Stochastic Approximation and Recursive Estimation

M. B. Nevel’son
R. Z. Has’minskiĭ
M. B. Nevel'son
R. Z. Has'minskiĭ

Stochastic Approximation and Recursive Estimation
Abstract. This book is devoted to sequential methods of solving a class of problems to which belongs, for example, the problem of finding a maximum point of a function if each measured value of this function contains a random error. Some basic procedures of stochastic approximation are investigated from a single point of view, namely the theory of Markov processes and martingales. Examples are considered of applications of the theorems to some problems of estimation theory, educational theory and control theory, and also to some problems of information transmission in the presence of inverse feedback.
# TABLE OF CONTENTS

Preface

Introduction

Chapter 1. Elements of probability and martingales
  1. Probability
  2. Random variables
  3. Conditional probabilities and conditional expectations
  4. Independence. Product measures
  5. Martingales and supermartingales

Chapter 2. Discrete time Markov processes
  1. Markov processes
  2. Markov processes and supermartingales
  3. Recursively defined processes
  4. Discrete model of diffusion
  5. Exit of sample functions from a domain
  6. Series of independent random variables
  7. Convergence of sample functions
  8. Teaching pattern recognition

Chapter 3. Markov processes and stochastic equations
  1. Continuous time Markov processes
  2. Stochastic differential equations. I
  3. Stochastic integrals
  4. Stochastic differential equations. II
  5. Itô's formula
  6. Supermartingales
  7. Existence of solutions in the large
  8. Exit from a domain. Convergence of sample functions

Chapter 4. Convergence of stochastic approximation procedures. I
  1. The Robbins-Monro procedure
  2. The Kiefer-Wolfowitz procedure
  3. Continuous procedures
  4. Convergence of the Robbins-Monro procedure
  5. Convergence of the Kiefer-Wolfowitz procedure
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5</td>
<td>Convergence of stochastic approximation procedures. II</td>
<td>101</td>
</tr>
<tr>
<td>1</td>
<td>Introductory remarks</td>
<td>101</td>
</tr>
<tr>
<td>2</td>
<td>General theorems</td>
<td>102</td>
</tr>
<tr>
<td>3</td>
<td>Auxiliary results (continuous time)</td>
<td>106</td>
</tr>
<tr>
<td>4</td>
<td>Auxiliary results (discrete time)</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>One-dimensional procedures</td>
<td>118</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Asymptotic normality of the Robbins-Monro procedure</td>
<td>123</td>
</tr>
<tr>
<td>1</td>
<td>Preliminary remarks</td>
<td>123</td>
</tr>
<tr>
<td>2</td>
<td>Asymptotic behavior of solutions</td>
<td>129</td>
</tr>
<tr>
<td>3</td>
<td>Investigation of the process $I_1(t)$</td>
<td>132</td>
</tr>
<tr>
<td>4</td>
<td>Investigation of the process $I_2(t)$</td>
<td>136</td>
</tr>
<tr>
<td>5</td>
<td>Asymptotic normality (continuous time)</td>
<td>140</td>
</tr>
<tr>
<td>6</td>
<td>Asymptotic normality (discrete time)</td>
<td>147</td>
</tr>
<tr>
<td>7</td>
<td>Convergence of moments</td>
<td>154</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>Some modifications of stochastic approximation procedures</td>
<td>161</td>
</tr>
<tr>
<td>1</td>
<td>Statement of the problem</td>
<td>161</td>
</tr>
<tr>
<td>2</td>
<td>General theorem</td>
<td>162</td>
</tr>
<tr>
<td>3</td>
<td>Auxiliary results</td>
<td>165</td>
</tr>
<tr>
<td>4</td>
<td>Theorems on convergence and asymptotic normality</td>
<td>167</td>
</tr>
<tr>
<td>5</td>
<td>Adaptive Robbins-Monro procedures</td>
<td>169</td>
</tr>
<tr>
<td>6</td>
<td>Asymptotic optimality</td>
<td>175</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>Recursive estimation (discrete time)</td>
<td>181</td>
</tr>
<tr>
<td>1</td>
<td>The Cramér-Rao inequality. Efficiency of estimates</td>
<td>181</td>
</tr>
<tr>
<td>2</td>
<td>The Cramér-Rao inequality in the multidimensional case</td>
<td>186</td>
</tr>
<tr>
<td>3</td>
<td>Estimation of a one-dimensional parameter</td>
<td>189</td>
</tr>
<tr>
<td>4</td>
<td>Asymptotically efficient recursive procedure</td>
<td>194</td>
</tr>
<tr>
<td>5</td>
<td>Estimation of a multidimensional parameter</td>
<td>197</td>
</tr>
<tr>
<td>6</td>
<td>Estimation with dependent observations</td>
<td>202</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Recursive estimation (continuous time)</td>
<td>207</td>
</tr>
<tr>
<td>1</td>
<td>The Cramér-Rao inequality</td>
<td>207</td>
</tr>
<tr>
<td>2</td>
<td>Application of the Robbins-Monro procedure</td>
<td>210</td>
</tr>
<tr>
<td>3</td>
<td>Time-dependent observations</td>
<td>212</td>
</tr>
<tr>
<td>4</td>
<td>Some applications</td>
<td>213</td>
</tr>
<tr>
<td>5</td>
<td>A modification</td>
<td>220</td>
</tr>
<tr>
<td>Chapter 10</td>
<td>Recursive estimation with a control parameter</td>
<td>221</td>
</tr>
<tr>
<td>1</td>
<td>Statement of the problem</td>
<td>221</td>
</tr>
<tr>
<td>2</td>
<td>Asymptotically optimal recursive plan</td>
<td>223</td>
</tr>
<tr>
<td>3</td>
<td>Two examples</td>
<td>226</td>
</tr>
<tr>
<td>4</td>
<td>Continuous case</td>
<td>229</td>
</tr>
<tr>
<td>Notes on the Literature</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>Bibliography</td>
<td>237</td>
<td></td>
</tr>
<tr>
<td>Subject Index</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>Main Notation</td>
<td>244</td>
<td></td>
</tr>
</tbody>
</table>
This page intentionally left blank
NOTES ON THE LITERATURE

Chapter 1

The proof of all theorems cited in §§1–4 may be found in Kolmogorov [1], Halmos [1], and Kolmorogov and Fomin [1]. The martingale and supermartingale concepts are due to Doob. Theorems 5.1 and 5.1’ and their proofs are taken from Doob [1].

Chapter 2

§1. For more details about properties of Markov processes and their transition functions, see Feller [1], Dynkin [1], Loève [1], and others.

§2. Theorem 2.1 is essentially a special case of Doob’s theorem (Doob [1]) on the stopping of a supermartingale. Kolomogorov’s inequality for supermartingales may also be found in Doob [1].

§3. Theorems 3.1 and 3.2 are well known. See, for example, Gihman and Skorohod [1].

§5. For related results, see, for example, ABR [1], Chapter IV.

§6. Theorems 6.1 and 6.2 were proved by Kolomogorov [1] in a more general setting.

§7. Theorem 7.1 is conceptually similar to Theorem 8 in Chapter 4 of ABR [1].

§8. For a detailed discussion of the questions related to this formulation of the problem, see ABR [1], Chapter 5.

Chapter 3

§1. For more details about the definition and properties of a continuous Markov process, see Dynkin [1].

§2. The limit passage from equation (2.1) to a continuous Markov process was considered by Bernstein [1], who, among other things, proved a theorem on the limit behavior of the transition probability as $h \to 0$.

§§3 and 4. This definition of the stochastic integral is due to Itô. For a proof of Theorem 4.1, see, for example, Gihman and Skorohod [2].

§7. A proof of Theorem 7.1 may be found in Has’minskii [1].

§8. Similar results are presented in the authors’ paper [1].

Chapter 4

§1. As mentioned in the text, s.a. procedures for location of the roots of regression equations were first suggested in the 1951 paper of Robbins and Monro [1], who proved convergence in mean square of the procedure (5.1) under certain assumptions. This result
was generalized by Wolfowitz [1], Blum [1], Chung [1], and others. Blum [1], in particular, presented the first proof of convergence with probability 1. Theorem 1.1 and the method of proof used in the text are due to Gladyšev [1]. Multidimensional RM procedures were introduced by Blum [2]. Blum was the first author to apply the theory of supermartingales to investigate the convergence of s.a. procedures.

§ 2. The procedure (2.5) was proposed by Kiefer and Wolfowitz [1] in 1952. Blum [2], Derman [1], Burkholder [1] and others studied the procedure in detail and, in particular, proved that it is convergent with probability 1. No account is given here of more general s.a. procedures such as those proposed in Burkholder [1], Dvoretzky [1], etc. More detailed bibliographies may be found in the review articles of Dvoretzky [1], Fabian [3], Sakriso [3], Schmetterer [1], Loginov [1], and others.

§ 3. The continuous analog of the RM procedure was introduced by Drmíl and Nedoma [1]. They proved the convergence of the procedure under fairly general assumptions on the stochastic processes appearing on the right of the equation. Under certain assumptions, Sakriso [1] proved the convergence of the continuous analog of the Kiefer-Wolfowitz procedure. The continuous procedures considered here were defined in Has’minskiǐ [1] and Nevel’son and Has’minskiǐ [1].

§ 4. The first convergence theorems for multidimensional RM procedures were proved by Blum [2]; see also ABR [1], Schmetterer [1] and further bibliography cited there. Theorem 4.4 is similar to the above-mentioned result of Gladyšev [1]; Theorem 4.5 is due to Braverman and Rozonoër (see ABR [1]). The formulation of the last theorem raises the question as to whether one can also weaken the condition \( \Sigma d^2(t) < \infty \) in the proof that \( X(t) \to x_0 \) a.s. It turns out that this is indeed possible. For example, in some cases it is sufficient to demand that \( \Sigma d^n(t) < \infty \) for some \( n > 0 \) (although one must then impose more stringent conditions on the “noise”). The theorems proved here for continuous RM procedures are due to the authors [1].

§ 5. Convergence theorems for discrete multidimensional KW procedures were proved by Blum [2], Dupač [1] and others. For the continuous case, see Nevel’son and Has’minskiǐ [1].

Chapter 5

The conjecture that a KW procedure cannot converge with positive probability to minimum points of the regression function was advanced by Fabian [1], [2]. The results presented here are based on Has’minskiǐ [1] and Nevel’son [1], [2]. The two last-mentioned papers also prove more general theorems. See also Krasulina [1].

Chapter 6

§ 1. Theorem 1.1 for the one-dimensional case is proved, e.g., in Loève [1]; for the multidimensional case, see Sacks [1].

§ 2. Conditions for the validity of the estimate \( \text{EX}^2(t) = O(1/t) \) as \( t \to \infty \), where \( X(t) \) is a discrete one-dimensional RM process, were first obtained by Chung [1]. An analog of Lemma 2.1 for discrete multidimensional RM processes was proved by Sacks [1] under the additional assumption \( E|G(t, x, \omega)|^2 < c < \infty \). The fact that (2.9) is a consequence of (2.8) follows from a well-known lemma of Chung [1].

§§ 3 and 4. Lemma 3.1 was proved by Has’minskiǐ [2]. Lemma 4.3 is due to Sacks [1].
§5. Theorem 5.3 for $\mu > 0$ is proved in Has'minskiĭ [2]. The asymptotic normality of a RM procedure was established, under certain assumptions, by Holevo [1].

§6. Theorem 6.1 was proved by Sacks under the following additional assumptions: a) the matrix $B$ may be reduced to diagonal form by a similarity transformation involving an orthogonal matrix; b) the norm of the matrix $A(t, x)$ is bounded for $t > 1$ and $x \in \mathcal{F}$. Theorem 6.3 is new. Asymptotic normality theorems for other s.a. procedures were proved by Fabian [5], Derman [1], Burkholder [1], Dupač [1] and others. The "truncation" method in asymptotic normality proofs for s.a. procedures was first applied by Hodges and Lehman [1].

§7. The first convergence theorems for moments of s.a. processes were proved by Chung [1] in the one-dimensional case. For the multidimensional case, the result of Theorem 7.2 is well known (see, for example, Schmetterer [1]) for $\lambda > \frac{1}{2}a$. But if Condition (a) of §7 holds with $\lambda < \frac{1}{2}a$, the only result known to us on convergence of moments is that of Sakrison [2]. Constructing a certain modification of the RM procedure, Sakrison proved that the matrix of second moments of the process $\sqrt{t}(X(t) - x_0)$ converges to the corresponding covariance matrix of a normal law, provided all moments $\mathbb{E}|G(t, x, \omega)|^p, p = 1, 2, \ldots$, are bounded.

Chapter 7

§§1–4. Albert and Gardner [1] established convergence and asymptotic normality for the procedure (1.1) and a more general one in which the factor $a(t)$ is allowed to depend on past observations. These authors also considered a multidimensional modification of (1.1). Recently, Calbi [1], [2] has proved convergence of truncated procedures for a broad range of truncations.

§5. There is a considerable literature on the optimal choice of parameters for RM procedures and their adaptive modifications. Among others, we mention Chung [1], Dvoretzky [1], Kesten [1], Cypkin [1], [2], and Stratonovič [2]. Our approach is based on the work of Venter [1], who proved Theorem 5.1 under slightly more restrictive assumptions.

§6. Theorem 6.1 slightly generalizes a result of Venter [1]; Theorem 6.2 is new.

Chapter 8

§1. Theorem 1.1 follows from results of Chapman and Robbins [1] and Kagan [1]. Theorem 1.2 was proved by Ibragimov and Has'minskiĭ [2]. These papers also derive inequalities for biased estimates.

§2. Theorem 2.1 and its analog for biased estimates were proved by Ibragimov and Has'minskiĭ [2].

§3. The use of RM procedures for parametric estimation is discussed in Albert and Gardner [1], where a larger class of estimates, not necessarily forming Markov processes, is considered.

§§4 and 5. Theorems 4.1 and 4.5 were proved by Sakrison [2], [3] under stronger conditions (and without the convergence of distributions). Other optimality criteria were considered by Albert and Gardner [1], Cypkin [1], and Stratonovič [2], but their procedures generally depend on past observations (see, for example, the procedure (0.9) in the Introduction).
Chapter 9

§1. For inequality (1.7), see, for example, Van Trees [1].

§2. A procedure equivalent to (2.4) for $m$ linear in the parameter $(m(x_0) = mx_0$, where $m$ is a nonsingular matrix) was considered by Holevo [1]. He also considered the case that $m$ is not a square matrix but $mm^*$ is nonsingular. For the nonlinear case, Holevo proposed a procedure more complicated than (2.4). Under certain assumptions, he established asymptotic normality of the estimation procedure.

§3. Recursive procedures similar to (3.1) may be derived from the Bayesian approach, using linearization. For linear systems, similar estimation procedures may be found in Kalman and Bucy [1], Lipcer and Širjaev [1], and elsewhere.

§4. The problem of estimating $x_0$ based on observations of type (4.7) (estimation of linear regression coefficients) has been considered from other points of view in many papers.

§5. With suitable initial conditions, the procedure (5.3), (5.4) for a linear function $m(t, x) = m(t)x$ coincides with the Kalman-Bucy optimal linear filtering equations. In other words, this system is satisfied, in particular, by the conditional expectation of $x_0$ given $Y(s)$, $s < t$, provided the a priori distribution of $x_0$ is Gaussian.
BIBLIOGRAPHY

M. A. Al'zerman, È. M. Braverman and L. I. Rozonoer (ABR)

Arthur E. Albert and Leland A. Garnder, Jr.

Serge Bernstein

David Blackwell and M. A. Girshick

Julius R. Blum

D. L. Burkholder

Douglas G. Chapman and Herbert Robbins

K. L. Chung

Harald Cramér

S. Csiby

Ja. Z. Cypkin

Cyrus Derman

A. G. Djačkov and M. S. Pinsker
J. L. Doob

Miloslav Driml and Jiří Nedomá

Václav Dušáč
1. On the Kiefer-Wolfowitz approximation method, Časopis Pěst. Mat. 82 (1957), 47–75. (Czech; Russian and English summaries) MR 19, 693.

Aryeh Dvoretzky

E. B. Dynkin

Václav Fabian

A. A. Fed'baum

William Feller

G. M. Fichtenhol'c

I. I. Gihman and A. V. Skorohod

E. G. Gladyshev

Paul R. Halmos
R.Z. Has’minskii

J. L. Hodges, Jr., and E. L. Lehmann

A. S. Holevo

I. A. Ibragimov and R. Z. Has’minskii

Kiyosi Itô

Kiyosi Itô and Makiko Nisio

A. M. Kagan

R. E. Kalman and R. S. Bucy

Harry Kesten

J. Kiefer and J. Wolfowitz

A. N. Kolmogorov

A. N. Kolmogorov and S. V. Fomin

N. N. Krasovskii
T. P. Krasulina
1. The Robbins-Monro process in the case of several roots, Teor. Verojatnost. i
MR 35 #5085.

L. M. Libkind
1. On the probability of large deviations for a quadratic functional of a stationary
Gaussian sequence, Problemy Peredači Informacii 5 (1969), vyp. 2, 72–78 = Problems

R. S. Lipcer and A. N. Širjaev
1. Nonlinear filtering of Markov diffusion processes, Trudy Mat. Inst. Steklov. 104
41 #6342.
2. The absolute continuity with respect to Wiener measure of measures that cor-
= Math. USSR Izv. 6 (1972), 839–882. MR 47 #1119.

Michel Loève

N. V. LogiNOV
1. Methods of stochastic approximation, Avtomat. i Telemeh. 27 (1966), no. 4, 185–

I. G. Malkin
transl. of 1st ed., AEC Transl. #3552, U. S. Atomic Energy Commission, Oak

M. B. Nevel’son
Verojatnost. i Primenen. 17 (1972), 310–319 = Theor. Probability Appl. 17 (1972),
298–308. MR 45 #6124.
2. The convergence of continuous and discrete Robbins-Monro procedures in the case
of several roots of the regression equation, Problemy Peredači Informacii 8 (1972),
48 #1320.

M. B. Nevel’son and R. Z. Has’minskii
1. Continuous procedures of stochastic approximation, Problemy Peredači Informacii
148. MR 48 #1319.
2. On the convergence of the moments of the Robbins-Monro procedure, Avtomat. i

M. S. Pinsker
1. The probability of error in block transmission in a memoryless Gaussian channel
with feedback, Problemy Peredači Informacii 4 (1968), vyp. 4, 3–19 = Problems

C. Radhakrishna Rao
36 #4668.

Herbert Robbins and Sutton Monro
MR 13, 144.

Jerome Sacks

David J. SakrisoN

J. Pieter M. Schalkwijk

J. Pieter M. Schalkwijk and T. Kailath

L. Schmetterer

C. E. Shannon

A. V. Skorohod

R. L. Stratonović

Harry L. Van Trees

J. H. Venter

M. T. Watan

J. Wolfowitz

John M. Wozencraft and Irwin Mark Jacobs

K. S. Zigangirov
SUBJECT INDEX

Admissible control, 221
    asymptotically optimal, 230
Asymptotic
    efficiency, 186
    normality, 123
    variance, 185
Borel set, 12
Chapman-Kolmogorov relation, 30
Condition (A), 36
    (B), 72
    (o), 155
Control parameter, 221
Convergence
    almost sure, 15
    in distribution, 15
    in probability, 15
Diffusion
    coefficient, 38
    matrix, 67
Distribution of a vector, 14
Drift
    coefficient, 38
    vector, 67
Estimate, 181
    asymptotically efficient, 186, 210
        strongly, 185, 188, 209
        weakly, 186, 188
    asymptotically unbiased, 182
    efficient, 185, 188
    maximum likelihood, 5
    recursive, 190, 210
    strongly consistent, 182
    unbiased, 182
    weakly consistent, 182
Expectation, 15
    conditional, 17, 18
    relative to a random vector, 19
    finite, 15
Fisher information, 182
    matrix, 186
Function
    Borel, 13
    characteristic (of a set), 17
    distribution, 14
    Liapunov, 89
    measurable, 13
Gaussian white noise, 63
Generating operator, 31
    differential, 67
Independent
    events, 20
    random variable and σ-algebra, 20
    sets of random vectors, 20
    σ-algebras, 20
    vectors, 20
Inequality
    Čebyšev's, 15
    Cramér-Rao (information), 182
    Kolmogorov's (for supermartingales), 33–34
Information
    content, 6, 182
    Fisher, 182
    inequality, 182
    matrix (Fisher), 186
Integral, Lebesgue, 14
    Itô's formula for stochastic differentials, 66–67
    Kiefer-Wolfowitz procedure, 83ff.
    Lebesgue integral, 14
    Lemma
        Fatou's, 16
        Liapunov's, 133
    Martingale, 22, 23, 69
    Measurable space, 12
    Measure, 12
        absolutely continuous, 16
        complete, 12
        finite, 12
        Lebesgue, 12
        probability, 13
        σ-finite, 12
    Observation process, 207
    Plan of experiment, 221
        asymptotically optimal, 223
    Probability, conditional, 17, 18
        relative to a random vector, 19
    Probability space, 13
    Product
        of measures, 21
        of spaces, 20
    Random variable, 13
        independent of the future, 61
        vector-valued, 13
Recursive procedure, 190
  truncated, 162
Regularity condition, 71
Regularization, 31
Robbins-Monro procedure, 4, 79ff.
Sample function of a stochastic process, 22, 53
Separating hyperplane, 50
Set of elementary events, 12
\(\sigma\)-algebras, 12
  Borel, 12
  generated by a random vector, 19
  intersection of, 12
  minimal, 12
Signalling rate, 216
Space of elementary events, 11, 12
Stable
  matrix, 132
  point, 106
Standard Wiener process, 55
  with values in \(E_I\), 56
Statistic, 181
Stochastic
  approximation procedures 4, 79ff.
  Kiefer-Wolfowitz, 83ff.
  Robbins-Monro, 4, 79ff.
differential (Ito’s), 63
equivalence, 14
  of stochastic processes, 54
  strong, 62
integral (Ito’s), 60
process, 53
discrete, 22
  Markov, 30, 54
  homogeneous, 54
measurable, 53
regular, 71
separable, 53
stationary (in the narrow sense), 144

Supermartingale, 23, 69

Theorem
  Fubini’s, 21
  Ito’s, 66
  Kolmogorov’s, 56
  Lebesgue’s, 16
  Radon-Nikodým, 16
Transition density of a Markov process, 54
Transition function, 30, 54
  homogeneous, 31
  temporally homogeneous, 54
Transition probability of a Markov process, 29
MAIN NOTATION

\[ \|B\| = \sqrt{\sum_{ij} b_{ij}^2} \] norm of matrix \( B = (b_{ij}) \), p. 125

\( C_2 \): class of functions \( V(t, x) \) continuously differentiable with respect to \( t \) and twice continuously differentiable with respect to \( x \), p. 57

\( C_2^0 \): set of functions \( V(x) \in C_2 \) with bounded second-order partial derivatives, p. 92

\( D_T \): domain of definition of generating operator, p. 31

\( D_L(t, x) \): domain of definition of generating operator \( L \) at point \( (t, x) \), p. 31

\( E_t \): euclidean l-space, p. 12

\( F_\varepsilon \): monotone family of \( \sigma \)-algebras, pp. 36, 59

\( J(x) \): Fisher information matrix, pp. 182, 186

\( J \): identity matrix, p. 124

\( L \): generating operator, pp. 31, 67

\( L_2[a, b] \): set of measurable random functions \( f(t, \omega), t \in [a, b], \omega \in \Omega, \mathcal{F}_t\)-measurable for every \( t \), such that \( \int_a^b f^2(t, \omega)dt \leq \infty \) with probability 1, p. 59

\( \mathcal{M}_\varphi \): \( \sigma \)-algebra of events generated by random variables \( X(u), u < t \), pp. 30, 73

\( \widetilde{U} \): closure of set \( U \), p. 162

\( U_\varepsilon(B) \): \( \varepsilon \)-neighborhood of set \( B \), p. 39

\( U_{\varepsilon,R}(B) = V_{\varepsilon}(B) \cap \{x: |x| < R\} \), p. 40

\( V_{\varepsilon}(B) = E_t \setminus U_\varepsilon(B) \), p. 39

\( \Xi \): domain of values of unknown parameter \( x \), p. 181

\( \mathcal{X}(t) \sim \mathcal{N}(m, \Sigma) \): asymptotic normality of \( \mathcal{X}(t) \), p. 123

\( \mathcal{Y}(n) = (y_1, \ldots, y_n) \), p. 202

\( \Xi \): domain of values of control parameter \( z \), p. 221

\( \nabla f(x) \): vector with coordinates \( [f(x + ce) - f(x - ce)]/2c \), p. 86

\( \partial \psi/\partial x \): matrix with elements \( \partial \psi_i/\partial x_j \), p. 107

\( \tau_C \): first exit time from domain \( G \), pp. 32, 73

\( \Phi(B) \): class of functions defined on p. 40

\[ 11^*1 = yfcjbf-fnor \]