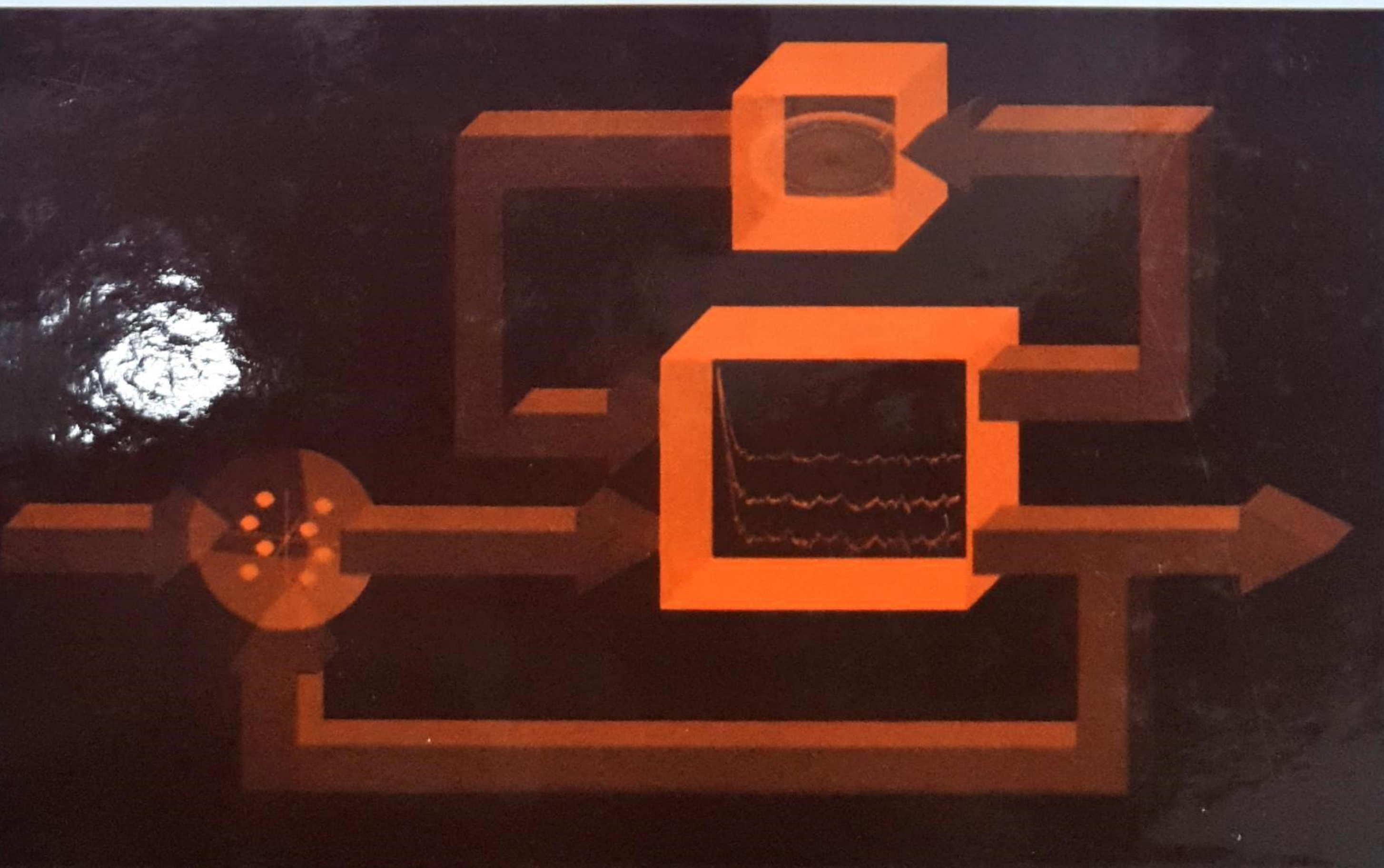
Fundamentals of Adaptive Filtering

ALI H. SAYED



Fundamentals of Adaptive Filtering

Ali H. Sayed
University of California, Los Angeles

IEEE
IEEE PRESS

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

CONTENTS

PREFACE

Adaptive filtering is a topic of immense practical relevance and deep theoretical challenges that persist even to this date. There are several notable texts on the subject that describe many of the features that have marvelled students and researchers over the years. In this textbook, we choose to step back and to take a broad look at the field. In so doing, we feel that we are able to bring forth, to the benefit of the reader, extensive commonalities that exist among different classes of adaptive algorithms and even among different filtering theories. We are also able to provide a uniform treatment of the subject in a manner that addresses some existing limitations, provides additional insights, and allows for extensions of current theory.

We do not have any illusions about the difficulties that arise in any attempt at understanding adaptive filters more fully. This is because adaptive filters are, by design, time-variant, nonlinear, and stochastic systems. Anyone of these qualifications alone would have resulted in a formidable system to study. Put them together and you face an almost impossible task. It is no wonder then that current practice tends to study different adaptive schemes separately, with techniques and assumptions that are usually more suitable for one adaptation form over another. It is also no surprise that most treatments of adaptive filters, including the one adopted in this textbook, need to rely on some simplifying assumptions in order to make filter analysis and design a more tractable objective.

Still, in our view, three desirable features of any study of adaptive filters would be (1) to attempt to keep the number of simplifying assumptions to a minimum, (2) to delay their use until necessary, and (3) to apply similar assumptions uniformly across different classes of adaptive algorithms. This last feature enables us to evaluate and compare the performance of adaptive schemes under similar assumptions on the data, while delaying the use of assumptions enables us to extract the most information possible about actual filter performance. In our discussions in this book we pay particular attention to these three features throughout the presentation.

In addition, we share the conviction that a thorough understanding of the performance and limitations of adaptive filters requires a solid grasp of the fundamentals of least-mean-squares estimation theory. These fundamentals help the designer understand what it is that an adaptive filter is trying to accomplish and how well it performs in this regard. For this reason, the first three chapters of the book are designed to provide the reader with a self-contained and easy-to-follow exposition of estimation theory, with a focus on topics that are relevant to the subject matter of the book. In these initial chapters, special emphasis is placed on geometric interpretations of several fundamental results. The reader is advised to pay close attention to these interpretations since it will become clear, time and again, that cumbersome algebraic manipulations can often be simplified by recourse to geometric constructions. These constructions not only provide a more lasting appreciation for the results of the book, but they also expose the reader to powerful tools that can be useful in other contexts as well, other than adaptive filtering and estimation theory.

The reader is further advised to master the convenience of the vector notation, which is used extensively throughout this book. Besides allowing a compact exposition of ideas and a compact representation of results, the vector notation also allows us to exploit to great effect several important results from linear algebra and matrix theory and to capture, in elegant ways, many revealing characteristics of adaptive filters. We cannot emphasize strongly enough the importance of linear algebraic and matrix tools in our presentation, as well as the elegance that they bring to the subject. The combined power of the geometric point of view and the vector notation are perhaps best exemplified by our detailed treatment later in this book of least-squares theory and its algorithmic variants. Of course, the reader is exposed to geometric and vector formulations in the early chapters of the book already, including the first chapter.

Style of the Book

Each chapter in the book consists generally of five distinctive parts in the following order: i) concepts, ii) bibliographic notes, iii) problems, iv) computer projects, and v) appendices.

i) **Concepts**. In the early chapters, each concept is motivated from first principles; starting from the obvious and ending with the more advanced. We follow this route of presentation until the reader develops enough maturity in the field. As the book progresses, we expect the reader to become more sophisticated and, therefore, we cut back on the "obvious". While for some advanced readers and researchers the "obvious" part in the initial chapters might seem at first unnecessary, please keep in mind that the primary readers of any textbook are novices to the field. From our experience over the years, teaching from early drafts of this manuscript, students have been particularly receptive to this line of presentation. In addition, for ease of reference, we have collected at the end of each chapter a summary of the key concepts and results.

ii) **Bibliographic Notes**. In the remarks at the end of each chapter we provide a wealth of references on the main contributors to the results discussed in the text. Rather than scatter references throughout the chapter, we find it useful to collect all references at the concluding section of each chapter in the form of a narrative. We believe that this way of presentation gives the reader a more focused perspective on how the references and the contributions relate to each other both in time and context.

iii) **Problems**. The book contains a significant number of problems, some more challenging than others and some more applied than others. The problems should be viewed as an *integral* part of the text, especially since many additional and interesting results appear in them. It was for this reason, and also for the benefit of the reader, that we have chosen to formulate and design all problems in a guided manner. Usually, and especially in the more challenging cases, a problem starts by stating its objective followed by a sequence of guided steps until the final answer is attained. The answer to each step appears stated in the body of the problem. In this way, a reader would know what the answer should be, even if he fails to solve the problem. Thus rather than ask the reader to "find an expression for x", we would instead ask him to "verify that x is given by $x = \ldots$" and then give the expression for x.

All instructors can request copies of a complete solutions manual from the publisher.

Moreover, several problems in the book have been designed to introduce readers to ideas of interest from related fields, such as multi-antenna receivers, cyclic-prefixing, maximal ratio combining, OFDM receivers, CDMA receivers, and so forth. Students are usually surprised to learn how classical concepts and ideas form the underpinnings of seemingly advanced techniques.

iv) **Computer Projects**. We have included several computer projects to show students, and also practitioners, how the results developed in the book can be useful in situations of practical interest (e.g., linear equalization, decision feedback equalization, channel estimation, beamforming, tracking fading channels, line echo cancellation, acoustic echo cancellation, active noise control, OFDM receivers, CDMA receivers, finite-precision implementations). In designing these projects, we have made an effort at choosing topics that are relevant to practitioners. We have also made an effort at illustrating to students how a solid theoretical understanding can guide them in challenging situations. All computer projects in the book are followed by extensive commentary and typical performance plots. Complete MATLAB[1] programs are available for solving all computer projects.

> Detailed MATLAB programs that solve all computer projects in the book
> can be downloaded by all readers from the publisher's website:
> ftp://ftp.wiley.com/public/sci_tech_med/filtering/

v) **Appendices**. Rather than collect all appendices at the end of the book, we have opted to place each appendix at the end of the chapter where it is called upon. In this way, the usefulness of the material in an appendix, and its relation to the discussion in the chapter, would become more evident to the reader. For example, although most students would have had some exposure to linear algebra and matrix theory before a course on adaptive filtering, we provide a handful of self-contained appendices that explain all the required concepts for the purposes of this book (e.g., rank and range spaces of matrices, solutions of linear equations, Schur complements, singular value decomposition, Cholesky decomposition, etc). Since each appendix is placed right where the concepts are first needed, students will be able to appreciate firsthand the elegance that such concepts bring to the presentation. Actually, after progressing sufficiently enough in the book, students will be able to master many useful concepts from linear algebra and matrix theory, in addition to adaptive filtering.

Organization of the Book

The material in the book can be categorized into five broad areas, as listed in Table P.1. Area I covers the fundamentals of least-mean-squares estimation theory with several application examples. Areas II and III deal mainly with LMS-type adaptive filters, while areas IV and V deal with least-squares-type adaptive filters. If an instructor wishes to focus mostly on LMS-type filters, then he can do so by covering only material from within areas II and III. Even in this case, students will still be exposed to the recursive-least-squares (RLS) algorithm and its performance results from the discussions in Chapter 5 and Area III. However, for a more-in-depth treatment of RLS and its many variants, instructors will need to select chapters from within Area IV as well.

Dependencies among chapters. Figure P.1 illustrates the dependencies among the chapters in the book. In the figure, the material in a chapter that is at the receiving end of an arrow requires some (but not necessarily all) of the material from the chapter at the origin of the

---

[1] MATLAB is a trademark of the MathWorks Inc.

Table P.1. A breakdown of the book chapters into general topic areas.

Category	Chapters
I. Introduction and Foundations	1. Optimal estimation. 2. Linear estimation. 3. Constrained linear estimation.
II. Stochastic–Gradient Methods	4. Steepest–descent algorithms. 5. Stochastic–gradient algorithms. 10. Block adaptive filters.
III. Performance Analyses	6. Steady-state performance of adaptive filters. 7. Tracking performance of adaptive filters. 8. Finite–precision effects. 9. Transient performance of adaptive filters.
IV. Least-Squares Methods	11. The least-squares criterion. 12. Recursive least–squares. 13. RLS array algorithms. 14. Fast fixed–order filters. 15. Lattice filters. 16. Laguerre adaptive filters.
V. Indefinite Least–Squares	17. Robust adaptive filters.

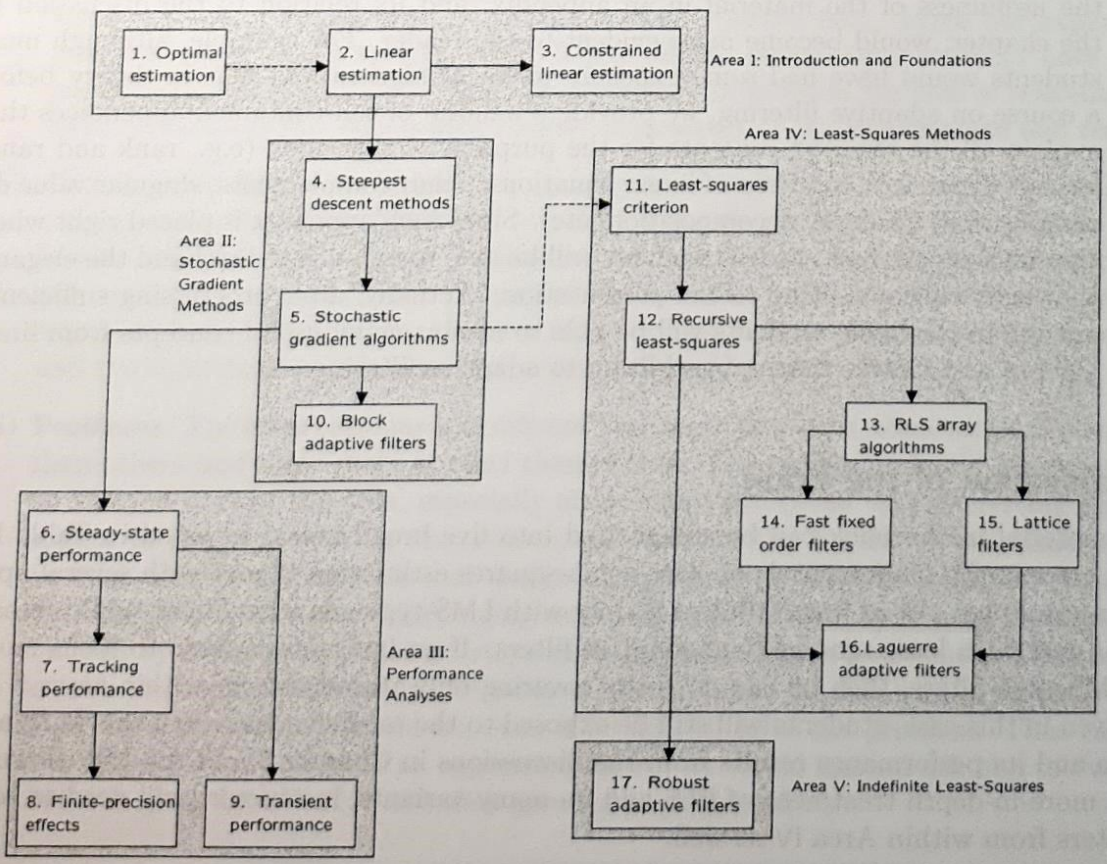


Figure P.1. Dependencies among the chapters. Instructors can design different course sequences in accordance with their needs and interests.

arrow. A dashed arrow indicates that the dependency between the respective chapters is weak and, if desired, the chapters can be covered independently of each other. For example, in order to cover Chapter 6, the instructor would need to cover Chapter 5, which, in turn, relies on Chapters 4 and 2. The material in Chapter 1 is not necessary for Chapter 2 but it is useful for a better understanding of it; actually, the material in Chapter 1 is presented in such a way that it also provides a useful review of basic probability theory concepts.

Figure P.1 can be used by instructors to design different course sequences according to their needs and interests. For example, if the instructor is interested in covering only LMS-type adaptive filters and in studying their performance, then one possibility is to cover material from within Chapters 2, 4, 5, 6, 7, and 9. Later, in Table P.4, we list the key sections from within each chapter; the other sections usually contain more advanced material, which students can read once they understand the key concepts from the main sections.

List of appendices. The book contains over 40 appendices that complement the material in the chapters and provide useful reviews and further analyses and connections. All appendices are listed in Table P.2. Readers interested in a quick review of basic linear algebra and matrix theory concepts may consult initially Apps. 1.A, 2.A, and 3.A, and subsequently Apps. 9.H, 11.B, 11.C, 13.A, 14.A, and 14.B for more advanced topics. These appendices are highlighted by the symbol * in Table P.2. Readers interested in a quick review of basic probability theory concepts should consult Secs. 1.1–1.4 and App. 1.B.

Computer projects. The book contains 24 computer projects that have been designed to reinforce the concepts discussed in the chapters. The projects are listed in Table P.3, and most of them cover topics of interest in communications and signal processing such as channel estimation, linear equalization (adaptive and channel-estimation based), decision-feedback equalization (also adaptive and channel-estimation based), adaptive blind equalization, CDMA and RAKE receivers, OFDM receivers, tracking of Rayleigh fading channels, line echo cancellation, acoustic echo cancellation, active noise control, beamforming, finite-precision effects, etc. Detailed MATLAB programs that solve all projects can be downloaded by all readers from the publisher's website (ftp://ftp.wiley.com/public/sci_tech_med/filtering/). These programs are offered without any guarantees. While we have found them to be effective for the instructional purposes of this textbook, the programs are not intended to be examples of full-blown or optimized designs; practitioners should use them at their own risk. For example, in order to keep the codes at a level that is easy to understand by students, we have often decided to sacrifice performance in lieu of simplicity.

Audience

The book is intended for a graduate-level course on adaptive filtering. Although it is beneficial that students have some familiarity with basic concepts from matrix theory, linear algebra, and random variables, the book includes several appendices on background material in these areas. The review is done in a motivated manner and is tailored to the needs of the presentation. From our experience, these reviews are sufficient for a thorough understanding of the discussions in the book. In addition, several of the problems reinforce the linear algebraic and matrix concepts, so much so that students will get valuable training in linear algebra and matrix theory, in addition to adaptive filtering, from reading (and understanding) this book.

The book is also intended to be a reference for researchers, which explains why we have chosen to include some advanced topics in several places. As a result, the book contains ample material for instructors to design courses according to their interests. Clearly, we do not expect instructors to cover all the material in the book in a typical course offering; such an objective would be counter-productive and even impossible. In our own teaching of

Table P.2. A listing of all appendices in the book. Appendices highlighted by the symbol * provide reviews of linear algebra and matrix theory concepts.

Appendix	Title
1.A*	Hermitian and positive-definite matrices.
1.B	Gaussian random variables.
2.A*	Range spaces and nullspaces of matrices.
2.B	Complex gradients and Hessians.
2.C	The Kalman filter.
3.A*	Schur complements.
3.B	A primer on channel equalization.
3.C	Causal Wiener-Hopf filtering.
6.A	Interpretations of the energy relation.
6.B	Relating $\epsilon-$NLMS to LMS.
6.C	Affine projection performance condition.
9.A	Stability bound.
9.B	Stability of $\epsilon-$NLMS.
9.C	Adaptive filters with error nonlinearities.
9.D	Convergence time of adaptive filters.
9.E	Learning behavior of adaptive filters.
9.F	Independence and averaging analysis.
9.G	Physical interpretation of energy relation.
9.H*	Kronecker products.
10.A	DCT-transformed regressors.
10.B	More constrained DFT block filters.
10.C	Overlap-add DFT-based block adaptive filters.
10.D	DCT-based block adaptive filters.
10.E	DHT-based block adaptive filters.
11.A	Equivalence results in linear estimation.
11.B*	The QR decomposition.
11.C*	The singular value decomposition.
12.A	Kalman filtering and recursive least-squares.
12.B	Extended RLS algorithms.
13.A*	Unitary transformations.
13.B	Array algorithms for Kalman filtering.
14.A*	Hyperbolic rotations.
14.B*	Hyperbolic basis rotations.
14.C	Backward consistency and minimality.
14.D	The Chandrasekhar filter.
16.A	Modeling with orthonormal basis functions.
16.B	Efficient matrix-vector multiplication.
16.C	Lyapunov equations.
17.A	Arbitrary coefficient matrices.
17.B	Total-least-squares.
17.C	$\mathcal{H}^{\infty}$ filters.
17.D	Stationary points.

Table P.3. A listing of all computer projects in the book. MATLAB programs that solve these projects can be downloaded by all readers from the publisher's website (ftp://ftp.wiley.com/public/sci_tech_med/filtering/).

Computer project	Topic
1	Comparing optimal and suboptimal estimators.
2	Linear equalization and decision devices.
3.1	Beamforming.
3.2	Decision-feedback equalization.
4.1	Constant-modulus criterion.
4.2	Linear prediction.
5.1	Constant-modulus algorithm.
5.2	Adaptive channel equalization.
5.3	Blind adaptive equalization.
6	Line echo cancellation.
7	Tracking Rayleigh fading channels.
8	Quantization effects in adaptive filtering.
9	Transient behavior of LMS and LMF.
10	Acoustic echo cancellation.
11.1	Amplitude tone detection.
11.2	An OFDM receiver.
11.3	CDMA and RAKE receivers.
12.1	Channel estimation with insufficient excitation.
12.2	Tracking a Rayleigh fading channel by extended RLS.
13	Performance of array implementations in finite precision.
14	Stability issues in fast least-squares.
15	Performance of lattice filters in finite precision.
16	Laguerre and FIR implementations.
17	Active noise control.

the material, we instead *focus on some key sections and request that students complement the discussions by means of reading and problem solving.* As explained below, several key sections in the chapters have been designed to convey the main concepts; while the remaining sections tend to include more advanced material and also illustrative examples. Once students understand the basic principles, you will be amazed at how well they can follow the other sections on their own and even solve the pertinent problems.

Guidelines to Instructors

As we explained before, instructors can use Fig. P.1 to design different course sequences according to their interests. For example, a course that is focused solely on LMS-type filters and their performance can be designed by covering only material from within Chapters 2 and 4–9. Even then, instructors do not need to cover the entire material from each one of these chapters. Instead, they need only cover some key sections and, if desired, ask students to complement the discussions in class with reading material from the other more advanced sections. To facilitate such a course planning, Table P.4 lists in boldface the key sections for the different chapters in the book for both lecturing and reading purposes.

For example, the key sections in Chapter 2 are Sec. 2.1 (Normal Equations), Sec. 2.4 (Orthogonality Condition), and Sec. 2.6 (Linear Models). These sections formulate and solve the linear least-mean-squares estimation problem and specialize the results to the important class of linear models (which is frequent in applications). The other sections in Chapter 2 complement the discussions with design examples, among other things.

As a second example, Chapter 9 studies the transient performance of a large family of adaptive filters in a uniform manner. The main idea is captured by the transient analysis of the LMS algorithm in Sec. 9.5, which uses the machinery developed in Sec. 9.4. Once students understand the framework as applied to LMS, they will be able to study the transient analysis of the other filters mostly on their own. This is one key advantage of adopting and emphasizing a uniform treatment of adaptive filter performance throughout our presentation. Similar remarks hold for the steady-state, tracking, and finite-precision performance analyses of Chapters 6–8. It is sufficient to illustrate how the methodology applies to the special case of LMS, for example, by covering Secs. 6.5, 7.5, and 8.5, which in turn rely on the machinery developed in Secs. 6.4, 7.4, and 8.4. The remaining sections in Chapters 6–8 extend the same type of analysis to other (more demanding) adaptive filters. Here again, students can do well in studying the extensions on their own if desired.

Table P.4. A suggested list of key sections (in boldface) for both lecturing and reading in all chapters along with relevant complementary sections (in normal font). At the instructor's discretion, some of the key sections for reading could, of course, be covered during lecturing as well; especially those dealing with basic review material on linear algebraic and matrix theory concepts.

Key sections for lecturing	Key sections for reading
Secs. 1.1, 1.2, **1.3**, **1.4**	App. **1.A**
Secs. **2.1**, 2.2, 2.3, **2.4**, 2.5, **2.6**	Apps. **2.A**, 2.B, 2.C
Secs. **3.1**, 3.2, **3.3**, 3.4	Apps. **3.A**, 3.B
Secs. 4.1, **4.2**, **4.3**, 4.4, 4.5	
Secs. 5.1, **5.2**, 5.3–5.6, **5.9**	Secs. **5.6**, 5.7, 5.8, 5.10
Secs. 6.1, 6.2, **6.3–6.5**	Secs. 6.6–6.10
Secs. 7.1,7.2, **7.3–7.5**	Secs. 7.6–7.11
Secs. **8.1–8.5**	Secs. 8.6–8.9
Secs. 9.1–9.3, **9.4**, **9.5**	Secs. 9.6–9.7 Apps. 9.A, 9.B, **9.C**, **9.D**, 9.E
Secs. **10.1**, 10.2, 10.3, **10.4**, **10.5**	Apps. 10.A–10.E
Secs. **11.1–11.4**	Sec. **11.5**, Apps. **11.A**–11.C
Secs. 12.1–**12.3**	Sec. **12.4**, Apps. **12.A–12.B**
Secs. 13.1–13.3, **13.5**, **13.6**, App. **13.A**	Sec. 13.7
Secs. **14.1**	Secs. **14.2**, **14.3**–14.6 Apps. 14.A–**14.B**
Secs. 15.1, **15.2–15.4**, 15.5, **15.6**, 15.7	Secs. 15.8, **15.9**–15.13
Secs. **16.1–16.2**, 16.8, 16.9	Secs. 16.3–16.7, 16.10–14 Apps. **16.A**–16.C
Secs. **17.1–17.5**	Apps. 17.A–17.D

Some Features of Our Treatment

There are some distinctive features in our treatment of adaptive filtering. Among other features, experts will be able to notice the following contributions:

(a) We treat a large variety of adaptive algorithms, as listed in Tables P.5 and P.6 for both LMS-type and RLS-type filters.

Table P.5. A list of LMS-type adaptive algorithms covered in the book.

Algorithm	Description
LMS	Least-mean-squares algorithm
NLMS	Normalized LMS
NLMS with power normalization	
leaky-LMS	
constrained LMS	
LMF	Least-mean-fourth algorithm
LMMN	Least-mean-mixed norm algorithm
sign-error LMS	
sign-regressor LMS	
sign-sign LMS	
FxLMS	Filtered-x LMS
FeLMS	Filtered-error LMS
APA	Affine projection algorithm
CMA	Constant modulus algorithm
NCMA	Normalized CMA
RCA	Reduced constellation algorithm
MMA	Multi-modulus algorithm
DFT-domain LMS	Transform-domain LMS
DCT-domain LMS	
DFT-based block LMS	
DCT-based block LMS	
DHT-based block LMS	
closed and open-loop subband LMS	
Robust filters	*A priori* and *a posteriori* forms

(b) Chapters 6–9 study the mean-square performance of adaptive filters by resorting to energy-conservation arguments. While the performance of different adaptive filters is usually studied separately in the literature, the framework adopted in these chapters applies uniformly across different classes of adaptive filters. In addition, the same framework is used for steady-state analysis, transient analysis, tracking analysis, fixed-point analysis, and robustness analysis.

(c) Chapter 10 studies block adaptive filters, and the related class of subband adaptive filters, in a manner that clarifies the connections between these two families more directly than prior treatments. Our presentation also indicates how to move beyond DFT-based transforms and how to use other classes of orthogonal transforms for block adaptive filtering.

(d) Chapters 11-15 provide a detailed treatment of least-squares adaptive filters that is distinct from prevailing approaches in a handful of respects. First, we focus on regularized least-squares problems from the onset and take the regularization factor into account in all derivations. Second, we insist on deriving time- and order-update relations independent of any structure in the regression data (e.g., we do not require the regressors to arise from a tapped-delay-line implementation). In this way, we are able to develop efficient least-squares filtering even for some non-FIR structures. Third, we emphasize the role and benefits of array-based schemes. And, finally, we highlight the role of geometric constructions and the insights they bring into least-squares theory.

Table P.6. A list of RLS-type adaptive algorithms covered in the book.

Algorithm	Description
RLS	Recursive least-squares algorithm
GN	Gauss-Newton algorithm
sliding-window RLS	
block RLS	Multi-channel RLS
extended RLS	
QR-RLS (square-root information RLS)	Array-based recursive least-squares
inverse-QR RLS (square-root RLS)	
Fast array RLS	
FTF	Fast transversal filter
FAEST	Fast *a posteriori* error sequential technique
Fast Kalman	
a posteriori-based LSL	Least-squares lattice
a priori-based LSL	Least-squares lattice
a posteriori error-feedback LSL	
a priori error-feedback LSL	
Normalized LSL	
Array-based LSL	
Laguerre FTF	Extended FTF
Laguerre FAEST	Extended FAEST
Laguerre fast Kalman	
a posteriori-based Laguerre LSL	Least-squares Laguerre lattice
a priori-based Laguerre LSL	
a posteriori error-feedback Laguerre LSL	
a priori error-feedback Laguerre LSL	
Normalized Laguerre LSL	
Array-based Laguerre LSL	
Kalman filter	
Robust filters	*A priori* and *a posteriori* forms

(e) Chapter 16 shows how the theory of fast least-squares methods (both for fixed-order and order-recursive problems) is not limited to tapped delay lines; an observation that extends classical derivations and developments. In the chapter, we illustrate this fact by studying Laguerre adaptive filters; they are obtained by replacing the delay operators in an FIR structure by first-order all-pass functions. Although the resulting regressors no longer possess shift-structure, it turns out that fast least-squares filters are still possible.

(f) Chapter 17 develops the theory of robust adaptive filters by studying indefinite least-squares problems and by relying on energy arguments as well. In the process, the robustness and optimality properties of several adaptive filters are clarified. The presentation in this chapter is developed in a manner that parallels our treatment of least-squares problems in Chapters 11–12 so that readers can appreciate the similarities and distinctions between both theories (classical least-squares versus indefinite least-squares).

Westwood, CA
December 2002

Ali H. Sayed