OPTIMAL SIGNAL PROCESSING UNDER UNCERTAINTY

OPTIMAL SIGNAL PROCESSING UNDER UNCERTAINTY

Edward R. Dougherty

Published by

SPIE.

To Edward Townes Dougherty

Contents

Preface

Whereas modern science concerns the mathematical modeling of phenomena, essentially a passive activity, modern engineering involves determining operations to actively alter phenomena to effect desired changes of behavior. It begins with a scientific (mathematical) model and applies mathematical methods to derive a suitable intervention for the given objective. Since one would prefer the best possible intervention, engineering inevitably becomes optimization and, since all but very simple systems must account for randomness, modern engineering might be defined as the study of optimal operators on random processes. The seminal work in the birth of modern engineering is the Wiener–Kolmogorov theory of optimal linear filtering on stochastic processes developed in the 1930s. As Newton's laws constitute the gateway into modern science, the Wiener–Kolmogorov theory is the gateway into modern engineering.

The design of optimal operators takes different forms depending on the random process constituting the scientific model and the operator class of interest. The operators might be linear filters, morphological filters, controllers, classifiers, or cluster operators, each having numerous domains of application. The underlying random process might be a random signal/ image for filtering, a Markov process for control, a feature-label distribution for classification, or a random point set for clustering. In all cases, operator class and random process must be united in a criterion (cost function) that characterizes the operational objective and, relative to the criterion, an optimal operator found. For the classical Wiener filter, the model is a pair of jointly distributed wide-sense stationary random signals, the objective is to estimate a desired signal from an observed signal via a linear filter, and the cost function to be minimized is the mean-square error between the filtered observation and the desired signal.

Besides the mathematical and computational issues that arise with classical operator optimization, especially with nonlinear operators, nonstationary processes, and high dimensions, a more profound issue is uncertainty in the scientific model. For instance, a long recognized problem with linear filtering is incomplete knowledge regarding the covariance functions (or power spectra in the case of Wiener filtering). Not only must optimization be

relative to the original cost function, be it mean-square error in linear filtering or classification/clustering error in pattern recognition, optimization must also take into account uncertainty in the underlying random process. Optimization is no longer relative to a single random process but instead relative to an uncertainty class of random processes. This means the postulation of a new cost function integrating the original cost function with the model uncertainty. If there is a prior distribution (or posterior distribution if data are employed) governing likelihood in the uncertainty class, then one can choose an operator from some class of operators that minimizes the expected cost over the uncertainty class. In the absence of a prior distribution, one might take a minimax approach and choose an operator that minimizes the maximum cost over the uncertainty class.

A prior (or posterior) distribution places the problem in a Bayesian framework. A critical point, and one that will be emphasized in the text, is that the prior distribution is not on the parameters of the operator model, but on the unknown parameters of the scientific model. This is natural. If the model were known with certainty, then one would optimize with respect to the known model; if the model is uncertain, then the optimization is naturally extended to include model uncertainty and the prior distribution should be on that uncertainty. For instance, in the case of linear filtering the covariance function might be uncertain, meaning that some of its parameters are unknown, in which case the prior distribution characterizes uncertainty relative to the unknown parameters.

A basic principle embodied in the book is to express an optimal operator under the joint probability space formed from the joint internal and external uncertainty in the same form as an optimal operator for a known model by replacing the mathematical structures characterizing the standard optimal operator with corresponding structures, called *effective* structures, that incorporate model uncertainty. For instance, in Wiener filtering the power spectra are replaced by effective power spectra, in Kalman filtering the Kalman gain matrix is replaced by the effective Kalman gain matrix, and in classification the class-conditional distributions are replaced by effective class-conditional distributions.

The first three chapters of the book review those aspects of random processes that are necessary for developing optimal operators under uncertainty. Chapter 1 covers random functions, including the moments and the calculus of random functions. Chapter 2 treats canonical expansions for random functions, a topic often left uncovered in basic courses on stochastic processes. It treats discrete expansions within the context of Hilbert space theory for random functions, in particular, the equivalence of canonical expansions of the random function and the covariance function entailed by Parseval's identity. It then goes on to treat integral canonical expansions in the framework of generalized functions. Chapter 3 covers the basics of

classical optimal filtering: optimal finite-observation linear filtering, optimal infinite-observation linear filtering via the Wiener–Hopf integral equation, Wiener filtering for wide-stationary processes, recursive (Kalman) filtering via direct-sum decomposition of the evolving observation space, and optimal morphological filtering via granulometric bandpass filters.

For the most part, although not entirely, the first three chapters are a compression of Chapters 2 through 4 of my book *Random Processes for Image and Signal Processing*, aimed directly at providing a tight background for optimal signal processing under uncertainty, the goal being to make a one-semester course for Ph.D. students. Indeed, the book has been developed from precisely such a course, attended by Ph.D. students, post-doctoral students, and faculty.

Chapter 4 covers optimal robust filtering. The first section lays out the basic definitions for intrinsically Bayesian robust (IBR) filtering, the fundamental principle being filter optimization with respect to both internal model stochasticity and external model uncertainty, the latter characterized by a prior distribution over an uncertainty class of random-process models. The first section introduces the concepts of effective process and effective characteristic, whereby the structure of the classical solutions is retained with characteristics such as the power spectra and the Wiener–Hopf equation generalized to effective power spectra and the effective Wiener–Hopf equation, which are relative to the uncertainty class. Section 4.2 covers optimal Bayesian filters, which are analogous to IBR filters except that new observations are employed to update the prior distribution to a posterior distribution. Section 4.3 treats model-constrained Bayesian robust (MCBR) filters, for which optimization is restricted to filters that are optimal for some model in the uncertainty class. In Section 4.4 the term "robustness" is defined quantitatively via the loss of performance and is characterized for linear filters in the context of integral canonical expansions, where random process representation is now parameterized via the uncertainty. Section 4.5 reviews classical minimax filtering and applies it to minimax morphological filtering. Sections 4.6 and 4.7 extend the classical Kalman (discrete time) and Kalman–Bucy (continuous time) recursive predictors and filters to the IBR framework, where classical concepts such as the Kalman gain matrix get extended to their effective counterparts (effective Kalman gain matrix).

When there is model uncertainty, a salient issue is the design of experiments to reduce uncertainty; in particular, which unknown parameter should be determined to optimally reduce uncertainty. To this end, Section 5.1 introduces the mean objective cost of uncertainty (MOCU), which is the expected cost increase relative to the objective resulting from the uncertainty, expectation being taken with respect to the prior (posterior) distribution. Whereas entropy is a global measure of uncertainty not related to any particular operational objective, MOCU is based directly on the engineering

objective. Section 5.2 analyzes optimal MOCU-based experimental design for IBR linear filtering. Section 5.3 revisits Karhunen–Loève optimal compression when there is model uncertainty, and therefore uncertainty as to the Karhunen–Loève expansion. The IBR compression is found and optimal experimental design is analyzed relative to unknown elements of the covariance matrix. Section 5.4 discusses optimal intervention in regulatory systems modeled by Markov chains when the transition probability matrix is uncertain and derives the experiment that optimally reduces model uncertainty relative to the objective of minimizing undesirable steady-state mass. The solution is computationally troublesome, and the next section discusses complexity reduction. Section 5.6 examines sequential experimental design, both greedy and dynamic-programming approaches, and compares MOCU-based and entropy-based sequential design. To this point, the chapter assumes that parameters can be determined exactly. Section 5.7 addresses the issue of inexact measurements owing to either experimental error or the use of surrogate measurements in place of the actually desired measurements, which are practically unattainable. The chapter closes with a section on a generalized notion of MOCU-based experimental design, a particular case being the knowledge gradient.

The optimal Bayesian filter paradigm was first introduced in classification with the design of optimal Bayesian classifiers (OBCs). Classically (Section 6.1), if the feature-label distribution is known, then an optimal classifier, one that minimizes classification error, is found as the Bayes classifier. As discussed in Section 6.2, when there are unknown parameters in the feature-label distribution, then there is an uncertainty class of feature-label distributions, and an optimal Bayesian classifier minimizes the expected error across the uncertainty class relative to the posterior distribution derived from the prior and the sample data. In order to compare optimal Bayesian classification with classical methods, Section 6.3 reviews the methodology of classification rules based solely on data. Section 6.4 derives the OBC in the discrete and Gaussian models. Section 6.5 examines consistency, that is, convergence of the OBC as the sample size goes to infinity. Rather than sample randomly or separately (randomly given the class sizes), sampling can be done in a nonrandom fashion by iteratively deciding which class to sample from prior to the selection of each point or by deciding which feature vector to observe. Optimal sequential sampling in these paradigms is discussed in Section 6.7 using MOCU-based experimental design. Section 6.8 provides a general framework for constructing prior distributions via optimization of an objective function subject to knowledge-based constraints. Epistemological issues regarding classification are briefly discussed in Section 6.9.

Clustering shares some commonality with classification in that both involve operations on points, classification on single points and clustering on point sets, and both have performances measured via a natural definition of

error, classification error for the former and cluster (partition) error for the latter. But they also differ fundamentally in that the underlying random process for classification is a feature-label distribution, and for clustering it is a random point set. Section 7.1 describes some classical clustering algorithms. Section 7.2 discusses the probabilistic foundation of clustering and optimal clustering (Bayes cluster operator) when the underlying random point set is known. Section 7.3 describes a special class of random point sets that can be used for modeling in practical situations. Finally, Section 7.4 discusses IBR clustering, which involves optimization over both the random set and its uncertainty when the random point set is unknown and belongs to an uncertainty class of random point sets. Whereas with linear and morphological filtering the structure of the IBR or optimal Bayesian filter essentially results from replacing the original characteristics with effective characteristics, optimal clustering cannot be so conveniently represented. Thus, the entire random point process must be replaced by an effective random point process.

Let me close this preface by noting that this book views science and engineering teleologically: a system within Nature is modeled for a purpose; an operator is designed for a purpose; and an optimal operator is obtained relative to a cost function quantifying the achievement of that purpose. Right at the outset, with model formation, purpose plays a critical role. As stated by Erwin Schrödinger (Schrödinger, 1957), "A selection has been made on which the present structure of science is built. That selection must have been influenced by circumstances that are other than purely scientific.... The origin of science [is] without any doubt the very anthropomorphic necessity of man's struggle for life." Norbert Wiener, whose thinking is the genesis behind this book, states (Rosenblueth and Wiener, 1945) this fundamental insight from the engineering perspective: "The intention and the result of a scientific inquiry is to obtain an understanding and a control of some part of the universe." It is not serendipity that leads us inexorably from optimal operator representation to optimal experimental design. This move, too, is teleological, as Wiener makes perfectly clear (Rosenblueth and Wiener, 1945): "An experiment is a question. A precise answer is seldom obtained if the question is not precise; indeed, foolish answers — i.e., inconsistent, discrepant or irrelevant experimental results — are usually indicative of a foolish question." Only in the context of optimization can one know the most relevant questions to ask Nature.

Edward R. Dougherty
College Station, Texas
June 2018

Acknowledgments

First, let me recognize the large contributions of my former students Lori Dalton and Roozbeh Dehghannasiri to the theory developed in this book. Many other students have made contributions: Mohammad Sharokh Esfahani, Shahin Boluki, Mahdi Imani, Amin Zollanvari, Jason Knight, Mohammadmahdi Yousefi, Ranadip Pal, Yidong Chen, Marcel Brun, Marco Benalcazar, and Artyom Grigoryan. Notable contributions have been made by my colleagues Xiaoning Qian and Byung-Jun Yoon. Let me also thank Corey Barrett, who created the unified file for the book, and Alireza Karbalayghareh, who updated that file as I modified the manuscript.