

---

# Contents

---

<b>Preface</b> . . . . .	ix
<b>Abbreviations</b> . . . . .	xi
<b>Chapter 1. The Minimum You Need to Know About Optimization</b> . . . . .	1
1.1. General concepts . . . . .	1
1.1.1. Convexity (set and objective) . . . . .	3
1.1.2. Smoothness of the objective . . . . .	3
1.1.3. Mathematical optimization problems . . . . .	5
1.2. Nonlinear optimization: standard optimization problems . . . . .	6
1.2.1. Optimization with no constraints . . . . .	6
1.2.2. Optimization with equality constraints . . . . .	7
1.3. Nonlinear optimization: convex programming problem . . . . .	9
1.4. Gradient descent algorithm (smooth objective) . . . . .	12
1.5. Non-smooth non-convex optimization from an integer problem . . . . .	13
1.5.1. Solutions of equation [1.7] are reached at the samples of $\mathbf{x}[\mathbf{i}, \mathbf{r}]$ , when $0 < p < 1$ . . . . .	14
1.5.2. Programming the solution of equation [1.7] for $0 < p < 1$ . . . . .	16
<b>Chapter 2. The Essentials to Know in Functional Analysis</b> . . . . .	17
2.1. Integration (the basics) . . . . .	17
2.1.1. Information measures (from examples) . . . . .	18
2.1.2. Basic integral evaluation methods . . . . .	19
2.2. Function spaces . . . . .	22
2.2.1. Function spaces $L^p$ , $p \in \{1, 2\}$ . . . . .	23
2.2.2. Function spaces $\mathcal{D}_0$ , $\mathcal{D}$ and $\mathcal{S}$ (test functions) . . . . .	24
2.2.3. Convergence results for spaces $L^p$ , $\mathcal{D}$ and $\mathcal{D}_0$ . . . . .	25
2.3. Generalized functions: distribution theory . . . . .	26

2.3.1. Definitions . . . . .	26
2.3.2. Properties of distributions . . . . .	29
2.3.3. Derivatives of a distribution . . . . .	30
2.3.4. Convergence in the sense of distributions . . . . .	31
2.3.5. Causal distributions . . . . .	31
2.4. Generalized functions: case of Schwartz distributions . . . . .	31
2.4.1. Regular and singular Schwartz distributions . . . . .	31
2.4.2. On translations for Schwartz distributions . . . . .	34
2.4.3. On periodicity for Schwartz distributions . . . . .	34
2.4.4. Transposition for Schwartz distributions . . . . .	35
2.4.5. Change of the scale for Schwartz distributions . . . . .	35
2.4.6. Product of a Schwartz distribution by $C^\infty$ function . . . . .	35
2.4.7. Derivatives for Schwartz distributions . . . . .	36
2.4.8. Convergence in the distributional sense (in $\mathcal{D}'$ ) . . . . .	36
2.5. Generalized functions: case of tempered distributions . . . . .	37
2.5.1. Fourier transform and regular distributions . . . . .	38
2.5.2. Fourier transform and Dirac distributions . . . . .	39
2.5.3. Fourier transform and derivation . . . . .	39
2.6. Convolution of distributions . . . . .	39
2.6.1. Existence of convolution operation . . . . .	40
2.6.2. Properties of convolution operations . . . . .	41
2.6.3. Derivation and convolution . . . . .	42
2.6.4. Primitive of a distribution in terms of convolution . . . . .	42
2.6.5. Fourier transform and convolution . . . . .	42
2.6.6. Inverse convolution . . . . .	42
2.6.7. Causal Schwartz distributions and convolution . . . . .	43
2.6.8. Fourier transform, periodicity and convolution . . . . .	43
<b>Chapter 3. Probability Entropy and Neural Statistical Parameterizations . . . . .</b>	<b>45</b>
3.1. Probabilistic parameterizations and entropy . . . . .	45
3.1.1. Models of random variables from probability distributions . . . . .	46
3.1.2. Entropies . . . . .	48
3.1.3. Relative entropy of random variables . . . . .	51
3.1.4. Extension to random vectors . . . . .	52
3.2. Statistical parameterizations and neural networks . . . . .	54
3.2.1. Autoregressive (AR), moving average (MA) models . . . . .	55
3.2.2. Nonlinear autoregressive (NAR and NARX) neural network models . . . . .	57
3.2.3. Long short-term memory (LSTM) models . . . . .	62
<b>Chapter 4. Convolutional Neural Networks . . . . .</b>	<b>63</b>
4.1. Motivation . . . . .	63
4.2. Multiserial convolution networks (filter bank architectures) . . . . .	66
4.2.1. Uniserial $M$ -band discrete wavelet packet transform ( $M$ -DWPT) . . . . .	66
4.2.2. First relaxation of $M$ -DWPT: multiserial splits and analogy with CNN . . . . .	68
4.2.3. Second relaxation: removing wavelet and orthogonality constraints . . . . .	69

4.3. Analytic expressions of CNNs . . . . .	71
4.3.1. Analytic expressions of expanded CNNs . . . . .	72
4.3.2. Analytic expressions of standard CNNs . . . . .	73
4.3.3. Summary of the remaining CNN operations performed prior to the decision . . . . .	75
4.4. Deep graphs, path description and decomposition-fusion issues . . . . .	76
4.4.1. Algebraic descriptions of deep network paths . . . . .	76
4.4.2. Open issues . . . . .	78
4.5. Learning analytical expansions . . . . .	83
<b>References</b> . . . . .	87
<b>Index</b> . . . . .	91