INNOVATION

The Kalman Filter:

Navigation's Integration Workhorse

Larry J. Levy

The Johns Hopkins University
Applied Physics Laboratory

Since its introduction in 1960, the Kalman filter has become an integral component in thousands of military and civilian navigation systems. This deceptively simple, recursive digital algorithm has been an early-on favorite for conveniently integrating (or fusing) navigation sensor data to achieve optimal overall system performance. To provide current estimates of the system variables -- such as position coordinates -- the filter uses statistical models to properly weight each new measurement relative to past information. It also determines up-to-date uncertainties of the estimates for real-time quality assessments or for off-line system design studies. Because of its optimum performance, versatility, and ease of implementation, the Kalman filter has been especially popular in GPS/inertial and GPS stand-alone devices. In this month's column, Larry Levy will introduce us to the Kalman filter and outline its application in GPS navigation.

Dr. Levy is chief scientist of the Strategic Systems Department of The Johns Hopkins University Applied Physics Laboratory. He received his Ph.D. in electrical engineering from Iowa State University in 1971. Levy has worked on applied Kalman filtering for more than 30 years, codeveloped the GPS translator concept in SATRACK (a GPS-based missile-tracking system), and was instrumental in developing the end-to-end methodology for evaluating Trident II accuracy. He conducts graduate courses in Kalman filtering and system identification at The Johns Hopkins University Whiting School of Engineering and teaches Navtech Seminars's Kalman Filtering short course.

"Innovation" is a regular column featuring discussions about recent advances in GPS technology and its applications as well as the fundamentals of GPS positioning. The column is coordinated by Richard Langley of the Department of Geodesy and Geomatics Engineering at the University of New Brunswick, who appreciates receiving

your comments as well as topic suggestions for future columns. To contact him, see the "Columnists"section on page 4 of this issue.

When Rudolf Kalman formally introduced the Kalman filter in 1960, the algorithm was well received: The digital computer had sufficiently matured, many pressing needs existed (for example, aided inertial navigation), and the algorithm was deceptively simple in form. Engineers soon recognized, though, that practical applications of the algorithm would require careful attention to adequate statistical modeling and numerical precision. With these considerations at the forefront, they subsequently developed thousands of ways to use the filter in navigation, surveying, vehicle tracking (aircraft, spacecraft, missiles), geology, oceanography, fluid dynamics, steel/paper/power industries, and demographic estimation, to mention just a few of the myriad application areas.

EQUATION-FREE DESCRIPTION

The Kalman filter is a multiple-input, multiple-output digital filter that can optimally estimate, in real time, the states of a system based on its noisy outputs (see Figure 1). These states are all the variables needed to completely describe the system behavior as a function of time (such as position, velocity, voltage levels, and so forth). In fact, one can think of the multiple noisy outputs as a multidimensional signal plus noise, with the system states being the desired unknown signals. The Kalman filter then filters the noisy measurements to estimate the desired signals. The estimates are statistically optimal in the sense that they minimize the mean-square estimation error. This has been shown to be a very general criterion in that many other reasonable criteria (the mean of any monotonically increasing, symmetric error function such as the absolute value) would yield the same estimator. The Kalman filter was a dramatic improvement over its minimum mean square error predecessor, in-vented by Norbert Wiener in the 1940s, which was primarily confined to scalar signals in noise with stationary statistics.

Figure 2 illustrates the Kalman filter algorithm itself. Because the state (or signal) is typically a vector of scalar random variables (rather than a single variable), the state uncertainty estimate is a variance-covariance matrix -- or simply, covariance matrix. Each diagonal term of the

WHAT GAUSS SAID

If the astronomical observations and other quantities, on which the computation of orbits is based, were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler), and, therefore, if other observations were used, they might be confirmed, but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and

matrix is the variance of a scalar random variable -- a description of its uncertainty. The term is the variable's mean squared deviation from its mean, and its square root is its standard deviation. The matrix's off-diagonal terms are the covariances that describe any correlation between pairs of variables.

The multiple measurements (at each time point) are also vectors that a recursive algorithm processes sequentially in time. This means that the algorithm iteratively repeats itself for each new measurement vector, using only values stored from the previous cycle. This procedure distinguishes itself from batch-processing algorithms, which must save all past measurements.

Starting with an initial predicted state estimate (as shown in Figure 2) and its associated covariance obtained from past information, the filter calculates the weights to be used when combining this estimate with the first measurement vector to obtain an updated "best" estimate. If the measurement noise covariance is much smaller than that of the predicted state estimate, the measurement's weight will be high and the predicted state estimate's will be low.

Also, the relative weighting between the scalar states will be a function of how "observable" they are in the measurement. Readily visible states in the measurement will receive the higher weights. Because the filter calculates an updated state estimate using the new measurement, the state estimate covariance must also be changed to reflect the information just added, resulting in a reduced uncertainty. The updated state estimates and their associated covariances form the Kalman filter outputs.

Finally, to prepare for the next measurement vector, the filter must project the updated state estimate and its associated covariance to the next measurement time. The actual system state vector is assumed to change with time according to a deterministic linear transformation plus an independent random noise. Therefore, the predicted state estimate follows only the deterministic transformation, because the actual noise value is unknown. The covariance

the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable, to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already attained, which is afterwards to be corrected so as to satisfy all the observations in the most accurate manner possible.

-- From Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections, a translation by C.H. Davis of C.F. Gauss's 1809 Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Davis's 1857 translation was republished by Dover Publications, Inc., New York, in 1963.

prediction ac-counts for both, because the random noise's uncertainty is known. Therefore, the prediction uncertainty will increase, as the state estimate prediction cannot account for the added random noise. This last step completes the Kalman filter's cycle.

 One can see that as the measurement vectors are recursively processed, the state estimate's uncertainty should generally decrease (if all states are observable) because of the accumulated information from the measurements. However, because information is lost (or uncertainty increases) in the prediction step, the uncertainty will reach a steady state when the amount of uncertainty increase in the prediction step is balanced by the uncertainty decrease in the update step. If no random noise exists in the actual model when the state evolves to the next step, then the uncertainty will eventually approach zero. Because the state estimate uncertainty changes with time, so too will the weights. Generally speaking, the Kalman filter is a digital filter with time-varying gains. Interested readers should consult "The Mathe-matics of Kalman Filtering" sidebar for a summary of the algorithm.

If the state of a system is constant, the Kalman filter reduces to a sequential form of deterministic, classical least squares with a weight matrix equal to the inverse of the measurement noise covariance matrix. In other words, the Kalman filter is essentially a recursive solution of the least-squares problem. Carl Friedrich Gauss first solved the problem in 1795 and published his results in 1809 in his Theoria Motus, where he applied the least-squares method to finding the orbits of celestial bodies (see the "What Gauss Said" sidebar). All of Gauss's statements on the effectiveness of least squares in analyzing measurements apply equally well to the Kalman filter.

A SIMPLE EXAMPLE

A simple hypothetical example may help clarify the concepts in the preceding section. Consider the problem of determining the actual resistance of a nominal 100-ohm resistor by making repeated ohmmeter measurements and processing them in a Kalman filter.

First, one must determine the appropriate statistical models of the state and measurement processes so that the filter can compute the proper Kalman weights (or gains). Here, only one state variable -- the resistance, x -- is unknown but assumed to be constant. So the state process evolves with time as

$$x_{k+1} = x_k . [1]$$

Note that no random noise corrupts the state process as it evolves with time. Now, the color code on a resistor indicates its precision, or tolerance, from which one can deduce -- assuming that the population of resistors has a Gaussian or normal histogram -- that the uncertainty (variance) of the 100-ohm value is, say, $(2 \text{ ohm})^2$. So our best estimate of x, with no measurements, is $x_{0/-} = 100$ with an uncertainty of $P_{0/-} = 4$. Repeated ohmmeter measurements,

$$z_k = x_k + v_k \ , \ [2]$$

directly yield the resistance value with some measurement noise, v_k (measurement errors from turn-on to turn-on are assumed uncorrelated). The ohmmeter manufacturer indicates the measurement noise uncertainty to be $R_k = (1 \text{ ohm})^2$ with an average value of zero about the true resistance.

Starting the Kalman filter at $k = 0$, with the initial estimate of 100 and uncertainty of 4, the weight for updating with the first measurement is

$$K_0 = \frac{P_{0/-}}{P_{0/-} + R_0} = \frac{4}{4+1} \ , \ [3]$$

with the updated state estimate as

$$x_{0/0} = (1-K_0)x_{0/-} + K_0 z_0$$

$$[4]$$

where $x_{0/0}$ denotes the best estimate at time 0, based on the measurement at time 0.

Note that the measurement receives a relatively high weight because it is much more precise (less uncertain) than the initial state estimate. The associated uncertainty or variance of the updated estimate is

$$P_{0/0} = (1 - K_0)P_{0/-} =$$

$$\frac{1}{4 + 1} \cdot 4 \quad = \quad \frac{4}{5}$$

. [5]

Also note that just one good measurement decreases the state estimate variance from 4 to 4/5. According to equation [1], the actual state projects identically to time 1, so the estimate projection and variance projection for the next measurement at time 1 is

$$x_{1/0} = x_{0/0} \; ; P_{1/0} = P_{0/0} = \frac{4}{5}$$

. [6]

Repeating the cycle over again, the new gain is

$$K_1 \quad = \quad \frac{P_{1/0}}{P_{1/0}+R_1} \quad = \quad \frac{4/5}{4/5+1}$$

[7]

and the new update variance is

$$P_{1/1} = (1 - K_1) P_{1/0} =$$

$$\left(\frac{1}{4/5 + 1} \right) \frac{4}{5} = \frac{4}{9}$$

. [8]

[Figure 3](#) represents a simulation of this process with the estimate converging toward the true value. The estimation uncertainty for this problem, which the Kalman filter provides, appears in [Figure 4](#). One can see that the uncertainty will eventually converge to zero.

A New Set of Conditions. Let's now change the problem by assuming that the measurements are taken one year apart with the resistor placed in extreme environmental conditions so that the true resistance changes a small amount. The manufacturer indicates that the small change is independent from year to year, with an average of zero and a variance of 1/4 ohms^2. Now the state process will evolve with time as

$$x_k + 1 = x_k + w_k. \text{ [9]}$$

where the random noise, w_k, has a variance of $Q_k = 1/4$. In the previous case, the variance prediction from time 0 to time 1 was constant as in equation [6]. Here, because of the random noise in equation [9], the variance prediction is

$$P_{1/0} = P_{0/0} + Q_0 = \frac{4}{5} + \frac{1}{4} = 1.05$$

. [10]

Now the gain and update variance calculations proceed on as in equations [7] and [8] but with larger values for the predicted variance. This will be repeated every cycle so that the measurement update will decrease the variance while the prediction step will increase the variance. Figure 5 illustrates this tendency. Eventually, the filter reaches a steady state when the variance increase in the prediction step matches the variance decrease in the measurement update step, with $P_{k+1/k} = 0.65$ and $P_{k/k} = 0.4$. The Q_k represents a very important part of the Kalman filter model because it tells the filter how far back in time to weight the measurements. An incorrect value of this parameter may dramatically affect performance.

GPS/INS INTEGRATION

We can see that the Kalman filter provides a simple algorithm that can easily lend itself to integrated systems and requires only adequate statistical models of the state variables and associated noises for its optimal performance. This fact led to its wide and enthusiastic use in aided inertial applications.

Integrating GPS with an inertial navigation system (INS) and a Kalman filter provides improved overall navigation perfor-
mance. Essentially, the INS supplies virtually noiseless outputs that slowly drift off with time. GPS has minimal drift but much more noise. The Kalman filter, using statistical models of both systems, can take advantage of their different error characteristics to optimally minimize their deleterious traits.

As shown in the "The Mathematics of Kalman Filtering" sidebar, the Kalman filter is a linear algorithm and assumes that the process generating the measurements is also linear. Because most systems and processes (including GPS and INS) are nonlinear, a method of linearizing the process about some known reference process is needed. Figure 6 illustrates the approach for integrating GPS and inertial navigators. Note that the true values of each system cancel out in the measurement into the Kalman filter so that only the GPS and inertial *errors* need be modeled. The reference trajectory, one hopes, is sufficiently close to the truth so that the error models are linear and the Kalman filter is optimal. For most GPS applications this is the case.

THE MATHEMATICS OF KALMAN FILTERING

The Kalman filter assumes that the system state vector, x_k, evolves with time as

$$x_{k+1} = \mathbf{F}_k x_k + w_k$$

with the measurement vector given by

$$z_k = \mathbf{H}_k x_k + v_k$$

where x_0, w_k, and v_k are mutually uncorrelated vectors: The latter two are white noise sequences, with means of $\mathbf{m}_0$, 0, and 0 and nonnegative definite covariances of S_0, Q_k, and R_k, respectively. The corresponding optimal Kalman filter is given by the recursive algorithm of Figure 7, which corresponds to the block

So, even though the overall systems are nonlinear, the Kalman filter still operates in the linear domain. Of course, the state variables for the Kalman filter must adequately model all error variables from both systems. GPS errors could include receiver clock, selective availability, ionospheric, tropospheric, multipath, and satellite ephemeris and clock errors. Inertial inaccuracies, on the other hand, could include position, velocity, orientation, gyro, accelerometer, and gravity errors. The equipment quality and the application requirements will determine how extensive the error models must be.

If the GPS outputs are user position, one terms the integration architecture as loosely coupled. A tightly coupled architecture depicts one in which the GPS outputs are pseudoranges (and possibly carrier phases) and the reference trajectory is used (along with the GPS ephemeris from the receiver) to predict the GPS measurements. In the tightly coupled system, the measurement errors would be in the range domain rather than the position domain. Usually, the tightly coupled arrangement is preferred because it is less sensitive to satellite dropouts, and adequate Kalman filter models are simpler and more accurate. One must employ the loosely coupled arrangement when the receiver outputs provide position without raw measurements.

The open-loop correction approach of [Figure 6] is termed linearized Kalman filtering. An alternate approach in which the algo-rithm feeds the estimates back to the inertial system to keep the reference trajectory close to the truth is an example of extended Kal-man filtering.

GPS-ONLY NAVIGATION

In some applications, an INS is not desired or may not be available, as in a stand-alone GPS receiver. In such cases, the Kalman filter resides within the receiver, and some known (or assumed) receiver *equations of motion* will replace the inertial system in a tightly coupled version of [Figure 6]. The extent to which the equations of motion (usually dead reckoning, for a moving receiver) faithfully

diagram of [Figure 2]. The vector $x_{k/j}$ denotes the optimal estimate of x at time t_k, based on measurements up to t_j, and $P_{k/j}$ is the corresponding "optimal" estimation error covariance matrix when the implemented filter model matches the real-world system that is actually generating the data.

One can derive the filter equations using a number of methods. Minimizing the generalized mean square error, $E[e^t_{k/j}Ae_{k/j}]$, where $e_{k/j}[x_k - x_{k/j}$ and A is any positive semidefinite weighting matrix, results in the Kalman equations if all variables and noises are Gaussian. For non-Gaussian cases, an additional restriction requires that there be a linear relationship between the state estimate, the measurements, and the predicted state.

model the receiver trajectory will determine the error model needed in the Kalman filter.

Simple equations of motion generally exhibit large errors that cause degraded performance relative to inertial-based reference trajectories in moving-receiver scenarios. Of course, fixed location equations of motion are trivial and very accurate. Here, the advantage of using Kalman filtering versus a single-point, least-squares fix is that the equations of motion can smooth the GPS noise, improving the performance.

PRACTICAL DESIGNS

Regardless of an application's equipment -- be it GPS, INS, or other devices -- developing a practical Kalman filter-based navigation system requires attention to a variety of design considerations.

The filter's covariance analysis portion (not requiring real data; see Figures 2 and 7) uses predetermined error models of potential systems (GPS, inertial, and so forth) to predict the particular configuration's performance. The filter designer repeats this for different potential equipment (models) until the requirements are satisfied. In some cases, one must implement the Kalman filter in a "small" computer with only a few states to model the process. This suboptimal filter must be evaluated by special covariance analysis algorithms that recognize the differences in the real-world model producing the measurements and the implemented filter model. Finally, once the filter meets all performance requirements, a few simulations of all processes should be run to evaluate the adequacy of the linearization approach and search for numerical computational errors.

In most cases, the extended Kalman filter (with resets after every cycle) will ameliorate any linearization errors. Numeric computational errors caused by finite machine word length manifest themselves in the covariance matrices, which become nonsymmetric or have negative diagonal elements, causing potentially

Further Reading

The literature on Kalman filtering abounds, with applications ranging from spacecraft navigation to the demographics of the French beef cattle herd. To ease you into it, here are a few suggestions.

For the seminal introduction of the Kalman filter algorithm, see

- "A New Approach to Linear Filtering and Prediction Problems," by R.E. Kalman in the Journal of Basic Engineering, the Transactions of the American Society of Mechanical Engineers, Series D, Vol. 83, No. 1, pp. 35-45, March 1960.

For an excellent, comprehensive introduction to Kalman filtering, including a GPS case study, see

- Introduction to Random Signals and Applied

disastrous performance. This problem can be alleviated by increasing the computational precision or by employing a theoretically equivalent but more numerically robust algorithm.

CONCLUSIONS

Because of its deceptively simple and easily programmed optimal algorithm, the Kalman filter continues to be the integration method of choice in GPS-based navigation systems. It requires sufficiently accurate multidimensional statistical models of all variables and noises to properly weight noisy measurement data. These models enable the filter to account for the disparate character of the errors in different systems, providing for an optimal integrated combination of large-scale systems. The recursive nature of the filter allows for efficient real-time processing. Off-line covariance studies enable the integrated system performance to be predicted before development, providing a convenient and easy-to-use system design tool.

Kalman Filtering (with Matlab exercises and solutions), 3d edition, by R.G. Brown and P.Y.C. Hwang, published by John Wiley & Sons, Inc., New York, 1997.

For discussions about various Kalman filter applications, see

- IEEE Transactions on Automatic Control, Special issue on applications of Kalman filtering, Vol. AC-28, No. 3 published by the Institute of Electrical and Electronics Engineers (IEEE), March 1983.

For a comprehensive selection of reprints of Kalman filter theory and application papers, including some of the germinal ones from the 1960s and those from the IEEE Transactions on Automatic Control special issue, see

- Kalman Filtering: Theory and Application, edited by H.W. Sorenson, published by IEEE Press, New York, 1985.

For a discussion about special covariance analysis and numerically robust algorithms, see the lecture notes

- Applied Kalman Filtering, Navtech Seminars, Course 457, presented by L.J. Levy, July 1997.

For an introductory discussion

about GPS and inertial navigation integration, see

- "Inertial Navigation and GPS," by M.B. May, in GPS World, Vol. 4, No. 9, September 1993, pp. 56-66.

Several good Web sites devoted to Kalman filtering exist, including

- "The Kalman Filter," a site maintained by G. Welch and G. Bishop of the University of North Carolina at Chapel Hill's Department of Computer Science: http://www.cs.unc.edu/~welch/kalmanLinks.html.

Copyright 1998 Advanstar Communications
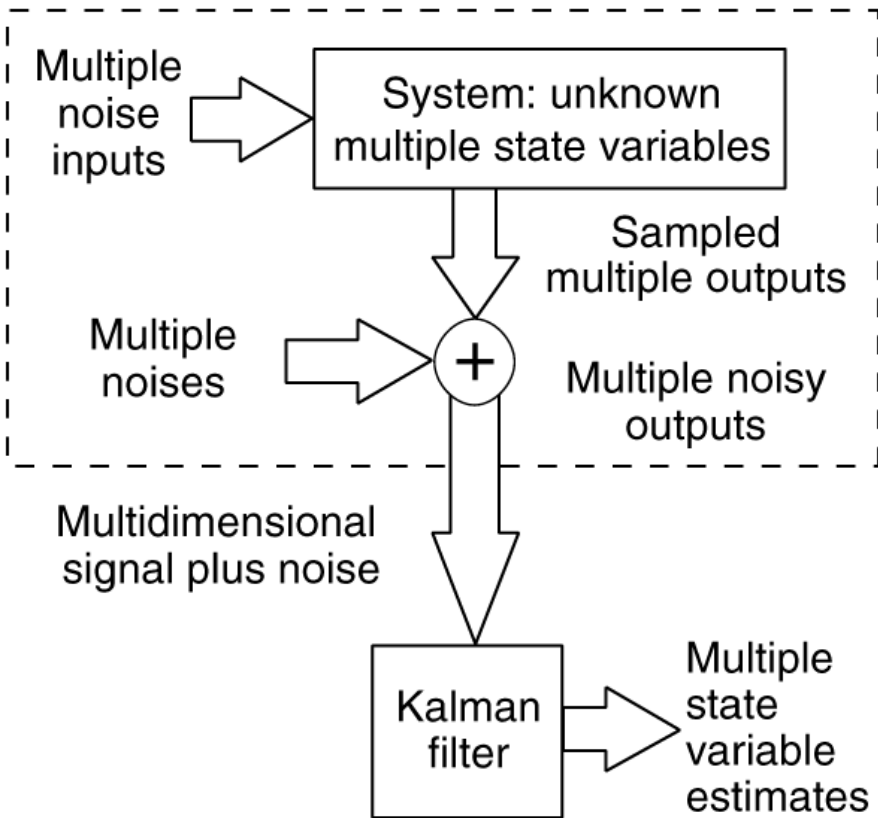
Figure 1.The purpose of a Kalman filter is to optimally estimate the values of variables describing the state of a system from a multidimensional signal contaminated by noise.
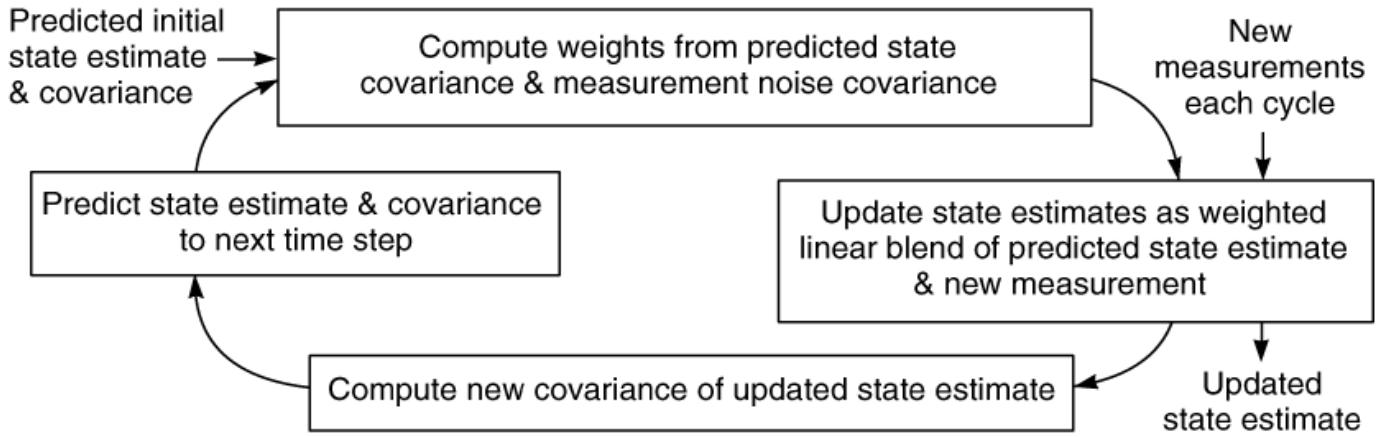
Figure 2.The Kalman filter is a recursive, linear filter. At each cycle, the state estimate is updated by combining new measurements with the predicted state estimate from previous measurements.
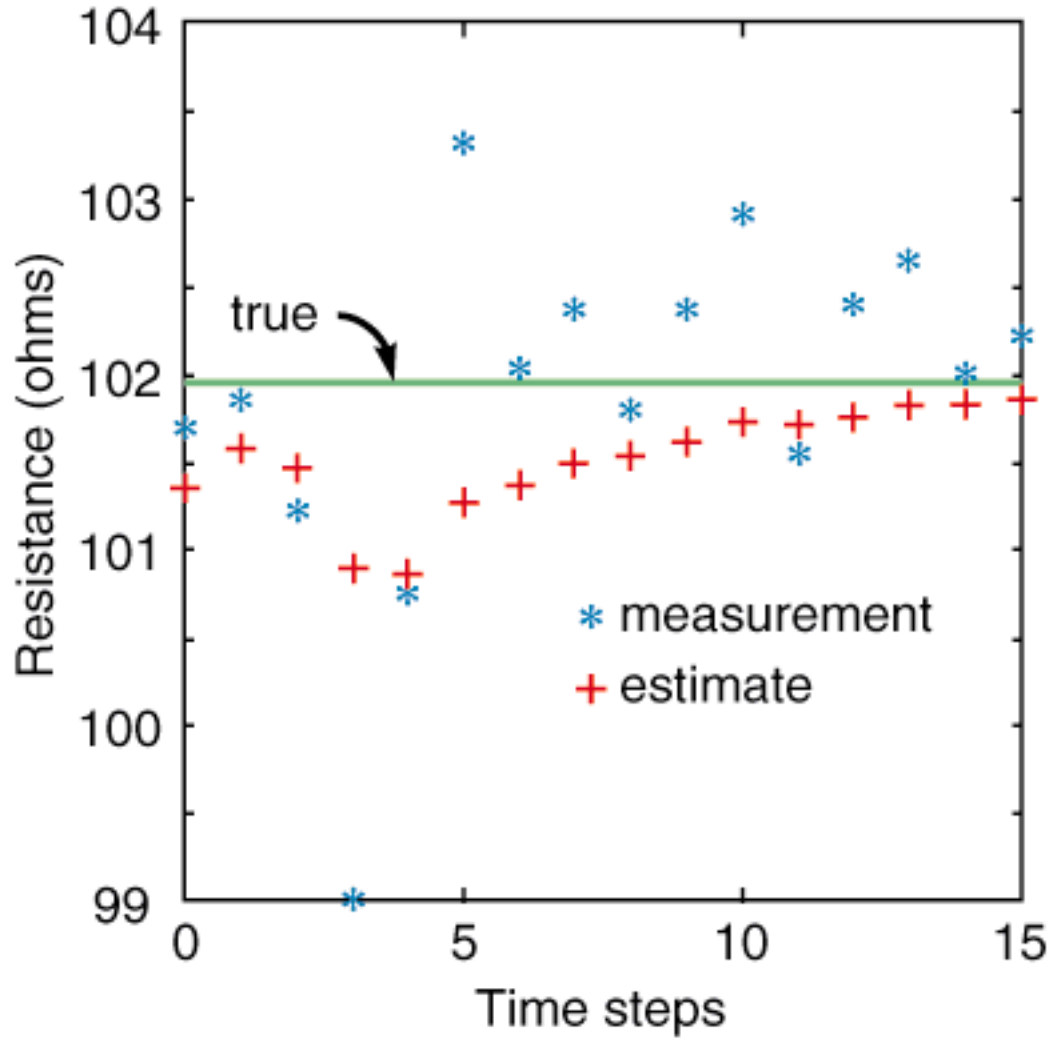
Figure 3.The individual resistance measurements of a nominally 100-ohm resistor are scattered about the true value of slightly less than 102 ohms. The Kalman filter estimate gradually converges to this value.
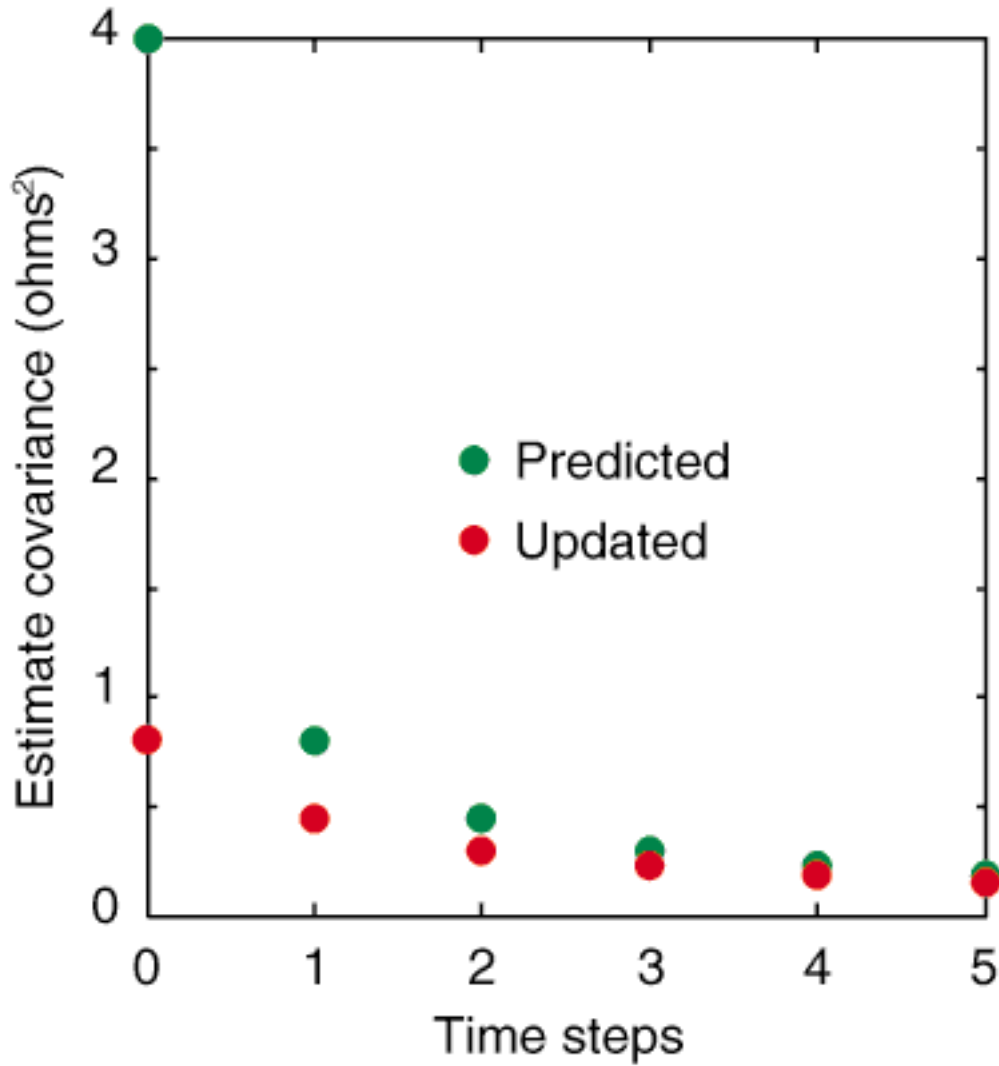
Figure 4. Initially, the predicted uncertainty (variance) of the resistor's value is 4 ohms^2 and is simply based on the manufacturer-provided tolerance value. However, after six measurements, the estimated variance drops to below 0.2 ohms^2.
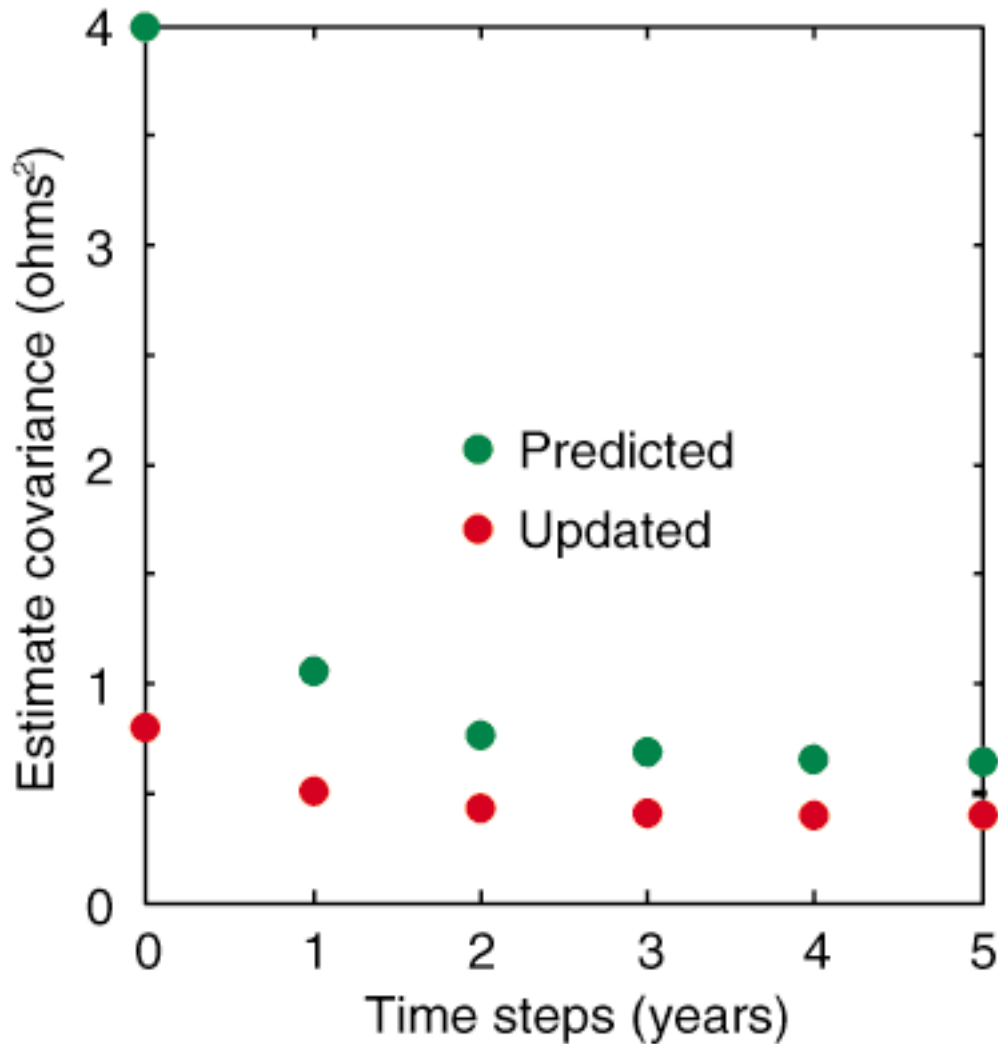
Figure 5.The true resistance of a resistor in an environment with a widely varying temperature is not quite constant. If this is modeled assuming a variation with an average of zero and a variance of 0.25 ohms^2, the Kalman filter estimate of the resistance variance converges to 0.4 ohms^2 after only a few measurements.
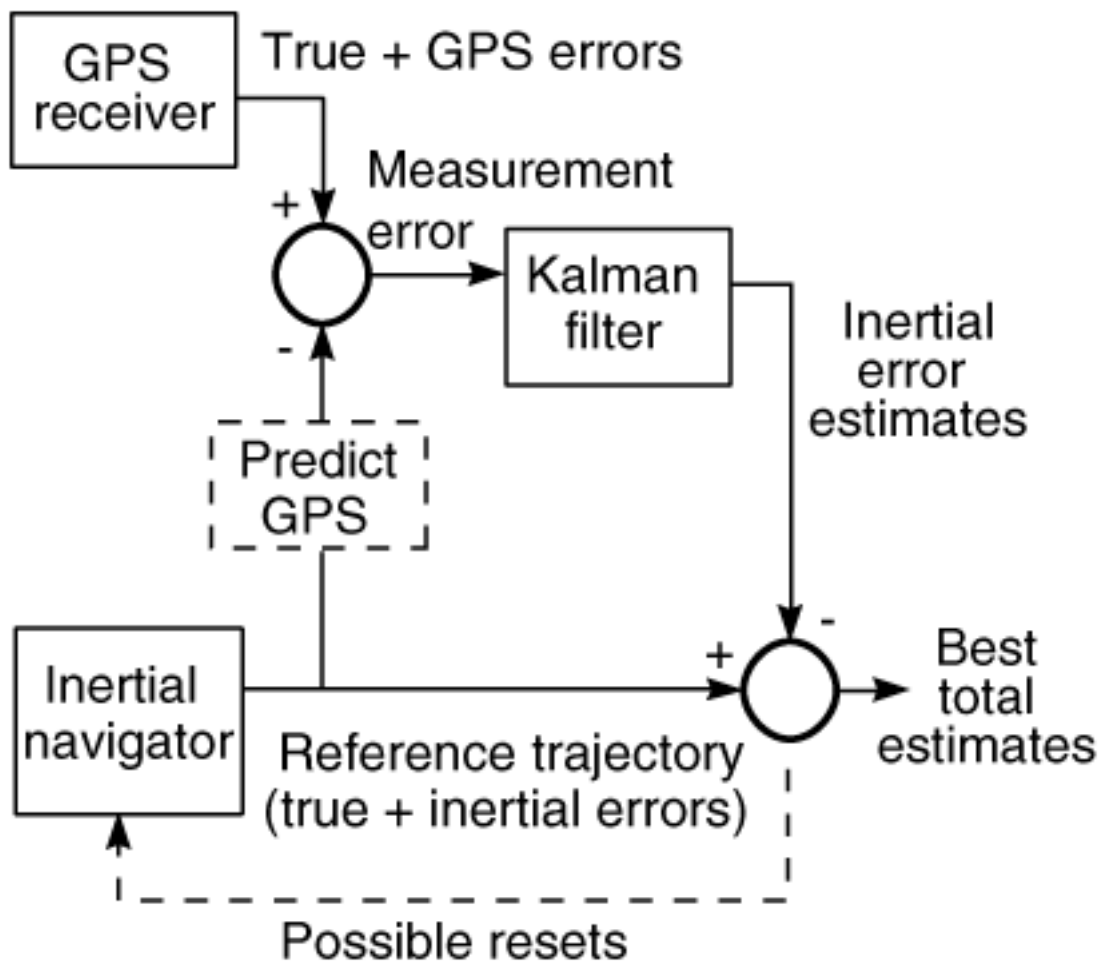
Figure 6.An integrated GPS receiver and inertial navigator use a Kalman filter to improve overall navigation performance
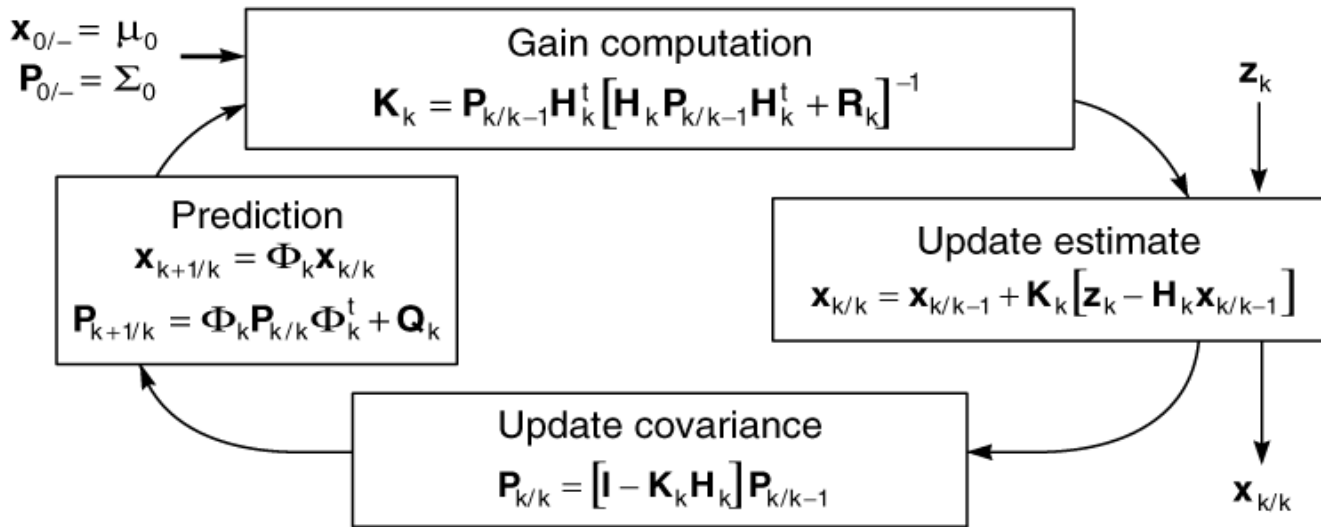
Figure 7.The Kalman filter algorithm involves four steps: gain computation, state estimate update, covariance update, and prediction.