Dual Contributions of Optimal Estimation Theory in Aerospace Applications

Keynote Speech to the 1985 American Control Conference

Arthur Gelb

ABSTRACT: Spectacular advances in applications of estimation theory have occurred in the last 25 years. This paper focuses on aerospace applications, highlighting the dual contributions: insights, first; implementation, second.

Background

It is asserted that there are two related but different contributions that optimal estimation theory makes in aerospace (and other) system applications. Clearly, the most noticeable contribution is in the implementation of any system. Optimal estimation theory contributes, here, in a variety of ways, depending on the stage of development at which the system exists. One can envision distinct contributions to implementation in system research and development, test and evaluation, and operational and postmission stages. Just what is the nature of contribution that optimal estimation makes? The answer is that optimal estimation theory enables “squeezing out” the accuracy inherent—but otherwise unattainable—in physical (i.e., electronic, electromechanical, optical,...) systems.

Another contribution—which is, in fact, the underpinning for any success achieved in implementation—chronologically occurs first, during the conceptual phase of system development. Here, through focus on mathematical models of the physical system and understanding physical processes, optimal estimation theory enables otherwise unreachable insights into system performance and performance potential. Moreover, through common state-space language and framework, optimal estimation enables meaningful comparison of system concepts.

The reader might ask: “Why the fuss?” This question can be answered by taking a retrospective look at what the communication issues are today against an image of what they were before optimal estimation theory was adopted into engineering practice. Before modern estimation theory, researchers working on alternative systems approaches to solving a given problem might have had different perspectives that centered around the following construct: “My system is better than yours because my information manipulation is cleverer than yours.” But there was generally no way to test such assertions short of building both systems, an impractical thing to do.

Today, postmodern estimation theory, the corresponding debate would sound more like the following: “My system is better than yours either because my mathematical models better capture the underlying physics or because my suboptimal implementation is superior.” The point is that since they all understand optimal data processing, their debate centers on mathematical models, and performance simulations can be made by different workers on truly comparable bases. It is hard to really overstate the value that optimal estimation theory has made in the communication process.

Table 1 is a brief depiction of recent optimal estimation-technology history. Wiener’s pioneering work [1] was undoubtedly the greatest single stimulus in bringing the statistical point of view to center stage in estimation and control theory. His were the first explicit solutions for least-squares estimates of stochastic processes. It is, nevertheless, arguable that Wiener’s work notwithstanding, aerospace-systems applications involving estimation theory were done on a rather ad hoc basis up to the late 1950s. This was replaced by a particularly structured approach as a consequence of Kalman’s work [2], [3], a major contribution to the study and development of systems of great diversity. The first serious applications of these new

Dr. Arthur Gelb’s principal field of technical activity has been estimation and control applied to aerospace systems. In 1966, Dr. Gelb was a founder of TASC (The Analytic Sciences Corporation), a research and development organization engaged in modern systems engineering, software engineering, and program management support. He has served as TASC’s President since its inception. Dr. Gelb is a Fellow of the IEEE and a Fellow of the AIAA. He has published over 25 papers in technical journals, is co-author of Multiple-Input Describing Functions and Nonlinear System Design (McGraw-Hill Book Company, 1968), and is editor and co-author of Applied Optimal Estimation (MIT Press, 1974). He received the B.E.E. degree at the City College of New York in 1958, the S.M. degree in applied physics at Harvard University in 1959, and the Sc.D. degree in systems engineering at MIT in 1961.

References that centered around the following construct: “My system is better than yours because my information manipulation is cleverer than yours.” But there was generally no way to test such assertions short of building both systems, an impractical thing to do.

Today, postmodern estimation theory, the corresponding debate would sound more like the following: “My system is better than yours either because my mathematical models better capture the underlying physics or because my suboptimal implementation is superior.” The point is that since they all understand optimal data processing, their debate centers on mathematical models, and performance simulations can be made by different workers on truly comparable bases. It is hard to really overstate the value that optimal estimation theory has made in the communication process.

Table 1 is a brief depiction of recent optimal estimation-technology history. Wiener’s pioneering work [1] was undoubtedly the greatest single stimulus in bringing the statistical point of view to center stage in estimation and control theory. His were the first explicit solutions for least-squares estimates of stochastic processes. It is, nevertheless, arguable that Wiener’s work notwithstanding, aerospace-systems applications involving estimation theory were done on a rather ad hoc basis up to the late 1950s. This was replaced by a particularly structured approach as a consequence of Kalman’s work [2], [3], a major contribution to the study and development of systems of great diversity. The first serious applications of these new

| Table 1 | Recent optimal estimation milestones. |
results, occurring in the early 1960s, included NASA’s Apollo system and the Navy’s Polaris system. A veritable explosion of the literature occurred in the mid-1960s; see Table 2. By the mid-1970s, it can be said that the technology had matured, as evidenced by the appearance of a number of books that developed the material at some length (for example, [4], [5]). The use of optimal estimation technology is currently standard engineering practice across the country, and we have thus seen in the 25-year period, from 1960 to the present, the transition from “brilliant new theory” to “business as usual.” In retrospect, a breathtaking sequence.

It is easy to demonstrate the breadth of applications stimulated by optimal estimation theory. Table 3 lists nonaerospace applications. The biomedical applications cited include electroencephalogram (EEG) signal processing and blood-flow estimation, but these are only two among many. There are applications to communication systems, digital imagery, power systems, process control, and resource exploration. The reader should broadly interpret each of these categories. For example, under process control, two rather different kinds of processes are cited, namely, steel-making furnaces and traffic flows. We could have included economic processes and others under this category, as well.

A listing of aerospace-systems applications is presented in Table 4. It should be noted that there is a degree of arbitrariness as to which applications go into which category. Nevertheless, Table 4 lists five categories: navigation systems, vehicle control systems, remote sensing systems, fire control systems, and satellites. Each category is itself rich with examples.

Bibliographies of aerospace and non-aerospace applications demonstrate the unusually broad space spanned by applications of optimal estimation theory, as evidenced by papers published during the period from 1973 to 1984. In the sequel, however, examples are drawn only from the first category in Table 4, namely, navigation systems.

### Theoretical Underpinning

Prior to 1959, the theoretical basis for the development of optimal systems to separate signals from noise had been provided by Wiener filter theory. Recall that in Wiener filter theory, one deals with a stationary signal, $s(t)$, in a stationary noise, $n(t)$. The Wiener filter for separating signal from noise is the linear filter that produces minimum mean-squared estimation error. It is given by the transfer function

$$
H_{opt}(\omega) = \frac{1}{2\pi \Phi_s(\omega)} \left[ \int_0^\infty e^{-jut} d\tau \int_{-\infty}^{\infty} \Phi_{u}(\omega) e^{jut} d\omega \right]$$

(1)

where

$$i(t) = s(t) + n(t) = \text{input}$$

$$\Phi_u(\omega) = \Phi_s^T(\omega)\Phi_s(\omega)$$

= factored PSD of $i(t)$

$$\Phi_u(\omega) = \text{cross-spectral density between } i(t) \text{ and desired output, } d(t)$$

and the Wiener-Hopf equation was solved using spectrum factorization [6]. Although the optimum transfer function is explicitly expressed in Eq. (1), it always remains for researchers in the field to devise means for realizing this particular filter in an actual application. It is also true that the minimum mean-squared error can be expressed without knowledge of the optimum system itself ($\Phi_u(0) =$ mean-squared value of desired output)

$$\text{Min. MSE} = \phi_{w}(0) - \frac{1}{2\pi} \int_0^\pi \psi^2(\tau) d\tau$$

$$\psi(\tau) = \int_{-\infty}^{\infty} \Phi_{u}(\omega) e^{-jut} d\omega$$

(2)

Wiener’s work represented a very important theoretical breakthrough and provided considerable guidance in the development of signal processing systems, particularly in applications like radar and communications. Although Eqs. (1) and (2) are characterized for the single-input, single-output system, Wiener filter theory has been generalized to the case of multiple inputs and outputs. It has also been applied to two-dimensional random fields and treated in terms of minimum-sensitivity (robust) design. These are areas of current research.

However, limitations to the widespread application of the Wiener filter theory have indeed existed. The applications to high-order systems are analytically very complex, and that has been something of a stumbling block. Recently, with the advent of symbolic manipulation computer programs such as MIT’s MACSYMA, this analytical complexity issue has become much more manageable. The far more significant limitation of Wiener filter theory (whether in scalar or vector-valued formulation) is the constraint that it be applied to stationary processes and stationary systems. Although work has been done in an attempt to broaden the application of the Wiener filter theory to time-varying systems, available results are cumbersome and quite limited. Moreover, there are time-invariant systems for which transient processes are primarily of interest and Wiener filter theory would not apply to those. Both time-varying systems and transient processes are of considerable importance to applications in aerospace engineering.

Now consider the estimation problem as developed by Kalman in state-space notation. For a system and measurement given by

$$x = Fx + Gw$$

$$z = Hx + \nu$$

(3)
where the system and measurement noises \( v \) and \( w \) are zero mean, white, uncorrelated Gaussian random processes (\( w \sim N(0, Q) \), \( v \sim N(0, R) \), \( E[ww^T] = 0 \)). The optimal filter—that is, the linear, minimum mean-square error filter—is given by \([5]\)

\[
\dot{x} = F\dot{x} + K[z - H\dot{x}]
\]

\( \dot{x}(0) = x_0 \) \[4\]

where

\[
K = PH^TR^{-1}
\]

\[ P = E[(\dot{x} - x)(\dot{x} - x)^T] \]

and

\[
\dot{P} = FP + PF^T + GQG^T - PH^TR^{-1}HP
\]

\[ P(0) = P_0 \] \[5\]

The covariance matrix, \( P \), is a statistical measure of uncertainty in the estimate of \( x \). Although, in practice, one almost always implements discretized forms of the Kalman filter, in this paper, the continuous version of the equations is used because of its pedagogical clarity. Equation (4) is often referred to as the Kalman-Bucy filter; Eq. (5), which is a nonlinear differential equation governing behavior of the covariance matrix, is the matrix Riccati equation.

Like Wiener filter theory, Kalman filter theory was an important theoretical breakthrough. And, like Wiener filter theory, various generalizations and extensions exist for the Kalman filter. For instance, much work has been done on suboptimal filter implementations and minimum sensitivity formulations. Most importantly, however, the Kalman filter is not limited to stationary processes. It is, therefore, equally applicable to transient processes in constant-coefficient systems and to time-varying systems.

The power of Kalman filter theory derives from its ease of implementation and, particularly, from its ability to support a line of scientific inquiry into the likely behavior of a system to be implemented. The point is that, through studies performed on the estimation-error covariance matrix, much can be learned about likely system performance before any hardware is physically built. This is done through the vehicle of system sensitivity analysis.

Suppose an \( m \)-th order system and an associated measurement are given by Eq. (3) and, further, suppose an \( m \)-th order imple-mented filter is given by

\[
\dot{x} = F\dot{x} + K[z - H\dot{x}]
\]

\[ \dot{x}(0) = x_0 \] \[6\]

Although it is not a requirement, in practice, \( n \) is often much greater than \( m \). Now define

\[
\Delta F = W^TF^*W - F
\]

\[ \Delta H = H^*W - H \] \[7\]

where \( W \) accounts for dimensional incompatibility between \( F \) and \( F^*, H \) and \( H^*, \) viz

\[
x_{\text{amp}}(m\text{th order}) = WX_{\text{actual}}(m\text{th order}) \] \[8\]

The corresponding (linear) sensitivity equations are \([5]\)

\[
\dot{P} = W^T(F^* - K^*H^*)WP
\]

\[ + PW^T(F^* - K^*H^*)^T W \]

\[ + V^T(\Delta F - W^T K^* \Delta H)^T \]

\[ + GQG^T + W^T K^* R K^* W \] \[9\]

\[
\dot{V} = FV + VW^T(F^* - K^*H^*)^T W
\]

\[ + U(\Delta F - W^T K^* \Delta H)^T - GQG^T \] \[10\]

\[
\dot{U} = FU + UF^T + GQG^T \] \[11\]

where

\[
P = E[(\dot{x} - x)(\dot{x} - x)^T] \]

\[ V = E[x(\dot{x} - x)^T] \]

\[ U = E[xx^T] \]

and \( P(0) = -V(0) = U(0) = P_0 \). These equations are very powerful for generation of insights and answers to "what if" questions.

Based on these equations, Fig. 1 provides the framework for inquiry into system performance. For an \( n \)-state "truth model" and an \( m \)-state filter design model, one proceeds as shown in the diagram—namely, given \( W, F^*, H^* \), the nonlinear filter-covariance equation is solved for \( P^*(t) \) and, consequently, \( K^*(t) \). The filter-indicated performance, \( P^*(t) \), may or may not be the least bit indicative of actual system performance. However, realistically projected performance, \( P(t) \), which is generated by processing the error sources associated with the truth model through the linear sensitivity covariance equations, can be quite meaningful. Error budgets and sensitivity curves are typical outputs of this process and can provide quite considerable insight into the behavior of systems in advance of their creation in hardware. A key issue here, of course, is whether the mathematical models employed capture the essence of the physical processes they are intended to describe.

**Applications**

In this section, five different applications of optimal estimation theory are presented to demonstrate the issues discussed above.

**Omega Navigation System**

The first application addressed occurs in improving the accuracy of the Omega navigation system; see Fig. 2. Omega is a worldwide VLF radio-navigation system comprised of eight transmitting stations that transmit four frequencies: 10.2, 11.05, 11.33, and 13.6 kHz. Continuous wave signals of the same frequency received from two different transmitters are used to derive a phase-difference measurement, which itself.

---

Fig. 1. Framework for inquiry into system performance.
is used to identify a line of position (LOP) on an Omega navigation chart. Using a different pair of transmitters (with perhaps one transmitter in common), another LOP is identified and the intersection of the two LOPs provides a solution to the navigation problem.

Propagation of Omega radio-frequency energy is well described by a waveguide model, as shown in Fig. 3. The received skywave signal bounces off the ionosphere and is thus subject to diurnal and seasonal variations, as well as influence of earth conductivity and earth's magnetic field. Each of the above-mentioned effects can be modeled by appeal to the physics involved, and predicted propagation corrections (PPCs) can be created. The received phase-difference signals are corrected for deviations from nominals using these PPCs, resulting in navigation accuracy as indicated in Fig. 3. Here, a 24-hour snapshot of the Omega navigation error (deviation from the true LOP) is given. It can be seen that the peak-to-peak error of the Omega LOP on the baseline connecting a transmitter pair is approximately 15 centicycles (CECs) or 1.2 mi.

Now suppose that some measurements are made as the result of a worldwide network of Omega monitor sites, as indicated in Fig. 4. A monitor site is comprised of an Omega receiver at a known location, whereby direct measurement of the Omega phase error is possible. Suppose measurements are made at monitor sites (2) and (3) and suppose a phase-propagation correction is desired for site (1). If there is sufficient spatial correlation between the phase errors measured at sites (2) and (3) and the error experienced at site (1), it should be possible to use the information gathered at the monitor sites to improve the PPCs. For this purpose, we assume a Markov process model for the residual phase errors; let us assume it is a first-order Markov process with spatial "correlation distance" \( d \). Furthermore, let us assume isotropic behavior of the errors; that is, phase-error behavior is taken to be independent of direction. With these assumptions, it is possible to create a Kalman-like estimator for the phase error. A typical result is shown in Fig. 4, in which it is seen that the improved navigation-error performance is quite considerably better than that without the mathematics-based portion of the PPCs.

The form of the estimator is very easily shown. The problem is to estimate \( \phi_1 \), given

\[
\phi_1 = \frac{1}{1 - e^{-2\pi d}} \left[ \phi_2 (e^{-\pi d} - e^{-(\pi d + \pi d)}) + \phi_3 (e^{-\pi d} - e^{-(\pi d + \pi d)}) \right]
\]

One reason for the success of this formulation is the apparent insensitivity of the result to the assumed correlation distance parameter, as shown in Fig. 5. It is seen here that the correlation distance parameter can be chosen at 1500 ± 750 nm with almost no impact on the performance of the navigation

\[
\begin{align*}
\phi_1 &= \frac{1}{1 - e^{-2\pi d}} \\
&\cdot \left[ \phi_2 (e^{-\pi d} - e^{-(\pi d + \pi d)}) + \phi_3 (e^{-\pi d} - e^{-(\pi d + \pi d)}) \right]
\end{align*}
\]

Fig. 2. Omega radio-navigation system.

Fig. 3. Omega propagation corrections (physics-based).
system. It should also be mentioned that further improvements are possible with more complex (path-dependent) models.

Electrostatically Supported Gyroscope

The second application considered is the electrostatically supported gyroscope (ESG); see Fig. 6. In the ESG, a spherical rotor is suspended electrostatically in an evacuated cavity. The rotor is spun at high rate, using magnetic induction motors and, in a gimbaled mechanization, the gyro housing is torqued to "follow" the rotor spin axis. The hope is that the rotor will itself be torque-free, such that the rate of change of its angular momentum in inertial space will be zero, and therefore the device would represent "a star in a bottle" for navigation purposes. In a strapdown mechanization, the gyro housing is fixed directly to the vehicle being navigated and a computer performs the stabilizing function equivalent to gimbals [8].

Imperfections in manufacture cause small gravity torques on the rotor. There are two kinds of error processes in this sensor, drift rate, $\dot{e}(t)$, and pickoff errors, $p(t)$; see Fig. 7. The equations governing behavior of this device are

$$\dot{e}(t) = e(t), \quad \dot{p}(t) = \theta(t) + p(t)$$

One set of simplified error models characterizing drift rates and pickoff errors is given below

$$e(t) = \varepsilon_0 + \varepsilon_{g_2}(t) + \varepsilon_{g_3}(t) + \varepsilon_{g_4}(t) g_2(t) + \varepsilon_{g_5}(t) g_3(t)$$

$$p(t) = p_0 + p_1 g_1(t) + p_2 g_2(t) + p_3 g_3(t) + p_4 g_4(t)$$

wherein $\varepsilon_i$ and $p_i$ are presumably stable coefficients, and $g_1(t)$, $g_2(t)$, $g_3(t)$ are gravity components expressed in an inertially stable reference frame.

With the system and measurement models described, it is a straightforward process to calibrate this ESG. A simulated result is shown in Fig. 8. It is clear that the result of calibration (or compensation) in this instance is that the error is reduced by orders of magnitude and that the frequency spectrum of the residual error is quite different from that of the dominant ramp and single frequency error that existed before calibration. In a real-time application, the ESG and its error model (as realized in real-time software) would both be implemented to produce a sensor with an effective drift rate, which is only the unpredictable part of the raw ESG drift behavior; see Fig. 9.

Inertial Navigation System

The third application is minimization of gyro-induced inertial navigation system (INS) errors. In a typical INS mechanization, Fig. 10, the gyroscopes and platform servos inertially stabilize an accelerometer triad, which measures specific force in a stable reference frame. The navigation computer then integrates the accelerometer outputs via Newton’s equations of motion, providing position and velocity indications while supplying orientation control to the platform.

Suppose an inertial platform physically implements the $\hat{\mathbf{P}}, \hat{\mathbf{E}}, \hat{\mathbf{L}}$ coordinate frame depicted in Fig. 11, where $\hat{\mathbf{P}}$ is a unit vector parallel to the earth’s polar axis, $\hat{\mathbf{E}}$ is a unit vector orthogonal to $\hat{\mathbf{P}}$ and lying in an equatorial plane, and $\hat{\mathbf{L}}$ is orthogonal to $\hat{\mathbf{P}}$ and $\hat{\mathbf{E}}$. Let us further assume that the otherwise present 84-minute (Schuler) oscillations are damped and that the vehicle in which this INS is mechanized is moving at low speed. It can be shown [9] that the long-period INS error dynamics (as characterized by the behavior of INS “error angles” $\psi_F, \psi_L, \psi_E$) are governed by a third-order system; see Fig. 12. Here we see an open-loop integration and a 24-hour oscillator as the dynamic parts of the error model. Azimuth (i.e., north-pointing) error, latitude error, and longitude error are comprised of these dynamic components.

Suppose, now, we wish to “reset” the INS, that is, to estimate its errors and thereby remove them from the system outputs. Consider latitude, for example. The inertial-system output, indicated latitude ($L_i$), can be thought of as being comprised of true latitude ($L_t$) plus the inertial navigation system latitude error ($\delta L$). The inertially indicated latitude is compared to the indicated latitude derived by reference to an external position device—for example, Omega—whose output can itself be thought of as comprised of true latitude plus the external reference error ($\epsilon$). The result is a measurement that is the
SPECIFIC FORCE
ACCELEROMETER NAVIGATION
VELOCITY
STABILIZED PLATFORM
c- ATTITUDE.

HEADING
VEHICLE BODY

Fig. 10. Inertial Navigation System (INS).

\[ \mathbf{P}, \mathbf{E}, \mathbf{L} = \text{UNIT ORTHOG. VECTORS} \]

\[ \Omega = \text{EARTH RATE VECTOR} \]

\[ L = \text{SYSTEM LATITUDE} \]

\[ \lambda = \text{SYSTEM LONGITUDE} \]

Fig. 11. Implemented coordinate frame.

Fig. 12. Long-period INS error dynamics.

difference between the inertial system error and the external reference error; see Fig. 13. The problem is to find a data processing scheme that will process the measurements to estimate the inertial system latitude error.

Twenty five years ago, this might have been done as follows. Let us assume that the gyro drift rates are constant. Let us further assume that position fixes are perfect. Under these assumptions, latitude and longitude errors are given as:

\[ \delta L(t_i) = A_1 + A_2 \sin \Omega t_i + A_3 \cos \Omega t_i \]

(18)

\[ \delta \lambda(t_i) = A_1 t_i + A_2 - A_3 \tan L \cos \Omega t_i + A_3 \tan L \sin \Omega t_i \]

(19)

where \( A_1 \) to \( A_3 \) are coefficients involving initial conditions and the three gyro drift rates. If we take three latitude position fixes at three different times \( (t_1, t_2, t_3) \), three equations in three unknowns are produced. If the times were so chosen that the equations are solvable, that is, that the matrix shown below is not singular,

\[
\begin{bmatrix}
A_1 \\
A_2 \\
A_3
\end{bmatrix} =
\begin{bmatrix}
1 & \sin \Omega t_1 & \cos \Omega t_1 \\
1 & \sin \Omega t_2 & \cos \Omega t_2 \\
1 & \sin \Omega t_3 & \cos \Omega t_3
\end{bmatrix}^{-1}
\begin{bmatrix}
\delta L(t_1) \\
\delta L(t_2) \\
\delta L(t_3)
\end{bmatrix}
\]

(20)

we can directly calculate the coefficients \( A_1 \), \( A_2 \), and \( A_3 \). Using \( A_2 \) and \( A_3 \) plus two longitude position fixes, \( A_4 \) and \( A_5 \) can be directly determined from Eq. (19) and the system can therefore be "reset." A simulation of this particular INS reset scheme is shown in Fig. 14, in which actual gyro data were used in the simulation of system performance [10]. Position fixes were taken every 8 hours and, beginning with the third position fix, the solution described in Fig. (20) was used. It is easy to see that the result of the perfect position fixes is to reduce the latitude error to zero at the instant just after the position fix and, on average, to improve the system somewhat. But the system is by no means "settled out" as a result of the resets. The reason for this, of course, is that gyro drift rates are not constant and, therefore, the fundamental assumption of this reset scheme is invalid.

If it were felt that position fix errors contained large but random components, an improvement to the reset scheme just described would be to use \( N \) position fixes and a least-squares fit. This approach does have the potential to reduce the effect of fix errors but, of course, does nothing for the constant gyro drift-rate assumption. A further sophistication to this reset scheme might be to perform a joint least-squares fit using both latitude and longitude error measurements simultaneously. And, continuing along, additional ad hoc thinking can be used to provide many different reset schemes aimed at improving the performance of the INS.

Let us now move forward in time to the point where optimal estimation theory has been disclosed. Suppose we wish to solve the same problem just treated. In the case of constant gyro drift rates, we would adopt a six-element state vector as shown below

\[ \mathbf{x} = [\psi_F, \psi_L, \psi_E, \epsilon_F, \epsilon_L, \epsilon_E]^T \]
INERTIALLY INDICATED LATITUDE INERTIAL Li = L, + 6L

SYSTEM NAVIGATION

Fig. 13. Resetting an INS.

Fig. 14. Simulation of simple INS reset scheme.

Fig. 15. INS performance projections: Kalman filter reset.

Fig. 16. INS performance projections, cont.

and immediately write the system dynamics and measurement matrices, \( F \) and \( H \), as

\[
F = \begin{bmatrix}
1 \\
-\Omega & 1 \\
\Omega & 1 \\
\vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}
\]

\[
H^T = \begin{bmatrix}
1 \\
-\tan L
\end{bmatrix}
\]

If we were to assume a vehicle latitude of 30 degrees, a latitude position fix error of 1000 ft rms, and two fixes per hour, solution of the Riccati equation would immediately give us the desired results; see Fig. 15. Here it is obvious that three fixes are required to pin down the essence of the oscillatory latitude error, after which the error rate slows down quite considerably. By the fourth position fix, latitude error is in statistical steady state. A similar result holds for longitude. Notice that the rate of error growth slows quite considerably after the third position fix. This must mean that we are learning some-thing about the gyro drift rates, which cause the error growth in the first place. This is clearly demonstrated in Fig. 16, in which the effect of calibration of the three gyro drift rates is displayed as a function of time.

At this point, the reader might say: "So what; these results aren't really new." The answer, as mentioned earlier, is that when viewed against the background of the simplified reset schemes available before Kalman's papers were published, these results are nothing short of astonishing. The insights they provide as to system statistical performance and the impacts of the various parameters that characterize the system simply have no parallel in prior theory. It is quite impossible to say too much about the practical value of this sort of result—if the mathematical modeling is properly attended to.

Space Shuttle Navigation

The fourth application considers Space Shuttle approach and landing navigation analysis. Here it is assumed that the Space Shuttle has on board an inertial measurement unit (IMU), distance measuring equipment (DME), a barometric altimeter, and an instrument landing system (ILS). See Fig. 17. It is desired that the behavior of this candidate precision navigation system be explored.

For this example, following the methodology outlined in the section on Theoretical Underpinning in this paper, an 82-state "truth model" and a 24-state implemented filter were considered [11]. The result of an error-budget calculation is shown in Fig. 18, in which the error states comprising the "truth model" were arrayed into 17 groups, each of which was used to generate one or several rows in the error budget. Note that this is a case of a system that is exceedingly complex but not very complicated to work with. The methodology calls for the error
budget to be built up one row at a time, with correlated error sources always in a single group; group-to-group errors are thus considered uncorrelated. It is then possible to root-sum-square (rss) "down the rows" to produce a total column rss error, which is one piece of the projected performance of the system. One might next reexamine each column and identify the "5 percent plus" contributors to the total error (shown circled), for further consideration.

As an example of the use of this result to develop a sensitivity curve, consider the case of Group 7: Gyro Bias Drifts. In this case, the nominal drift-rate value is 0.01 deg/hr, and its nominal contribution to down-range velocity error can be seen to be 0.85 ft/sec. The total down-range velocity error is 1.07 ft/sec. By scaling the nominal value of gyro drift rates and thereby the corresponding position, velocity, and misalignment angles that linearly result from the scaling, a new row in the error budget is directly developed, and new column rss total-performance projections are computed with no new simulation required to produce the result! By this mechanism, the sensitivity curves of Fig. 19, the one on the left having just been discussed, are easily developed. They, clearly, have a great deal of insight value and would have been nothing but impossible to produce, in any practical way, by an alternative analytical method. The insight gained using this approach was, in fact, employed in the Space Shuttle program.

**Testing Inertial Navigation Systems**

A fifth and final application is drawn from the arena of testing inertial navigation systems. The test scenario is generally depicted in Fig. 20. If we wish to consider a single test, it is possible to do the obvious quantitative "trades," that is, of alternative instrumentation suites, test geometries, test durations, and so forth. However, modern inertial navigation systems often outperform the available test instrumentation systems, being themselves among the most precise electromechanical systems built. It is possible to employ multiple, nonidentical tests to yield observability enhancement.

There are two kinds of observability issues to be considered. The first is structural observability, in which parameter(s) are unobservable with no measurement noise, and the second is stochastic observability, in which measurement accuracy plus finite test time are insufficient to average the errors down. The framework for multiple testing is shown in Fig. 21. Actual processing of data from multiple tests, for example, using residuals analysis, leads directly to model verification. Additionally, and most importantly, predicting benefits gained by multiple tests, particularly under different scenarios, is possible in advance.

Multiple test-data processors are usually formulated in the "information" domain. Let \( P[i] \) and \( \tilde{X}[i] \), respectively, represent the estimated state and estimation-error covariance matrix associated with the ith test. Thus, in the case of \( N \) independent tests where \( P[i] \) is invertible (but this is not a requirement in general), the resulting covariance matrix and system state estimate are given by

\[
P^{-1} = \sum_{i=1}^{N} P^{-1}[i]
\]

\[
\tilde{X} = P^{-1}[i] \tilde{X}[i] + \sum_{i=1}^{N} P^{-1}[i] \tilde{X}[i]
\]

The look of which will be familiar to those acquainted with optimal smoothing formulations. Once again, the role played by optimal estimation theory is to provide a kind of insight simply not otherwise available, in advance of the actual need to commit resources (i.e., to "buy or bend metal"), potentially...
Fig. 19. System sensitivity to ("Truth model") parameter variations (time of touchdown).

Fig. 20. INS testing.

Fig. 21. Framework for multiple INS tests.

Fig. 22. Importance models.

Overview and Summary

Optimal estimation theory provides for a systematic treatment of high-order, non-stationary, mixed discrete-continuous and even nonlinear systems. However, real problems have been associated with the application of optimal estimation theory, particularly those of mismodeling and computer limitations. A variety of engineering solutions have been proposed to cope with both of these issues, for example, system/parameter identification, adaptive formulations, minimum-sensitivity formulations, square-root formulations, reduced-order models, ... [12]. At the present time, with continually improving computer hardware (i.e., more memory, shorter cycle times, longer word lengths), the computer limitations are being eliminated. However, the mismodeling issue remains a very real one.

In all applications, the importance of models needs to be highlighted. Consider, for example, the simple system illustrated in Fig. 22. Here we assume the long-term INS position-error dynamics are simply given by the open-loop integration, which is part of the longitude-error behavior. The gyro drift-rate model is allowed to vary from white noise to random ramp. Each model listed is in common use today, but the consequences of different drift-rate models are predictions that rms position error grows either as \( t^{1/2} \) (white-noise model), or \( t \) (bias model), or \( t^2 \) (random-walk model), or \( t^2 \) (random-ramp model). It is clearly apparent that the estimated inertial-system position error is highly model-dependent and that the consequence of incorrect choice of model would, over the long term, be substantial. The question is: "How do you cope with this issue?" The answer is: "You ask good questions!"

For example, one might inquire as to system sensitivity to "truth model" parameter variations. Figure 23 illustrates this case for a system such as that considered in the third application, namely an INS with periodic position fixes. In this case, the design value of a parameter \( \sigma \) is held constant, while the value of the same parameter in the "truth model" is allowed to vary. The results are latitude and longitude error curves that are monotonic with the increasing values of the parameter, which is presumed to be one of the driving noises in the system.

Next, one might inquire as to system sensitivity to filter parameter mismodeling, as shown in Fig. 24. Here, the actual value of the parameter \( \sigma \) is held constant, and the design value—that is, the value implemented in the filter—is allowed to vary. When the ratio of the design value of the parameter to the actual value is unity, mini-
truth model whose structure is different in that the order assumed is wrong. Or, the truth model might have some nonlinearities, or pure time delays, or distributed-parameter characteristics, etc. In any case, characterizing truth-model structural variations in the present context is problematic and perhaps impossible. Frosch [13] admonishes that "Given a set of inputs, the outputs of a particular model are probably mathematically pathological with regard to changes in the structure of the model." Whether pathological or not, it is clearly an issue of concern.

Having said all the above, there are a few realities we ought to accept. First, the optimal filter is optimal only for linear systems, linear measurements, and rational power spectra. Second, the filter model almost always contains approximations. Third, the "optimal filter" in real life is probably never optimal. And, furthermore, the "truth model" is never the "truth." And yet, hopefully, the "realistically" projected performance \((\hat{\mathbf{x}}, \mathbf{P})\) will be quite close to the actual system performance, notwithstanding that the filter performance \((\mathbf{x}^*, \mathbf{P}^*)\) may be quite far away. Experience shows that when the right questions are asked by someone who has an appropriate grasp of the physical underpinning of the mathematical models, this is, in fact, the case.

In summary, two points might be made. First, is that enormous and positive changes in optimal estimation theory have occurred in the last 25 years. They are central to aerospace-system applications and are not likely to be eclipsed soon, in the author's estimation. Secondly, the state of the art is now very sophisticated, and the dual contributions — insight, first, implementation, second — are alive and well and being made in practice across the country.
Armed with optimal estimation tools, modern systems engineers can and do contribute greatly to system conceptual design and development, and, in the process, develop absolutely remarkable insights (or, conversely, remarkable errors, if insufficient attention is paid to the underlying physics of the situation).

Acknowledgment

The author would like to acknowledge helpful discussions with the following individuals at TASC, particularly with regard to the applications described in this paper: Joe Kasper, Joe D’Apolitto, Bard Crawford, Steve Alter, Ray Gupta, and Dave Whitney.

References


Out of Control

"Adaptive control seems to be really catching on!"