Smoothing, Filtering and Prediction:
Estimating the Past, Present and Future
Second Edition
Garry A. Einicke

Prime Publishing – Amazon.com
Second Edition published February 2019

ISBN 978-0-6485115-1-9 Kindle format
ISBN 978-0-6485115-0-2 Printable pdf format

Preface

Scientists, engineers and the like are a strange lot. Unperturbed by societal norms, they direct their energies to finding better alternatives to existing theories and concocting solutions to unsolved problems. Driven by an insatiable curiosity, they record their observations and crunch the numbers. This tome is about the science of crunching. It's about digging out something of value from the detritus that others tend to leave behind. The described approaches involve constructing models to process the available data. Smoothing entails revisiting historical records in an endeavour to understand something of the past. Filtering refers to estimating what is happening currently, whereas prediction is concerned with hazarding a guess about what might happen next.

The basics of smoothing, filtering and prediction were worked out by Norbert Wiener, Rudolf E. Kalman and Richard S. Bucy *et al* over half a century ago. This book describes the classical techniques together with some more recently developed embellishments for improving performance within applications. Its aims are threefold. First, to present the subject in an accessible way, so that it can serve as a practical guide for undergraduates and newcomers to the field. Second, to differentiate between techniques that satisfy performance criteria versus those relying on heuristics. Third, to draw attention to Wiener's approach for optimal non-causal filtering (or smoothing).

Optimal estimation is routinely taught at a post-graduate level while not necessarily assuming familiarity with prerequisite material or backgrounds in an engineering discipline. That is, the basics of estimation theory can be taught as a standalone subject. In the same way that a vehicle driver does not need to understand the workings of an internal combustion engine or a computer user does not need to be acquainted with its inner workings, implementing an optimal filter is hardly rocket science. Indeed, since the filter recursions are all known – its operation is no different to pushing a button on a calculator. The key to obtaining good estimator performance is developing intimacy with the application at hand, namely, exploiting any available insight, expertise and *a priori* knowledge to model the problem. If the measurement noise is negligible, any number of solutions may suffice. Conversely, if the observations are dominated by measurement noise, the problem may be too hard. Experienced practitioners are able recognise those intermediate sweet-spots where cost-benefits can be realised.

Systems employing optimal techniques pervade our lives. They are embedded within medical diagnosis equipment, communication networks, aircraft avionics, robotics and market forecasting – to name a few. When tasked with new problems, in which information is to be extracted from noisy measurements, one can be faced with a plethora of algorithms and techniques. Understanding the performance of candidate approaches may seem unwieldy and daunting to novices. Therefore, the philosophy here is to present the linear-quadratic-Gaussian results for smoothing, filtering and prediction with accompanying proofs about performance being attained, wherever this is appropriate. Unfortunately, this does require some maths which trades off accessibility. The treatment is little repetitive and may seem trite, but hopefully it contributes an understanding of the conditions under which solutions can value-add.

Science is an evolving process where what we think we know is continuously updated with refashioned ideas. Although evidence suggests that Babylonian astronomers were able to predict planetary motion, a bewildering variety of Earth and universe models followed. According to lore, ancient Greek philosophers such as Aristotle assumed a geocentric model of the universe and about two centuries later Aristarchus developed a heliocentric version. It is reported that Eratosthenes arrived at a good estimate of the Earth's circumference, yet there was a revival of flat earth beliefs during the middle ages. Not all ideas are welcomed - Galileo was famously incarcerated for knowing too much. Similarly, newly-appearing signal processing techniques compete with old favourites. An aspiration here is to publicise that the oft forgotten approach of Wiener, which in concert with Kalman's, leads to optimal smoothers. The ensuing results contrast with traditional solutions and may not sit well with more orthodox practitioners.

Kalman's optimal filter results were published in the early 1960s and various techniques for smoothing in a state-space framework were developed shortly thereafter. Wiener's optimal smoother solution is less well known, perhaps because it was framed in the frequency domain and described in the archaic language of the day. His work of the 1940s was borne of an analog world where filters were made exclusively of lumped circuit components. At that time, computers referred to people labouring with an abacus or an adding machine – Alan Turing's and John von Neumann's ideas had yet to be realised. In his book, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiener wrote with little fanfare and dubbed the smoother "unrealisable". The use of the Wiener-Hopf factor allows this smoother to be expressed in a time-domain state-space setting and included alongside other techniques within the designer's toolbox.

A model-based approach is employed throughout where estimation problems are defined in terms of state-space parameters. I recall attending Michael Green's robust control course, where he referred to a distillation column control problem competition, in which a student's robust low-order solution out-performed a senior specialist's optimal high-order solution. It is hoped that this text will equip readers to do similarly, namely: make some simplifying assumptions, apply the standard solutions and back-off from optimality if uncertainties degrade performance.

Both continuous-time and discrete-time techniques are presented. Sometimes the state dynamics and observations may be modelled exactly in continuous-time. In the majority of applications, some discrete-time approximations and processing of sampled data will be required. The material is organised as an eleven-lecture course.

- Chapter 1 introduces some standard continuous-time fare such as the Laplace Transform, stability, adjoints and causality. A completing-the-square approach is then used to obtain the minimum-mean-square error (or Wiener) filtering solutions.

- Chapter 2 deals with discrete-time minimum-mean-square error filtering. The treatment is somewhat brief since the developments follow analogously from the continuous-time case.

- Chapter 3 describes continuous-time minimum-variance (or Kalman-Bucy) filtering. The filter is found using the conditional mean or least-mean-square-error formula. It is shown for time-invariant problems that the Wiener and Kalman solutions are the same.

- Chapter 4 addresses discrete-time minimum-variance (or Kalman) prediction and filtering. Once again, the optimum conditional mean estimate may be found via the least-mean-square-error approach. Generalisations for missing data, deterministic inputs, correlated noises, direct feedthrough terms, output estimation and equalisation are described.

- Chapter 5 simplifies the discrete-time minimum-variance filtering results for steady-state problems. Discrete-time observability, Riccati equation solution convergence, asymptotic stability and Wiener filter equivalence are discussed.

- Chapter 6 covers the subject of continuous-time smoothing. The main fixed-lag, fixed-point and fixed-interval smoother results are derived. It is shown that the minimum-variance fixed-interval smoother attains the best performance.

- Chapter 7 is about discrete-time smoothing. It is observed that the fixed-point fixed-lag, fixed-interval smoothers outperform the Kalman filter. Once again, the minimum-variance smoother attains the best-possible performance, provided that the underlying assumptions are correct.

- Chapter 8 attends to parameter estimation. As the above-mentioned approaches all rely on knowledge of the underlying model parameters, maximum-likelihood estimation techniques and expectation-maximisation algorithms are described. The addition of a correction term yields unbiased, consistent state-space parameter estimates that attain the Cramer-Rao Lower Bounds.

- Chapter 9 is concerned with robust techniques that accommodate uncertainties within problem specifications. An extra term within the design Riccati equations enables designers to trade-off average error and peak error performance.

- Chapter 10 applies the afore-mentioned linear techniques to nonlinear estimation problems. It is demonstrated that step-wise linearisations can be used within predictors, filters and smoothers, albeit by forsaking optimal performance guarantees.

- Chapter 11 rounds off the course by exploiting knowledge about transition probabilities. HMM, minimum-variance-HMM and high-order minimum-variance-HMM filters and smoothers are derived. The improved performance offered by these techniques needs to be reconciled against the significantly higher calculation overheads.

The foundations are laid in Chapters 1 – 2, which explain minimum-mean-square-error solution construction and asymptotic behaviour. In single-input-single-output cases, finding Wiener filter transfer functions may have appeal. In general, designing Kalman filters is more tractable because solving a Riccati equation is easier than pole-zero cancellation. Kalman filters are needed if the signal models are time-varying. The filtered states can be updated via a one-line recursion but the gain may require to be re-evaluated at each step in time. Extended Kalman filters are contenders if nonlinearities are present. Smoothers are advocated when better performance is desired and some calculation delays can be tolerated. Using additional transition probability information can be advantageous wherever measurements exhibit reoccurring patterns.

This book elaborates on several articles published in *IEEE* journals and I am grateful to my collaborators and the anonymous reviewers who have improved my efforts over the years. Lang White continues to teach and motivate me to this day. The great people at the CSIRO, such as David Hainsworth and George Poropat generously make themselves available to anglicise my engineering jargon. Sometimes posing good questions is helpful, for example, Paul Malcolm once asked “is it stable?” which led down to fruitful paths. During a seminar at HSU, Udo Zoelzer provided the impulse for me to undertake this project. My sources of inspiration include interactions at the CDC meetings - thanks particularly to Dennis Bernstein whose passion for writing has motivated me along the way.

Garry Einicke,
CSIRO Australia

Contents

1. Continuous-Time, Minimum-Mean-Square-Error Filtering

1.1 Introduction

Optimal filtering is concerned with designing the best linear system for recovering data from noisy measurements. It is a model-based approach requiring knowledge of the signal generating system. The signal models, together with the noise statistics are factored into the design in such a way to satisfy an optimality criterion, namely, minimising the square of the error.

A prerequisite technique, the method of least-squares, has its origin in curve fitting. Amid some controversy, Kepler claimed in 1609 that the planets move around the Sun in elliptical orbits [1]. Carl Freidrich Gauss arrived at a better performing method for fitting curves to astronomical observations and predicting planetary trajectories in 1799 [1]. He formally published a least-squares approximation method in 1809 [2], which was developed independently by Adrien-Marie Legendre in 1806 [1]. This technique was famously used by Giusseppe Piazzi to discover and track the asteroid Ceres using a least-squares analysis which was easier than solving Kepler's complicated nonlinear equations of planetary motion [1]. Andrey N. Kolmogorov refined Gauss's theory of least-squares and applied it for the prediction of discrete-time stationary stochastic processes in 1939 [3]. Norbert Wiener, a faculty member at MIT, independently solved analogous continuous-time estimation problems. He worked on defence applications during the Second World War and produced a report entitled *Extrapolation, Interpolation and Smoothing of Stationary Time Series* in 1943. The report was later published as a book in 1949 [4].

Wiener derived two important results, namely, the optimum (non-causal) minimum-mean-square-error solution and the optimum causal minimum-mean-square-error solution [4] – [6]. The optimum causal solution has since become known at the Wiener filter and in the time-invariant case is equivalent to the Kalman filter that was developed subsequently. Wiener pursued practical outcomes and attributed the term "unrealisable filter" to the optimal non-causal solution because "it is not in fact realisable with a finite network of resistances, capacities, and inductances" [4]. Wiener's unrealisable filter is actually the optimum linear smoother.

"All men by nature desire to know." *Aristotle*

The optimal Wiener filter is calculated in the frequency domain. Consequently, Section 1.2 touches on some frequency-domain concepts. In particular, the notions of spaces, state-space systems, transfer functions, canonical realisations, stability, causal systems, power spectral density and spectral factorisation are introduced. The Wiener filter is then derived by minimising the square of the error. Three cases are discussed in Section 1.3. First, the solution to general estimation problem is stated. Second, the general estimation results are specialised to output estimation. The optimal input estimation or equalisation solution is then described. An example, demonstrating the recovery of a desired signal from noisy measurements, completes the chapter.

1.2 Prerequisites

1.2.1 Signals

Let $(.)^T$ denote the transpose operator. Consider a continuous-time, stochastic (or random) signal $w(t) = [w_1(t), w_2(t), \ldots, w_n(t)]^T$, with $w_i(t) \in \mathbb{R}$, $i = 1, \ldots n$, which belongs to the space $\mathbb{R}^n$, or more concisely $w(t) \in \mathbb{R}^n$. In general, where $n > 1$, the set of $w(t)$ over all time t is denoted by the matrix $w = [\, w(-\infty), \ldots, w(\infty)]$. Often the focus is on scalar signals, in which case $w = [\, w(-\infty), \ldots, w(\infty)]$ is a row vector.

1.2.2 Elementary Functions Defined on Signals

The inner product $\langle v, w\rangle$ of two continuous-time signal vectors v and w is defined by

$$\langle v, w\rangle = \int_{-\infty}^{\infty} v^T w \, dt . \qquad (1)$$

The 2-norm or Euclidean norm of a continuous-time signal vector w, is defined as $\|w\|_2 = \sqrt{\langle w, w\rangle} = \sqrt{\int_{-\infty}^{\infty} w^T w dt}$. The square of the 2-norm, that is, $\|w\|_2^2 = \langle w^T w\rangle = \int_{-\infty}^{\infty} w^T w \, dt$ is commonly known as energy of the signal w.

"Scientific discovery consists in the interpretation for our own convenience of a system of existence which has been made with no eye to our convenience at all." *Norbert Wiener*

1.2.3 Spaces

The Lebesgue 2-space, defined as the set of continuous-time signals having finite 2-norm, is denoted by $\mathcal{L}_2$. Thus, $w \in \mathcal{L}_2$ means that the energy of w is bounded. The following properties hold for 2-norms.

(i) $\|v\|_2 = 0 \Rightarrow v = 0$.

(ii) $\|\alpha v\|_2 = |\alpha| \; \|v\|_2$.

(iii) $\|v + w\|_2 \le \|v\|_2 + \|w\|_2$, which is known as the triangle inequality.

(iv) $\|vw\|_2 \le \|v\|_2 \, \|w\|_2$.

(v) $|\langle v, w\rangle| \le \|v\|_2 \, \|w\|_2$, which is known as the Cauchy-Schwarz inequality.

See [8] for more detailed discussions of spaces and norms.

1.2.4 Linear Systems

A linear system is defined as having an output vector which is equal to the value of a linear operator applied to an input vector. That is, the relationships between the output and input vectors are described by linear equations, which may be algebraic, differential or integral. Linear time-domain systems are denoted by upper-case script fonts. Consider two linear systems $\mathcal{G} : w \to \mathcal{G}w$, $\mathcal{H} : w \to \mathcal{H}w$, that is, they operate on an input w and produce outputs $\mathcal{G}w$, $\mathcal{H}w$. The following properties hold.

$$(\mathcal{G} + \mathcal{H})\, w = \mathcal{G}w + \mathcal{H}w, \tag{2}$$

$$(\mathcal{G}\mathcal{H})\, w = \mathcal{G}(\mathcal{H}w), \tag{3}$$

$$(\alpha\mathcal{G})\, w = \alpha(\mathcal{G}w), \tag{4}$$

where $\alpha \in \mathbb{R}$. An interpretation of (2) is that a parallel combination of $\mathcal{G}$ and $\mathcal{H}$ is equivalent to the system $\mathcal{G} + \mathcal{H}$. From (3), a series combination of $\mathcal{G}$ and $\mathcal{H}$ is equivalent to the system $\mathcal{G}\mathcal{H}$. Equation (4) states that scalar amplification of a system is equivalent to scalar amplification of a system's output.

1.2.5 Polynomial Fraction Systems

The Wiener filtering results [4] – [6] were originally developed for polynomial fraction descriptions of systems which are described below. Consider an n^{th}-order linear, time-invariant system $\mathcal{G}$ that operates on an input $w(t) \in \mathbb{R}$ and produces an output $y(t) \in \mathbb{R}$. Suppose that the differential equation model for this system is

---

"Science is a way of thinking much more than it is a body of knowledge." *Carl Edward Sagan*

$$a_n \frac{d^n y(t)}{dt^n} + a_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + ... + a_1 \frac{dy(t)}{dt} + a_0 y(t)$$

$$= b_m \frac{d^m w(t)}{dt^m} + b_{m-1} \frac{d^{m-1} w(t)}{dt^{n-1}} + ... + b_1 \frac{dw(t)}{dt} + b_0 w(t) , \quad (5)$$

where a_0, … a_n and b_0, … b_m are real-valued constant coefficients, $a_n \neq 0$, with zero initial conditions. This differential equation can be written in the more compact form

$$\left(a_n \frac{d^n}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + ... + a_1 \frac{d}{dt} + a_0 \right) y(t)$$

$$= \left(b_m \frac{d^m}{dt^m} + b_{m-1} \frac{d^{m-1}}{dt^{n-1}} + ... + b_1 \frac{d}{dt} + b_0 \right) w(t) . \quad (6)$$

1.2.6 The Laplace Transform of a Signal

The two-sided Laplace transform of a continuous-time signal $y(t) \in \mathbb{R}$ is denoted by $Y(s)$ and defined by

$$Y(s) = \int_{-\infty}^{\infty} y(t) e^{-st} dt , \quad (7)$$

where $s = \sigma + j\omega$ is the Laplace transform variable, in which σ, $\omega \in \mathbb{R}$ and $j = \sqrt{-1}$. Given a signal $y(t)$ with Laplace transform $Y(s)$, $y(t)$ can be calculated from $Y(s)$ by taking the inverse Laplace Transform of $Y(s)$, which is defined by

$$y(t) = \int_{\sigma - j\infty}^{\sigma + j\infty} Y(s) e^{st} ds . \quad (8)$$

Theorem 1 Parseval's Theorem [7]:

$$\int_{-\infty}^{\infty} \left| y(t) \right|^2 dt = \int_{-j\infty}^{j\infty} \left| Y(s) \right|^2 ds . \quad (9)$$

Proof. *Let* $y^H(t) = \int_{\sigma - j\infty}^{\sigma + j\infty} Y^H(s) e^{-st} ds$ *and* $Y^H(s)$ *denote the Hermitian transpose (or adjoint) of* $y(t)$ *and* $Y(s)$*, respectively. The left-hand-side of (9) may be written as*

"No, no, you're not thinking; you're just being logical." *Niels Henrik David Bohr*