation	111
(ii)	Diffusion through a spherical shell	114
(iii)	Sphere covered by a thin membrane, mass consumption in the interior	115
2.5.5	Time-dependent diffusion processes	117
2.5.5.1	An extended initial distribution (Boltzmann's trick)	117
2.5.5.2	Diffusion from a region with constant concentration	123
2.5.5.3	Duhamel's integral	124
2.5.5.4	An instantaneous surface distribution	127
2.5.5.5	Green's function	130
(i)	A varying initial distribution in space	130
(ii)	Initial uniform distribution in the infinite half-space	133
(iii)	The effect of an impermeable barrier	133
(iv)	The effect of a matter-absorbing wall	135
(v)	A variable flux into one half-space	137
2.5.6	Molecular description of diffusion	139
2.5.6.1	Brownian motion	140

2.5.6.2	Diffusion from a statistical point of view	143
2.5.6.3	Random walk	145
(i)	The distribution function	146
(ii)	The mean displacement	153
(iii)	The mean displacement in one direction	154
(iv)	The root mean square displacement (the Einstein–Smoluchowski relation)	154
( $\alpha$ )	Two-dimensional random walk	157
( $\beta$ )	Three-dimensional random walk	159
2.5.6.4	Random walk and Fick’s law	161
(i)	Einstein’s simplified treatment	161
(ii)	A more exact derivation of Fick’s law	164
2.5.6.5	Random walk and the diffusion equation	169
2.5.6.6	Random walk over an energy barrier	175
2.6	Diffusion and migration superimposed	180
2.6.1	The Smoluchowski equation	180
2.6.1.1	An instantaneous plane source in infinite space	182
(i)	The concentration change with time at a fixed point in space	184
(ii)	Driving the swarm towards a reflecting barrier: a case of sedimentation	184
2.6.2	“Random walk” considerations	190
2.6.2.1	The flux equation	190
2.6.2.2	Random walk and the diffusion–migration equation	193
2.6.2.3	Migration over an energy barrier	195
2.6.3	Kramers’ equation	199
2.6.4	Diffusion coefficient and mobility	201
2.6.4.1	The Einstein relation	202
2.6.4.2	Einstein–Stokes relation	203
2.6.4.3	The “driving force” behind the diffusion process	206
2.7	Diffusion through membranes	208
2.7.1	Permeability coefficient	209
2.7.2	Kinetics of exchange	210
2.7.2.1	Outer concentration kept at zero	211
2.7.2.2	Outer concentration kept constant: cell concentration initially zero	212
2.7.2.3	Both phases comparable in size	213
2.7.3	Compartment analysis	218

2.7.3.1	Transport with passive membrane permeabilities	218
(i)	A step change in outer concentration	219
(ii)	Outer concentration grows asymptotically	220
2.7.3.2	One-way transport	221
(i)	Unidirectional flux	221
(ii)	Unidirectional transfer	222
(iii)	Passive influx and unidirected efflux	224
2.7.4	Stationary diffusion with superimposed migration	224
2.7.4.1	Determination of the flux	225
2.7.4.2	Unidirectional fluxes and flux ratio	228
2.7.4.3	Concentration profile	229
2.8	Convective and osmotic water movement through membranes	230
2.8.1	Convective water movement	231
2.8.2	Osmotic water movement	233
2.8.2.1	Osmotic pressure	233
2.8.2.2	Colligative properties	235
2.8.2.3	The underlying mechanism of osmotic water movement	236
2.8.2.4	The equation for the osmotic pressure	237
2.8.2.5	Osmotic coefficient	242
2.8.2.6	A simple dynamic model of osmotic equilibrium	243
2.8.3	The freezing-point depression	250
2.8.3.1	The freezing-point depression and osmotic pressure	251
2.8.3.2	Osmolarity	252
2.8.3.3	Reflection coefficient	254
2.8.4	Water movement across cell membranes	255
3	Membrane potentials	259
3.1	Introduction	259
3.2	Electric field and potential	260
3.3	Transport of ions in solutions	265
3.3.1	Migration	266
3.3.2	Electrodiffusion (Nernst–Planck equations)	271
3.3.2.1	Equivalent forms	272
3.3.2.2	Poisson’s equation	274
3.3.2.3	Electroneutrality	276
3.3.2.4	The constant field	277
3.4	The equilibrium potential	277

3.4.1	A qualitative description of the origin of the membrane potential across an ion-selective permeable membrane	278
3.4.2	The Nernst equation	281
3.4.2.1	The charge density of the excess charges on the two membrane sides	283
3.4.2.2	Derivation of Nernst's equation	283
3.4.3	Establishing the electric contact to the electrolyte solution: electrodes	286
3.4.3.1	The galvanic cell	286
(i)	Half-cells	288
(ii)	Electrode potentials	289
(iii)	Standard electrode potentials	294
(iv)	Non-equivalent electrode current	295
(v)	Reversibility	298
3.4.3.2	Two recording electrodes	299
(i)	The silver–silver chloride electrode	299
(ii)	The calomel electrode	301
3.4.4	The equivalent electric circuit for the ionic-selective membrane	302
3.4.4.1	Measurement of the current–voltage characteristic	302
3.4.4.2	Membrane current and membrane conductance	304
3.4.4.3	The equivalent circuit diagram	308
3.4.4.4	Membrane conductance and membrane permeability	311
3.5	The Donnan potential	313
3.5.1	Qualitative description of the Donnan distribution	314
3.5.2	Quantitative treatment of the Donnan system	316
3.5.2.1	Low polyelectrolyte concentration	320
3.5.2.2	High polyelectrolyte concentration	321
3.5.3	Concentration and potential profiles	322
3.5.3.1	The Poisson–Boltzmann equations	324
3.5.3.2	Solving the Poisson–Boltzmann equations	326
(i)	The solution for $x \geq 0$	326
(ii)	The solution for $x \leq 0$	329
(iii)	Binding the two solutions together	331
3.5.3.3	A numerical example	333
3.5.3.4	The total space charge	335
3.6	Diffusion potentials	336

3.6.1	Qualitative description of the diffusion potential	336
3.6.1.1	Collapse of the Donnan regime	336
3.6.1.2	A binary electrolyte	337
3.6.1.3	The salt bridge to eliminate the diffusion potential	338
3.6.2	Calculation of the diffusion potential for a binary monovalent electrolyte	339
3.6.3	Diffusion potential between solutions of different composition	341
3.6.3.1	The Planck regime	342
(i)	Planck's general relations	343
(ii)	The electrical equivalent for the Planck regime	345
(iii)	Planck's expression for the diffusion potential	348
3.6.3.2	The Henderson regime	348
3.6.3.3	The salt bridge once again	351
3.7	Electrodifusion through membranes	353
3.7.1	A single salt	353
3.7.1.1	The diffusion potential	354
3.7.1.2	The membrane resistance	356
3.7.1.3	The potential profile	357
3.7.1.4	The equivalent electric circuit	357
3.7.1.5	Electroneutrality	359
3.7.2	Ion-selective membranes	360
3.7.3	The Goldman regime	363
3.7.3.1	Derivation of the Goldman equation	365
3.7.3.2	Concentration profiles	370
3.7.3.3	Membrane conductance and membrane permeability	372
(i)	Concentrations equal on both sides	373
(ii)	Different surrounding concentrations: $V \approx V_j^{(eq)}$	374
(iii)	$V \neq V_j^{(eq)}$	376
3.7.3.4	Total current and membrane potential	377
3.7.4	The mosaic membrane (the Millman equation)	379
3.8	The membrane potential of a biological cell	384
3.8.1	Measuring the membrane potential	385
3.8.2	The origin of the membrane potential	388
3.8.3	Membrane potential and ionic concentrations in the extracellular medium	391
3.8.3.1	Sudden changes of both $V_K^{(eq)}$ and $V_{Cl}^{(eq)}$	392
3.8.3.2	Membrane potential with varying $K^+$ in extracellular fluid in the absence of $Cl^-$	396

3.8.3.3	Membrane potential with varying $\text{Na}^+$ in the extracellular fluid	398
3.8.4	Membrane potential and active $\text{Na}^+/\text{K}^+$ transport	399
3.8.4.1	Absolute stationarity	400
3.8.4.2	Total current is zero	402
3.8.4.3	The mosaic membrane	403
3.9	Flux ratio analysis	403
3.9.1	Flux ratio and electrodiffusion	406
3.9.2	Flux ratio and convective diffusion	409
4	The nerve impulse	411
4.1	Introduction	411
4.1.1	Excitability	411
4.1.2	The communication system	412
4.2	Historical background	413
4.3	The nerve signal recorded with external electrodes	417
4.3.1	The nerve signal	417
4.3.1.1	The diphasic action potential	419
4.3.1.2	The monophasic action potential	420
4.3.2	Some elementary properties of the nerve signal	422
4.3.2.1	“All or nothing” law	423
4.3.2.2	Subliminal stimuli	424
4.3.2.3	The refractory period	426
4.4	Results from the giant axon of the squid	427
4.4.1	Recording the resting membrane potential and action potential	427
4.4.2	The resting membrane potential	431
4.4.2.1	Membrane potential and the Goldman regime	432
4.4.2.2	Mobility of $\text{K}^+$ ions in axoplasm	433
4.4.3	The action potential	437
4.4.3.1	Stimulation with an intracellular microelectrode	437
(i)	The internal electrode is a cathode	437
(ii)	The internal electrode is an anode	438
4.4.3.2	Conductance changes attending the action potential	441
4.4.3.3	The effect of extracellular sodium concentration on the action potential	442
4.4.3.4	Perfused axons	443
4.4.4	The experimental substantiation of the sodium hypothesis	447
4.4.4.1	Net movement of radioactive sodium and potassium during electric activity	447

4.4.4.2	The temporal resolution of the separate ionic current: voltage clamp technique	449
(i)	The control system	450
(ii)	Identification of the early inward-directed current	454
(iii)	Identification of the outward-directed current	459
4.4.4.3	Calculation of the partial membrane conductances for sodium and potassium	461
(i)	The time course of the changes in sodium and potassium conductance during a voltage clamp	461
(ii)	The dependence of the conductance changes on the displacement of the membrane potential	463
(iii)	The inactivation process	464
4.4.4.4	The membrane action potential: qualitative synthesis of the voltage clamp experiments	470
(i)	The action potential	472
(ii)	Threshold phenomena	473
4.4.4.5	Investigations on single $\text{Na}^+$ channels	474
4.4.4.6	Selective effects on the $\text{Na}^+$ and $\text{K}^+$ channels	479
(i)	Blocking the $\text{Na}^+$ channel	480
(ii)	Blocking the $\text{K}^+$ channel	481
(iii)	Destruction of $\text{Na}^+$ inactivation	483
4.4.5	The propagated action potential	484
4.4.5.1	The local current loops	484
4.4.5.2	The nerve as an electric cable	486
(i)	Derivation of the cable equation	486
(ii)	The stationary state	491
(iii)	Time-dependent solutions	492
4.4.5.3	Reconstruction of the action potential	500
(i)	The Hodgkin–Huxley equations	501
(ii)	The propagated action potential	503
4.4.5.4	Nerve impulse recorded with external electrodes	506
4.4.5.5	The conduction velocity	510
4.5	Myelinated nerves	511
4.6	Repetitive impulse transmission	516
4.7	The Hodgkin–Huxley equations	520
4.7.1	Empirical equations for the sodium and potassium conductances	520
4.7.2	Voltage clamp	522
4.7.3	The membrane action potential	528
4.7.4	The propagated action potential	532

5	Impulse transmission across cell boundaries	537
5.1	General characteristics of impulse transmission	537
5.1.1	The synapse	537
5.1.2	Mechanisms of transmission	539
5.2	Neuromuscular transmission: impulse transmission from nerve to muscle	542
5.2.1	Structure of the end-plate	542
5.2.2	The course of the electrical transmission	544
5.2.2.1	The synaptic delay	546
5.2.2.2	The postsynaptic potential	546
5.2.3	The end-plate potential	548
5.2.3.1	End-plate potential and acetylcholine	549
5.2.3.2	The time course of the end-plate potential	553
5.2.3.3	Cholinesterase and anti-cholinesterase	554
5.2.3.4	Ionic movements associated with the course of the end-plate potential	556
5.2.4	Quantal release of acetylcholine	565
5.2.4.1	Miniature end-plate potentials	565
5.2.4.2	End-plate potential and miniature potentials	571
5.2.4.3	The vesicle hypothesis	582
Appendix A	About the functions $\text{Erf}\{x\}$ , $\text{Erfc}\{x\}$ and calculation of the integral $\int_{-\infty}^{\infty} e^{-x^2} dx$	584
Appendix B	Solving the integral $\int_0^t \frac{1}{\sqrt{u}} e^{-x^2/4Du} du$	588
Appendix C	Evaluation of $\frac{N!}{[(N+m)/2]! [(N-m)/2]!} \left(\frac{1}{2}\right)^N = \sqrt{\frac{2}{\pi N}} e^{-m^2/2N}$	590
Appendix D	To demonstrate that the mean value $\langle \xi^2 \rangle$ of the displacements squared is proportional to the number $N$ of the displacements	593
Appendix E	Evaluation of the integral Eq. (2.5.191)	595
Appendix F	Evaluation of the integral $\int_{-\infty}^x e^{mX - (X+x_0)^2/B} dX$	599
Appendix G	Example of the application of the theory of Brownian motion	601
Appendix H	A note on the physical meaning of the pressure gradient	604
Appendix I	About hyperbolic functions	606
Appendix J	Evaluation of an integral of Duhamel's type	612
Appendix K	Calculation of the potential profiles resulting from injection of a constant current $I_0$ in the axon	616
Appendix L	A note on the method of images	618
Appendix M	Cable analysis of the end-plate potential (Fatt & Katz, 1951)	623

*Contents*

xvii

Appendix N	Measuring the electric parameters in a spherical cell	628
Appendix O	Measuring the electric parameters in a cylindrical cell	631
Appendix P	Electric parameters for a cylindrical unmyelinated axon	638
Appendix Q	Surface recorded action potential and membrane action potential	641
Appendix R	About concentration scales	646
Appendix S	Units and physical constants	648
	<i>List of symbols</i>	650
	<i>References</i>	655
	<i>Index</i>	662

# Chapter 1

## Mathematical prelude

For more than two thousand years some familiarity with mathematics has been regarded as an indispensable part of the intellectual equipment of every cultured person.

*(Richard Courant, 1941)*

### 1.1 Introduction

In biological research there is a steadily increasing trend to describe functions and mechanisms *quantitatively* by applying ideas and concepts from physics and physical chemistry. This tendency is found in large areas of biology, extending from ecology over the function of the integrated organism to processes taking place at the cellular and molecular level. This development will doubtless continue.

However, a quantitative treatment of any phenomenon in physics or physical chemistry requires an adequate command of the mathematical tools that are needed to formulate and solve the particular problem that is subject to such close scrutiny. For that reason, mastery of certain elements of mathematical analysis is an indispensable element in the arsenal of tools that are loaded into the knapsack of the serious student of general physiology or cell biology.

The sections that follow in this chapter are not presented as a self-contained mathematical text. The intention is to present a summary – short in some places, more detailed in others – of the mathematical concepts and techniques that are used in this book. It is presumed that the reader is already familiar with these concepts. Thus, a cursory reading of this chapter may have the effect of acting as a reminder of items that are known but perhaps not immediately recalled from memory.

**1.2 Basic concepts of differential calculus****1.2.1 Limits**

A collection of numbers

$$a_1; a_2; a_3; a_4; \dots a_n;$$

that follow each other according to a given law is called a *sequence* of numbers. If the number of elements  $n$  increases without bound the sequence is an *infinite* sequence. The elements of the sequence are said to *converge* to a *limit*  $L$  if the elements beyond that of  $a_\mu$  behave in such a way that the difference

$$|L - a_n| \quad \text{for } n > \mu$$

is smaller than any arbitrarily small positive number  $\varepsilon$ . If the elements  $a_n$  do not pile up in this manner, the sequence is made up of elements that *diverge*. When the elements of a sequence are added they constitute a *series*

$$S_n = a_1 + a_2 + a_3 + a_4 + \dots a_n,$$

which may be *finite* or *infinite* according to whether the number of elements  $n$  is bounded or not. An infinite series may converge to a definite value  $S_n$  when  $n$  increases beyond the boundary. This value  $S_\infty = L$  is called the *limit* of the series. This is generally written as

$$S_n \rightarrow L, \quad \text{for } n \rightarrow \infty, \quad \text{or} \quad \lim_{n \rightarrow \infty} S_n = L.$$

**1.2.2 Functions**

Let  $x$  and  $y$  represent two arbitrary quantities that are coupled together in such a way that to each value of  $x$  there exists a definite value of  $y$ . We say then that the quantity  $y$  is a *function* of the quantity  $x$ . Usually this is represented as

$$y = f(x), \tag{1.2.1}$$

where  $x$  is called the *independent variable* and  $y$  is called the *dependent variable*\*. Of course one could equally well have considered the *inverse function*

$$x = g(y), \tag{1.2.2}$$

where  $y$  is now the independent variable and  $x$  is the dependent variable. The condition that the inverse function  $x = g(y)$  is so well-behaved that there exists in the interval  $a \leq x \leq b$  one and only one value of  $x$  for a given value of  $y$ , is

\* To facilitate the readability of this text, mathematical and physical variable quantities are printed in *italics*. Similarly, mathematical operators are printed in Roman type.

that the function  $y = f(x)$  is increasing or decreasing *monotonically* in the same domain. Thus, the function  $y = x^2$  is monotonically decreasing in the region  $-a \leq x \leq 0$ , and to every value of  $y$  there corresponds only one value  $x = -\sqrt{y}$ . In the region  $0 \leq x \leq a$  the function  $y = x^2$  increases monotonically, and to every value of  $y$  there corresponds likewise only one value  $x = \sqrt{y}$ . With increasing values for  $x$  in the region  $-a \leq x \leq a$  the function  $y = x^2$  both decreases monotonically as well as increasing, and for a given value of  $y$  we have the corresponding values  $x = -\sqrt{y}$  and  $x = \sqrt{y}$ . A function that suddenly *jumps* from one value to another is said to be a *discontinuous function*. Thus, the function

$$y = f(x) = \begin{cases} 2 & \text{for } x \geq 1 \\ 1 & \text{for } x < 1 \end{cases}$$

is a discontinuous function for  $x = 1$ , since

$$f(1 + \varepsilon) - f(1 - \varepsilon) = 1$$

no matter how small we make the positive quantity  $\varepsilon$ . A *continuous function* is, roughly speaking, a function that does not do such things. Thus, the function

$$y = f(x) = \begin{cases} x^2 & \text{for } x \geq 1 \\ x & \text{for } x \leq 1 \end{cases}$$

is continuous at the point  $x = 1$  since

$$f(1 + \varepsilon) - f(1 - \varepsilon) = (1 + \varepsilon)^2 - (1 - \varepsilon) = 3\varepsilon + \varepsilon^2 \rightarrow 0 \quad \text{for } \varepsilon \rightarrow 0,$$

although the formula displays changes for  $x = 1$ .

### 1.2.3 The derivative

Consider the function  $y = f(x)$  that is continuous in the range  $a < x < b$ . If the quantity, denoted the *difference quotient*, for the function  $y = f(x)$  at the point  $x$

$$\frac{f(x + h) - f(x)}{h}, \quad (1.2.3)$$

converges towards a *definite limit* as  $h$  approaches zero in an arbitrary manner 0, the value of this limit

$$\lim_{h \rightarrow 0} \left[ \frac{f(x + h) - f(x)}{h} \right] \stackrel{\text{def}}{=} f'(x), \quad (1.2.4)$$

is called the *first derivative* of the function  $y = f(x)$ \*. Another name for  $f'(x)$  is the *differential quotient* of  $f(x)$ . We can illustrate this limiting process geometrically as follows: Eq. (1.2.3) represents the value of the slope of a straight line that is anchored at the curve point  $P_0$  with coordinates  $(x, f(x))$  and makes another section with the curve at the point  $P_1$  at  $(x + h, f(x + h))$ . This line is called a *secant* to the curve. When we let  $h$  decrease in an arbitrary manner, the point  $P_1$  approaches the point  $P_0$  from either side according to the sign of  $h$ , and when  $h \rightarrow 0$  the slope of the secant attains a limiting value that is equal to the slope of the line that, at the point  $P_0$ , has only one point in common with the curve  $y = f(x)$ , namely the *tangent* of the curve at  $P_0$ , or

$$\lim_{P_1 \rightarrow P_0} (\text{Slope of secant anchored at } P_0) = (\text{Slope of tangent at } P_0)$$

always provided there is a tangent with a well-defined direction at the point  $P_0$  on the curve. This occurs if the limit of the ratio  $(f(x + h) - f(x))/h$  in Eq. (1.2.4) converges to the definite value  $f'(x)$  when  $h \rightarrow 0$ . In many physical applications involving the derivative it may be useful to keep in mind this geometrical representation of  $f'(x)$ .

The expression  $y' = f'(x)$  goes back to the work of J.-L. Lagrange†. Another way of writing the derivative  $f'(x)$  is

$$f'(x) \stackrel{\text{def}}{=} \frac{dy}{dx}, \quad (1.2.5)$$

which was introduced by G.W. Leibniz (1646–1716)‡, has many practical advantages, and is almost always used in applied mathematics.

The quantity  $(dy/dx)$  is not a fraction in the usual sense but a compact *symbol* meaning that the function  $y = f(x)$  has been subjected to the operation that is defined by Eq. (1.2.4). To emphasize the character of  $dy/dx$  as a mathematical operation many people prefer to use the typographical convention

$$\frac{dy}{dx} \stackrel{\text{def}}{=} \frac{d y}{d x}, \quad (1.2.6)$$

to distract one's thoughts from a fraction. This notation will be used in this book.

\* The symbol  $\stackrel{\text{def}}{=}$  is used in this text to emphasize that it is a definition.

† J.-L. Lagrange (1736–1813) was a Professor at École Polytechnique. He was one of the greatest mathematicians of the eighteenth century, who made fundamental contributions to the development of differential and integral calculus, calculus of variation, theory of numbers and to mechanics (*Mécanique analytique*) and astronomy.

‡ This is a remainder of the derivative being obtained from the difference quotient which he wrote as

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{\Delta y}{\Delta x}, \quad \text{for } \Delta x \rightarrow 0.$$

As an illustration we consider the function  $y = f(x) = x^2$ . We have

$$\frac{(x+h)^2 - x^2}{h} = \frac{(x^2 + 2hx + h^2) - x^2}{h} = \frac{2hx + h^2}{h} = 2x + h.$$

Hence

$$\lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = 2x.$$

Thus, the limit exists, giving

$$f'(x) = \frac{dy}{dx} = 2x.$$

Continuing this argument to  $y = f(x) = x^n$ , where  $n$  is any real number, one gets

$$\frac{d}{dx}(x^n) = n x^{n-1}.$$

Naturally the operations of Eq. (1.2.3) and Eq. (1.2.4) can be applied to the function  $f'(x)$ . If the limit exists it is called the *second derivative* of the function  $f(x)$ . The notation for this limit is

$$f''(x) \stackrel{\text{def}}{=} \frac{d}{dx} \left( \frac{dy}{dx} \right) \stackrel{\text{def}}{=} \frac{d^2 y}{dx^2}. \quad (1.2.7)$$

Some mathematicians have never become reconciled to Leibniz's notation and have instead replaced the operator  $d(\ )/dx$  by the symbol  $D$  to denote the operation\*

$$D f(x) \stackrel{\text{def}}{=} \lim_{h \rightarrow 0} \left[ \frac{f(x+h) - f(x)}{h} \right] \stackrel{\text{def}}{=} f'(x).$$

The  $D$  notation will not be used in this text.

The requirement for the limit of Eq. (1.2.4) to exist is that the function  $f(x)$  is *continuous*. However, this condition is not sufficient, because a continuous function may exhibit a sudden break at a point  $x_0$ . In this case  $f'(x_0 - \varepsilon)$  and  $f'(x_0 + \varepsilon)$  both exist no matter how small we make  $\varepsilon$ , but they may differ drastically from each other in value, leaving  $f'(x)$  to have a discontinuity at the point  $x_0$ .

\* This was introduced in 1808 by Brisson and gained a footing owing to the extensive use of the operator  $D$  made by A.L. Cauchy (1789–1857).

## 1.2.3.1 A few derived functions

Using the operations that are defined by Eq. (1.2.4) on the elementary mathematical functions one obtains explicit expressions for the derivatives of the functions in question. Below are a few important elementary examples\*

- (a) If  $f(x) = A$ , where  $A$  is a constant,  $f'(x) = 0$ .
- (b) If  $f(x) = Au(x)$ ,  $f'(x) = Au'(x)$ .
- (c) If  $f(x) = u(x) + v(x)$ ,  $f'(x) = u'(x) + v'(x)$ .
- (d) If  $f(x) = u(x)v(x)$ ,  $f'(x) = u'(x)v(x) + u(x)v'(x)$ .
- (e) If  $f(x) = \frac{u(x)}{v(x)}$ ,  $f'(x) = \frac{u'(x)v(x) - u(x)v'(x)}{v(x)^2}$ .
- (f) If  $f(x) = x$ ,  $f'(x) = 1$ .
- (g) If  $f(x) = x^n$ ,  $f'(x) = nx^{n-1}$ .
- (h) If  $f(x) = \sin x$ ,  $f'(x) = \cos x$ .
- (i) If  $f(x) = \cos x$ ,  $f'(x) = -\sin x$ .
- (j) If  $f(x) = \tan x$ ,  $f'(x) = 1/\cos^2 x$ .

1.2.4 Approximate value of the increment  $\Delta y$ 

In physics many relations are described in terms of the *rate* of change of a quantity. This change may depend upon time, position in space, or both. With hardly a single exception it is sufficient initially to express this change with an approximate accuracy that may be improved later as occasion requires. In this context, differential calculus is a very useful tool. One proceeds as follows. The curve in Fig. 1.1 shows an arbitrary differentiable function  $y = f(x)$ . The line AB denotes the tangent to the curve on the point  $(x, y)$  having a slope that is equal to the value of the derivative  $f'(x)$  taken at the point  $(x, y)$ . Let  $x + h$  be a neighboring point to  $x$  that corresponds to assigning a finite increment  $h = \Delta x$  to the value  $x$  of the independent variable. We denote the value of the function at the neighboring point  $x + h$  as  $f(x + h) = y + \Delta y$ , where  $\Delta y$  is the increment in  $y = f(x)$  due to the change  $h$  in the argument. According to Eq. (1.2.3) and Eq. (1.2.4), that defines the derivative  $f'(x)$ , the increment can be written as

$$\Delta y = f(x + h) - f(x) = f'(x)h + \varepsilon\Delta x, \quad (1.2.8)$$

or

$$y + \Delta y = f(x + \Delta x) = f(x) + f'(x)\Delta x + \varepsilon\Delta x, \quad (1.2.9)$$

\* For more about hyperbolic functions, see Appendix I.

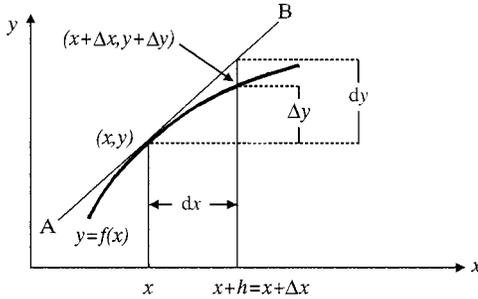


Fig. 1.1. Approximation of the increment  $\Delta y$  of a function  $y = f(x)$  by a linear function. The figure also illustrates the geometrical meaning of the differentials  $dy$  and  $dx$ .

where  $\varepsilon = \varepsilon(\Delta x)$  depends on the magnitude of  $\Delta x$  and approaches zero when  $h = \Delta x \rightarrow 0$ .

We now regard the variable  $x$  as fixed and let the increment  $h = \Delta x$  vary in an arbitrary manner. Equation (1.2.9) now states that the increment  $\Delta y$  to the value  $y$  of  $f(x)$  at a given value of  $x$  is made up of two terms:

- (i) a term  $f'(x)h = f'(x)\Delta x$  that is proportional to the increment  $h = \Delta x$  with  $f'(x)$  as the proportionality coefficient that is a constant at a fixed value of  $x$ , and
- (ii) a correction term  $\varepsilon h = \varepsilon \Delta x$ , which can be made as small as we wish relative to  $h$  by making the increment  $h = \Delta x$  sufficiently small. Thus, the smaller we make the interval in question  $h = \Delta x$  around  $x$  the more precisely will the function  $f(x + h)$ , being a function of  $h$ , be represented by its linear part

$$f(x + h) \approx f(x) + f'(x)h, \quad (1.2.10)$$

where both  $f(x)$  and  $f'(x)$  are two fixed numbers for a given value of  $x$ . From a geometrical viewpoint this approximate description of the value  $f(x + h)$  of the function  $y = f(x)$  at the point  $(x, y)$  means that the curve of  $f(x)$  is *replaced* by the tangent and that the expression for the increment of the function

$$\Delta y = \Delta f = f(x + h) - f(x),$$

corresponding to the increment  $\Delta x$  of the independent variable, can be written approximately as

$$\Delta y = \Delta f \approx f'(x)\Delta x, \quad (1.2.11)$$

provided  $\Delta x$  is sufficiently small to make the term  $\varepsilon \Delta x$  negligible relative to the term  $f'(x)\Delta x$ .

## 1.2.5 Differential

The approximate description of the increment  $\Delta y$  by the linear part  $f'(x)h = f'(x)\Delta x$  can also be used to put the term *differential* on a firmer logical basis. The original meaning of differentials as infinitely small quantities – different from zero – very soon turned out to have no precise meaning. One of the founders of differential calculus G.W. Leibniz (1646–1716) tried, without success, around 1680 to define the differential quotient as the ratio between two infinitely small increments  $dy$  and  $dx$  that were considered just before both quantities assumed the value zero. More than 100 years passed before the Bohemian priest B. Bolzano (in 1817) sharpened the definitions of such concepts as limits, continuity, etc., and then described the derivative by the limiting process in Eq. (1.2.4). However, Leibniz's notation has turned out to be the most suitable for handling calculations in physics and chemistry. For that reason, it is of value to attempt to give an unambiguous description of the identity

$$f'(x) \stackrel{\text{def}}{=} \frac{dy}{dx},$$

in such a way that the expression  $dy/dx$  need not be regarded only as a symbol for the limiting process

$$\frac{dy}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h},$$

but can also be considered as a quotient between two actual, well-defined, quantities.

Starting from the definition of the derivative  $f'(x)$  as a limiting process, as in Eq. (1.2.4), we then assign a fixed value to the independent variable  $x$  and consider the increment  $h = \Delta x$  as the variable (see Fig. 1.1). The quantity  $h = \Delta x$  is then called the *differential* of  $x$ , and is designated as  $dx$ . We then *define* the quantity

$$dy \stackrel{\text{def}}{=} f'(x) dx, \quad (1.2.12)$$

as the *differential*  $dy$  of the function  $y = f(x)$  corresponding to the differential  $dx$  of the independent variable. Thus, by means of this definition the derivative  $f'(x)$  is regarded as the ratio between two quantities  $dy$  and  $dx$ , which can have any value provided their ratio is constant and equal to  $f'(x)$ . Comparing Eq. (1.2.9) with Eq. (1.2.10) shows that the differential  $dy$  is equal to the linear portion of the increment  $\Delta y$  that corresponds to the increment  $dx$  of the independent variable  $x$  (compare Fig. 1.1).

The introduction of the differentials  $dy$  and  $dx$  due to S.-F. Lacroix (1765–1843) and A.L. Cauchy (1789–1857) does not represent a new idea. But their

merit is to make more precise the wording of “infinitesimal quantity”: these quantities are now of finite magnitude, and not quantities “just differing from zero”. Hence, when considering a particular problem, they may be chosen to be small enough so that one can, with confidence, replace the increment  $\Delta y$  of the function with its differential  $dy$  and write

$$\Delta y \approx dy = f'(x) dx = \left( \frac{dy}{dx} \right) dx, \quad (1.2.13)$$

and

$$f(x + dx) \approx f(x) + f'(x) dx = f(x) + \left( \frac{dy}{dx} \right) dx. \quad (1.2.14)$$

The validity of the above approximation depends on the special character of the physical situation in question. In general, the error introduced will be insignificant for the solution of the physical problem as long the infinitesimal quantities introduced are smaller than the actual error of measurement that are related to the physical situation.

### 1.2.5.1 The chain rule

One often finds that the dependent variable  $y$  is a function of the independent variable  $u$  that again is a function of the independent variable  $x$ , e.g.

$$y = u^3 \quad \text{and} \quad u = \sin x.$$

This situation is described by saying that  $y$  is a *function of a function* or that  $y$  is a *compound function* of  $x$ . In general we write this as

$$y = f(x) = F(u) = F\{u(x)\}.$$

If both derivatives

$$\frac{dF}{du} \quad \text{and} \quad \frac{du}{dx}$$

exist it can be shown that

$$f'(x) = F'(u) u'(x),$$

or, in terms of Leibniz's notation,

$$\frac{dy}{dx} = \frac{dF}{du} = \frac{dF}{du} \frac{du}{dx}, \quad (1.2.15)$$

which illustrates both the flexibility and suggestive strength of this notation. It appears as if the symbols  $dy$  and  $dx$  are quantities that can be considered

and manipulated as if they were real numerical quantities. In fact, they can. According to Eq. (1.1.10) we have

$$dF = \frac{dF}{du} du, \quad \text{and} \quad du = \frac{du}{dx} dx,$$

so that

$$dF = \frac{dF}{du} \frac{du}{dx} dx,$$

which on division on both sides by  $dx$  becomes Eq. (1.2.15). In the above example we have  $dy/du = 3u^2$  and  $du/dx = \cos x$ . Hence

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx} = 3 \sin^2 x \cos x.$$

For the function  $y = \sin^3 \alpha x$  we obtain  $dy/dx = 3\alpha \sin^2 \alpha x \cos \alpha x$ , since

$$\frac{d(\sin \alpha x)}{dx} = \frac{d(\sin \alpha x)}{d(\alpha x)} \frac{d(\alpha x)}{dx} = \alpha \cos \alpha x.$$

If  $y = \sin \sqrt{x} = \sin u$ , where  $u = \sqrt{x} = x^{1/2}$  we have

$$\frac{dy}{dx} = \frac{d \sin u}{du} \frac{du}{dx} = \cos \sqrt{x} \frac{d}{dx} (\sqrt{x}) = \cos \sqrt{x} \left( \frac{1}{2} \right) x^{-\frac{1}{2}} = \frac{1}{2} \frac{\cos \sqrt{x}}{\sqrt{x}}.$$

### 1.2.5.2 The derivative of the inverse function

It has previously been stated that if a continuous function  $y = f(x)$  is either increasing or decreasing monotonically in an interval (say  $a \leq x \leq b$ ) then the *inverse* function  $x = g(y)$  also exists as a single-valued function that is continuous and monotonic in the same interval. If the function  $y = f(x)$  is differentiable in the interval, the function increases monotonically if  $f'(x) > 0$  in the interval and, correspondingly, can decrease monotonically if  $f'(x) < 0$ . Knowledge of the differentiability of a function in a given interval provides a tool for deciding whether the function also possesses an unambiguous inverse function as expressed in the following statement.

*If the function  $y = f(x)$  is differentiable in the interval  $a < x < b$  and  $f'(x) > 0$  everywhere or  $f'(x) < 0$  everywhere, then the inverse function  $x = g(y)$  also has a derivative  $x' = g'(y)$  in the whole interval. The derivative of the original function  $y = f(x)$  and that of the inverse function  $x = g(y)$  are for the values of  $x$  and  $y$  belonging together connected by the following relation:*

$$f'(x) \cdot g'(y) = 1, \quad (1.2.16)$$

or written in the form

$$\frac{dy}{dx} = \frac{1}{(dx/dy)}. \quad (1.2.17)$$

This is demonstrated by applying the definition for the derivative (Eq. (1.2.4)) on  $y = f(x)$  and its inverse function  $x = g(y)$ . For the function  $y = f(x)$  we have

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The numerator is written as  $f(x + \Delta x) - f(x) = \Delta y$ . As  $x = g(y)$  we can express  $\Delta x$  by means of the increment  $\Delta y$ , since  $\Delta x = g(y + \Delta y) - g(y)$ . Hence, the above difference quotient can also be written as

$$\frac{\Delta y}{g(y + \Delta y) - g(y)} = \frac{1}{[g(y + \Delta y) - g(y)]/\Delta y}.$$

Since the two functions  $f(x)$  and  $g(y)$  are continuous we have  $\Delta y \rightarrow 0$  when  $\Delta x \rightarrow 0$ , and vice versa. This implies that

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta y \rightarrow 0} \frac{1}{[g(y + \Delta y) - g(y)]/\Delta y} = \frac{1}{g'(y)},$$

provided that  $f'(x) \neq 0$  and  $g'(y) \neq 0$  in the interval  $a \leq x \leq b$ .

### 1.3 Basic concepts of integral calculus

Integral calculus emerged from the need to determine areas of surfaces differing from those of rectangles and to find equations for curves where the behavior of their tangents were known. The basic method was known to the Greek mathematicians\*, for example in their attempts to find the area of a circle, which was confined between the  $n$ -sided regular inscribed and circumscribed polygons, whose areas are known from Euclidian geometry. As  $n$  increases, the difference between the two areas becomes smaller. We can make this difference as small as we please by choosing  $n$  sufficiently large, and so the value of the area can be estimated to any degree of accuracy that is required. This method of exhaustion is essentially that of integral calculus.

\* It was known in particular by Archimedes (287–212 BC), who, in addition to his great contributions to mathematics, is also regarded as the founder of the laws of equilibrium in rigid and fluid bodies. He was also an imaginative inventor.

## 1.3.1 Definite and indefinite integral

Let  $y = f(x)$  be a function represented by a finite, positive value in the interval  $a \leq x \leq b$ . The *definite integral* of the function  $y = f(x)$  from  $x = a$  to  $x = b$  is defined by the following operation. The interval  $a \leq x \leq b$  is divided in  $n$  subintervals

$$\Delta x_1, \Delta x_2, \dots, \Delta x_i, \dots, \Delta x_n.$$

Let  $f(x_i)$  be the value of the function somewhere in the subinterval  $\Delta x_i$ . One then introduces the sum

$$S_n = f(x_1)\Delta x_1 + f(x_2)\Delta x_2 + \dots + f(x_i)\Delta x_i + \dots + f(x_n)\Delta x_n, \quad (1.3.1)$$

or

$$S_n = \sum_{i=1}^{i=n} f(x_i)\Delta x_i = \sum_{i=1}^{i=n} \Delta A_i, \quad (1.3.2)$$

where  $\sum$  stands for “sum of elements of the form ...”, in this case

$$f(x_i)\Delta x_i = \Delta A_i,$$

where  $\Delta A_i$  is the area of the rectangle with sides  $f(x_i)$  and  $\Delta x_i$ .

If this sum  $S_n$  assumes a definite value, *the limit* of  $S_n$ , when all intervals  $\Delta x_i$  approach zero as the number of intervals  $n \rightarrow \infty$ , the function  $f(x)$  is said to be *integrable* in the interval between  $x = a$  and  $x = b$ . The value of this limit for  $S_n$  is denoted the *definite integral* of  $y = f(x)$  from  $x = a$  to  $x = b$ . The symbolism that reflects this operation is

$$\lim_{\substack{n \rightarrow \infty \\ \Delta x_i \rightarrow 0}} \sum_{i=1}^{i=n} f(x_i)\Delta x_i \stackrel{\text{def}}{=} \int_a^b f(x) dx. \quad (1.3.3)$$

The symbol  $\int$  – an elongated S – was introduced by Leibniz\* to make an association to the “sum of infinitely large number of infinitely small subelements”, and the symbol has retained its value of convenience ever since and is called the *integral sign*. We denote  $x = a$  as the *lower limit* of the definite integral and  $x = b$  as the *upper limit*. The arithmetic definition above also holds if  $a > b$ , as the only change that arises is that the differences  $\Delta x_i = f(x_{i+1}) - f(x_i)$  now become negative when the interval is traversed from  $a$  to  $b$ . This suggests the relation

$$\int_a^b f(x) dx = - \int_b^a f(x) dx, \quad (1.3.4)$$

\* In a manuscript dated 29th October 1675.

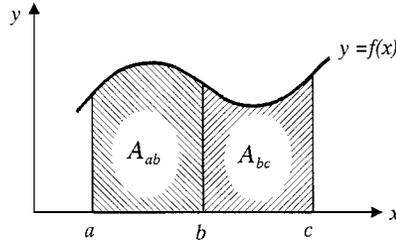


Fig. 1.2. Illustration of the definite integral as an area.

and the definition

$$\int_a^a f(x) dx = 0. \quad (1.3.5)$$

From a geometric point of view, Eq. (1.3.1) gives the value of the area between the curve  $y = f(x)$  and the  $x$ -axis that is delimited by the lines  $x = a$  and  $x = b$ . An example of such an area  $A_{ab}$  is shown in Fig. 1.2 together with the adjacent area  $A_{bc}$  that is delimited by the curve  $y = f(x)$  and by the lines  $x = b$  and  $x = c$ . Denoting the total area between the lines  $x = a$  and  $x = c$  as  $A_{ac}$ , we have:  $A_{ab} + A_{bc} = A_{ac}$ , or

$$\int_a^b f(x) dx + \int_b^c f(x) dx = \int_a^c f(x) dx. \quad (1.3.6)$$

On account of Eq. (1.3.4) and Eq. (1.3.5) this relation will hold for any mutual positions the three points  $a$ ,  $b$  and  $c$  may assume.

In Fig. 1.2 it is assumed that the function  $f(x)$  is positive in the whole range considered. However, the integral that is defined by Eq. (1.3.1) as the limit of the sum of elements  $f(x_i)\Delta x_i$  is independent of such an assumption. If  $f(x) < 0$  in part of the range from  $a$  to  $b$  it only results in making the summation elements in question negative, thereby assigning a negative value to the area where the curve of  $f(x)$  is located *below* the  $x$ -axis. Thus, the total area that is enveloped by an arbitrary curve  $y = f(x)$ , will in general comprise positive as well as negative areas.

Let  $y = f(t)$  represent a function of the independent variable that, for reasons of convenience, we shall denote by  $t$ . Next we consider the integral of this function taken from a fixed point  $t = a$  to another point  $t = x$ , which we allow to vary on the  $t$ -axis. The value of this integral is then determined by the value

that is assigned to  $x$ . Thus, the integral will be a function  $F(x)$  of its upper limit  $t = x$ , namely

$$F(x) = \int_a^x f(t) dt. \quad (1.3.7)$$

The function  $F(x)$  is the area between the curve  $y = f(t)$  and the  $t$ -axis that is delimited by the fixed line  $t = a$  and the line  $t = x$  that may vary as we please. For that reason an integral  $F(x)$  with a variable upper limit is called an *indefinite integral*. The condition for the existence of an indefinite integral  $F(x)$  is that the function  $y = f(t)$  is *continuous*.

### 1.3.2 The fundamental law

The fundamental law of integral and differential calculus\* states: *the derivative of the indefinite integral  $F(x)$  of the function  $y = f(t)$  with respect to  $x$  is equal to the value of  $f(t)$  for  $t = x$ , namely*

$$F'(x) = \frac{dF}{dx} = f(x), \quad (1.3.8)$$

that is *the process of integration that leads from the function  $f(x)$  to  $F(x)$  can be reversed by taking the derivative of the function  $F(x)$  with respect to  $x$ .*

This important theorem can be demonstrated by applying the limiting procedure Eq. (1.2.4) to the difference quotient (Eq. (1.2.3)) of the indefinite integral, i.e.

$$F'(x) = \lim_{h \rightarrow 0} \left[ \frac{F(x+h) - F(x)}{h} \right].$$

From Eq. (1.3.4) and Eq. (1.3.6) it follows that the denominator can be written as

$$\begin{aligned} F(x+h) - F(x) &= \int_a^{x+h} f(t) dt - \int_a^x f(t) dt \\ &= \int_x^a f(t) dt + \int_a^{x+h} f(t) dt \\ &= \int_x^{x+h} f(t) dt. \end{aligned}$$

The right-hand side of Fig. 1.3 can be visualized as the area between the curve  $y = f(t)$  and the  $t$ -axis that is delimited by the lines  $t = x$  and  $t = x + h$ . Furthermore it is seen that this area is contained between the two rectangles of

\* This theorem was discovered around 1670 by Isaac Newton (1642–1727) and by G.W. Leibniz (1646–1716), independently of each other.

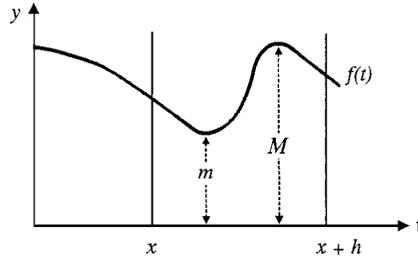


Fig. 1.3. To the derivation of the fundamental law of integral and differential calculus.

areas  $hm$  and  $hM$ , where  $m$  and  $M$  are the smallest and largest value respectively of  $y = f(t)$  in the interval  $x \leq t \leq x + h$ . Thus, we have

$$m \leq \frac{F(x+h) - F(x)}{h} \leq M.$$

As the function  $y = f(t)$  is continuous both  $m$  and  $M$  will approach the value  $f(x)$  when  $h \rightarrow 0$ . At the same time the difference quotient  $(f(x+h) - f(x))/h$  will approach  $F'(x)$ . Thus, the above limit becomes

$$F'(x) = \lim_{h \rightarrow 0} \left[ \frac{F(x+h) - F(x)}{h} \right] = f(x).$$

This version of the derivation of Eq. (1.18) is due to Cauchy\* (1823).

Thus, to obtain an *indefinite integral* or a *primitive function* of the function  $y = f(x)$  one has to find a function  $F(x)$ , whose derivative is equal to  $f(x)$ , i.e. find a function with the property

$$F'(x) = f(x). \quad (1.3.9)$$

### 1.3.3 Evaluation of a definite integral

Having at our disposal *one* primitive function  $F(x)$  – an indefinite integral – that satisfies Eq. (1.3.8), we can construct any number of primitive functions, such as the function

$$G(x) = F(x) + C, \quad (1.3.10)$$

where  $C$  is a constant that will also satisfy Eq. (1.3.8), because the derivative of the function  $y = C$  is equal to zero. This property leads to an important rule

\* Augustin Louis Cauchy (1789–1857) was one of the greatest mathematicians. He was the founder of the modern theory of functions of complex variables, and was responsible for further development of the theory of differential equations, difference equations and infinite series.

for finding the value of a definite integral of the function  $f(x)$  taken between the limits  $a$  and  $b$ , if a primitive function  $G(x)$  of  $f(x)$  is known.

Consider the primitive function

$$F(x) = \int_a^x f(t) dt,$$

of the function  $y = f(x)$ . Equation (1.3.10) can then be written as

$$G(x) = \int_a^x f(t) dt + C. \quad (1.3.11)$$

This expression is also valid for  $x = a$ , namely

$$G(a) = \int_a^a f(t) dt + C.$$

But according to Eq. (1.3.5) we have

$$\int_a^a f(t) dt = 0$$

and hence

$$G(a) = \int_a^a f(t) dt + C = 0 + C.$$

Inserting  $C = G(a)$  in Eq. (1.3.11) and putting  $x = b$  gives

$$G(b) = \int_a^b f(t) dt + G(a),$$

or

$$\int_a^b f(t) dt = G(b) - G(a), \quad (1.3.12)$$

no matter which of the many possible forms for  $G(x)$  one may choose to use. We then have the following important result: *to calculate the value of the definite integral*

$$\int_a^b f(x) dx,$$

*we have only to find a function  $G(x)$  with the property  $G'(x) = f(x)$  and then form the difference  $G(b) - G(a)$ .*

To simplify the notation it has been found to be convenient to remove the limits from the integral sign in Eq. (1.3.11) and modify the graphics for the

indefinite integral to

$$G(x) = \int f(x) dx + C, \quad (1.3.13)$$

where  $\int \cdots dx$  means: find a function  $F(x)$  with the property  $F'(x) = f(x)$ , and have the additive constant  $C$  in mind. Sometimes it may be useful to remember the above formula in this way

$$G(x) = \int \left( \frac{dF}{dx} \right) dx + C = F(x) + C, \quad (1.3.14)$$

in particular in those cases where it is almost directly obvious that the function  $f(x)$  can be written as the derivative of a function  $F(x)$ . The indefinite integral on the form  $\int dx$  sometimes leads to difficulties in understanding until one realizes that the integrand in this case is  $f(x) = 1$ , which again is the derivative of the function  $F(x) = x$ . Hence we have:  $\int dx + C = x + C$ .

### 1.3.4 The mean value theorem

There are several ways for estimating the value of a definite integral. We shall consider the simplest. Let  $y = f(x)$  represent a continuous non-negative function – either positive or zero – in the interval  $a \leq x \leq b$ , i.e.  $f(x) \geq 0$ . For the definite integral it holds that

$$\int_a^b f(x) dx = \lim_{\substack{n \rightarrow \infty \\ \Delta x_i \rightarrow 0}} \sum_{i=1}^{i=n} f(x_i) \Delta x_i \geq 0,$$

as the sum contains only positive elements. Let  $M$  denote a number such that  $M \geq f(x)$  for every value of  $x$  in the interval  $a \leq x \leq b$ . Furthermore, let  $m$  denote another number such that  $m \leq f(x)$  for every  $x$  in the interval  $a \leq x \leq b$ . Hence we have

$$\int_a^b m dx \leq \int_a^b f(x) dx \leq \int_a^b M dx.$$

This double inequality is illustrated geometrically in Fig. 1.4.

But we have

$$\int_a^b m dx = m \int_a^b dx = m(b-a), \quad \text{and also} \quad M \int_a^b dx = M(b-a),$$

and hence

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a).$$

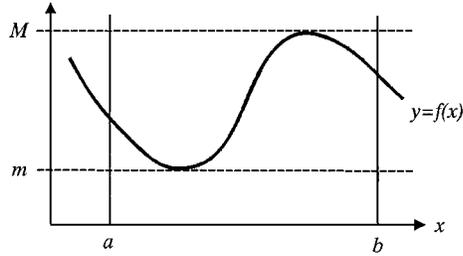


Fig. 1.4. Illustration of the mean value theorem for the definite integral.

Therefore, the value of the definite integral can be represented as the product of  $(b - a)$  and some number  $\mu$  that is located between  $m$  and  $M$ :

$$\int_a^b f(x) dx = \mu(b - a), \quad m \leq \mu \leq M, \quad (1.3.15)$$

where we can regard  $\mu$  as the *mean value* of  $f(x)$  in the interval  $a \leq x \leq b$ . The function  $y = f(x)$  is continuous in the interval considered, and will therefore assume all values between the largest and smallest value of  $f(x)$  in the interval. Therefore, we can put  $\mu = f(\xi)$  where  $\xi$  is located somewhere in the interval. The last expression can therefore also be written as

$$\int_a^b f(x) dx = (b - a)f(\xi), \quad a \leq \xi \leq b. \quad (1.3.16)$$

This formula is called the *mean value theorem of the integral calculus*.

## 1.4 The natural logarithm

### 1.4.1 Definition of the natural logarithm

After this recapitulation of the fundamentals of the integral calculus we consider the function

$$y = f(x) = x^n.$$

If  $n$  is different from  $-1$  there exists an indefinite integral

$$G(x) = \int x^n dx = \frac{1}{n+1}x^{n+1} + C, \quad (1.4.1)$$

since  $G'(x) = x^n$ . If  $n = -1$ , the function assumes the form

$$f(x) = \frac{1}{x} = x^{-1}.$$

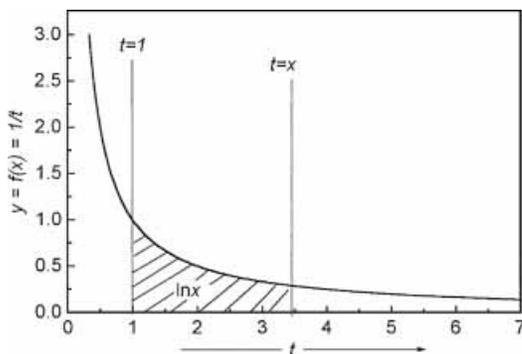


Fig. 1.5. Illustration of the geometric meaning of the natural logarithm  $y = \ln x$  as an area.

The right-hand side of Eq. (1.4.1) then becomes indeterminate since  $1/(n+1) = 1/(-1+1) = 1/0$ . Thus, in this case the integral of  $f(x) = 1/x$  cannot be expressed by Eq. (1.4.1). It turns out to be impossible to find an indefinite integral of the function  $y = 1/x$  that is expressed in terms of elementary functions, i.e. polynomials, fractional rational functions (the ratio between two polynomials) or algebraic functions (e.g. the square root of a polynomial). Because of the frequent occurrence of the integral  $\int dx/x$ , mathematicians found it convenient to define a *new* function by means of this integral. This function is called the *natural logarithm* and is denoted as  $\ln x$ . This function is *defined* by the integral

$$\ln x = \int_1^x \frac{1}{t} dt, \quad (1.4.2)$$

i.e. as the area between the rectangular hyperbola  $y = 1/t$  and the  $x$ -axis that is delimited between the line  $t = 1$  and the line  $t = x$  (Fig. 1.5). The variable  $x$  can be any positive number, but  $x = 0$  is excluded because the integral diverges as the integrand  $y = 1/t$  becomes infinite when  $x \rightarrow 0$ .

### 1.4.2 Elementary properties of the logarithm

The function  $y = \ln x$  is useful for several reasons. The first follows from the fundamental theorem Eq. (1.3.8). We have

$$f'(x) = \frac{d \ln x}{dx} = \frac{1}{x}. \quad (1.4.3)$$

Thus, the derivative of  $y = \ln x$  is always positive, but it decreases for increasing values of  $x$ . In accordance with this we see that the area under the rectangular

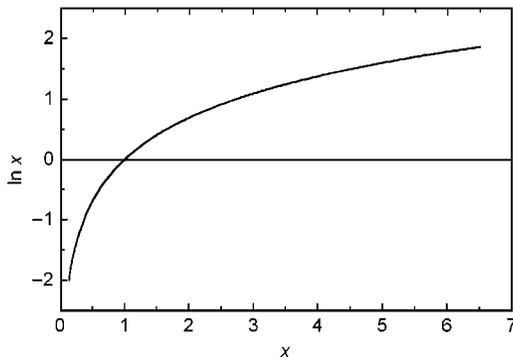


Fig. 1.6. The course of the function  $y = \ln x$ . The number  $x = e$  satisfies the relation  $\ln e = 1$ .

hyperbola  $y = 1/x$  taken between the two lines at  $x$  and  $x + \Delta x$  decreases monotonically with increasing values of  $x$ . The course of the function  $y = \ln x$  is illustrated in Fig. 1.6. Below we shall recapitulate the three basic properties of the logarithmic function.

#### 1.4.2.1 Logarithm of a product

The main property of the logarithmic function is given by the formula

$$\ln a + \ln b = \ln(ab). \quad (1.4.4)$$

To demonstrate this theorem we consider the function  $F(x) = \ln x$  together with another function

$$G(x) = \ln(ax) = \ln w = \int_1^w \frac{1}{t} dt, \quad (1.4.5)$$

where  $w = ax$ . Taking the derivative of  $G(x)$  with respect to  $x$  yields (see Eq. (1.2.15))

$$G'(x) = \frac{d \ln w}{dw} \frac{dw}{dx} = \frac{1}{w} \frac{d(ax)}{dx} = \frac{1}{ax} a = \frac{1}{x}.$$

We also have

$$F'(x) = \frac{1}{x}.$$

The two functions  $F(x)$  and  $G(x)$  have exactly the same derivative and, consequently, can only differ from each other by a constant number. Thus,

$$G(x) = F(x) + C,$$