

CONTENTS

NUMERICAL METHODS I – BASIS AND FUNDAMENTALS

1. NUMERAL SYSTEMS	11
1.1 Decimal system	21
1.2 Binary system	22
1.2.1 Conversions from the binary to the decimal system	23
1.2.2 Conversions from the decimal to the binary system	24
1.3 Octal system	26
1.4 Hexadecimal system	27
2. NUMBERS IN DIGITAL COMPUTERS	29
2.1 Binary coded decimal (BCD)	30
2.2 Densely packed decimal (DPD)	31
2.3 Negative integers	31
2.3.1 Second complement	31
2.3.2 Biased integers	33
2.4 Formats of numbers in digital computers	34
2.5 Binary integer formats	36
2.5.1 Binary integer in byte	37
2.5.2 Short integer (16 bits)	39
2.5.3 Integer (32 bits)	39
2.5.4 Long integer (64 bits)	40
2.5.5 Double long integer (128 bits)	40
2.6 Decimal integer formats	41
2.7 Introduction to floating-point formats	42
2.7.1 Types of floating-point formats	43

2.7.2 Binary significand	44
3.7.3 Decimal significand	45
2.8 Binary floating-point formats	47
2.8.1 Binary16 (Half-precision binary format)	51
2.8.2 Binary32 (Single-precision binary format)	55
2.8.3 Binary64 (Double-precision binary format)	58
2.8.4 Binary80 (Extended double-precision binary format)	60
2.8.5 Binary128 (Quadruple-precision binary format)	63
2.9 Decimal floating-point formats	65
2.9.1 Decimal32 (Single-precision decimal format)	71
2.9.2 Decimal64 (Double-precision decimal format)	76
2.9.3 Decimal128 (Quadruple-precision decimal format)	80
3. VARIABLES AND NUMERICAL ERRORS	85
3.1 Common terminology	86
3.1.1 Errors and uncertainty	86
3.1.2 Precision vs accuracy	88
3.1.3 Significant digits vs accuracy	89
3.2 Numerical errors	90
3.2.1 Rounding errors	91
3.2.2 Representation error	92
3.3 Propagation of numerical error	97
3.3.1 Propagated error of a single variable	104
3.3.2 Propagated error of multiple variables	107
3.4 Propagated error vs representation error	113
3.5 Numerical instability	115
3.6 Method error	118
3.7 Approximation error	119
4. RANDOM VARIABLES AND PROCESSES	121
4.1 Random processes	122
4.1.1 Definition of random variables	123
4.1.2 Quantification of random processes	125
4.1.3 Equivalent non-random process	127
4.1.4 Probability distribution	129
4.1.5 Probability density	130
4.1.6 Probability function	131
4.1.7 Sampling	132
4.2 Parameters and statistics	135
4.2.1 Expected value	135
4.2.2 Arithmetical average and mean	137
4.2.3 Population covariance	140

4.2.4 Sample covariance	141
4.2.5 Variance	142
4.2.6 Standard deviations	145
4.2.7 Correlation coefficient	146
4.2.8 Linear combinations of random variables	149
4.2.9 Nonlinear combinations of random variables	151
5. RANDOM ERRORS AND UNCERTAINTY	153
5.1 Normal distribution	154
5.2 Error in the mean (repeated observations)	158
5.3 Uncertainty vs sample deviation	159
5.4 Propagation of uncertainty	163
5.4.1 Function of a single variable	165
5.4.2 Propagated random error of multiple variables	166
5.5 Randomisation of constant variables	170
6. REGRESSION	171
6.1 Linear regression	171
6.1.1 Simple linear regression	172
6.1.2 Simple linear regression through fixed point	176
6.1.3 Simple linear regression in multiple variables	178
6.1.4 Orthogonal linear regression in two variables	179
6.1.5 Orthogonal linear regression in multiple variables	185
6.2 Polynomial regression.....	187
6.3 Logarithmic regression	189
6.4 Exponential regression	193
6.5 Power regression	197
7. ROOTS OF UNIVARIATE POLYNOMIALS	201
7.1 Descartes' rule of signs	202
7.2 Roots of second degree polynomials	203
7.2.1 Floating-point implementation	204
7.3 Roots of third degree polynomials	206
7.3.1 Polynomials with three real roots	209
7.3.2 Polynomials with two equal real roots	212
7.3.3 Polynomials with three equal real roots	213
7.3.4 Algorithm for finding three real roots	213
7.3.5 Polynomials with one real and a pair of complex roots	216
7.3.6 Algorithm for finding real and complex roots	220
7.4 Roots of fourth degree polynomials	222
7.4.1 Real roots	226
7.4.2 Complex-conjugated roots	228

7.5 Newton's method	233
7.6 Weierstrass method	235
APPENDIX	239
A.1 Events	239
A.2 Probability	243
A.3 Typical probability distributions	246
A.4 Sum of binomially distributed random variables	249
INDEXES	251
INDEX OF ALGORITHMS	251
INDEX OF EXAMPLES	252
INDEX	255
REFERENCES	256

SOME MATHEMATICAL SYMBOLS

\forall	<i>for all</i>
\exists	<i>exist(s)</i>
\setminus	<i>except, excluding</i>
$ $	<i>with property,</i>
\wedge	<i>and</i> (binary logical operator)
\vee	<i>or</i> (binary logical operator)
$\{ \}$	<i>set of elements</i>
\in	<i>is element of a set</i>
\notin	<i>is not element of a set</i>
\subset	<i>subset of a set</i>
\subseteq	<i>subset or equal to a set,</i>
\cup	<i>union of two (or more) sets</i>
\cap	<i>common elements of two (or more) sets</i>
Σ	<i>sum</i>

Events

\overline{A}	complementary event of the event A
\oplus	sum of events (creates compound event)
\otimes	product of events (creates sub-event contained in both events)
\times	product of subsequent events (product of the second kind)
\setminus	difference of events (excludes sub-events)
\subseteq	inclusion of events
$A B$	event A occurs given that B has occurred