

14

Navigation and Tracking

James L. Farrell
VIGIL, Inc.

14.1 Introduction

14.2 Fundamentals

14.3 Applications

Position and Velocity along a Line • Position and Velocity in Three-Dimensional Space • Position, Velocity, and Acceleration of a Tracked Object • Position, Velocity, and Attitude in Three-Dimensional Space (INS Aiding) • Individual GPS Measurements as Observables

14.4 Conclusion

References

Further Information

14.1 Introduction

The task of navigation (“Nav”) interacts with multiple avionics functions. To clarify the focus here, this chapter will not discuss tight formations, guidance, steering, minimization of fuel/noise/pollution, or managing time of arrival. The accent instead is on determining position and velocity (plus, where applicable, other variables such as acceleration, verticality, heading) with maximum accuracy reachable from whatever combination of sensor outputs are available at any time. Position can be expressed as a vector displacement from a designated point or in terms of latitude/longitude/altitude above mean sea level, above the geoid—or both. Velocity can be expressed in a locally level coordinate frame with various choices for an azimuth reference (e.g., geodetic North, Universal Transverse Mercator [UTM] grid North, runway centerline, wander azimuth with or without Earth sidereal rate torquing). In principle any set of axes could be used — such as an Earth Centered Earth Fixed (ECEF) frame for defining position by a Cartesian vector; or velocity in Cartesian coordinates or in terms of groundspeed, flight path angle, and ground track angle — in either case it is advisable to use accepted conventions.

Realization of near-optimal accuracy with any configuration under changing conditions is now routinely achievable. The method uses a means of dead-reckoning — preferably an Inertial Navigation System (INS) — which can provide essentially continuous position, velocity, and attitude in three dimensions by performing a running accumulation from derivative data. Whenever a full or partial fix is available from a nav sensor, a discrete update is performed on the entire set of variables representing the state of the nav system; the amount of reset for each state variable is determined by a weighting computation based on modern estimation. In this way, “initial” conditions applicable to the dead-reckoning device in effect are reinitialized as the “zero” time is advanced (and thus kept current) with each update. Computer-directed operations easily accommodate conditions that may arise in practice (incomplete fixes, inconsistent data rates, intermittent availability, changing measurement geometry, varying accuracies) while providing complete flexibility for backup with graceful degradation. The approach inherently combines

short-term accuracy of the dead-reckoning data with the nav aids' long-term accuracy. A commonly cited example of synergy offered by the scheme is a tightly coupled GPS/INS wherein the inertial information provides short-term aiding that vastly improves responsiveness of narrowband code and/or carrier tracking, while GPS information counteracts the long-term accumulation of INS error.

The goal of navigation has progressed far beyond mere determination of geographic location. Efforts to obtain double and triple "mileage" from inertial instruments, by integrating nav with sensor stabilization and flight control, are over a decade old. Older yet are additional tasks such as target designation, precision pointing, tracking, antenna stabilization, imaging sensor stabilization (and therefore transfer alignment). Digital beamforming (DBF) for array antennas (including graceful degradation to recover when some elements fail), needs repetitive data for instantaneous relative position of those elements; on deformable structures this can require multiple low-cost transfer-aligned Inertial Measuring Units (IMUs) and/or the fitting of spatial data to an aeroelastic model. The multiplicity of demands underlines the importance of integrating the operations; the rest of this chapter describes how integration should be done.

14.2 Fundamentals

To accomplish the goals just described, the best available balance is obtained between old and new information — avoiding the both extremes of undue clinging to old data and jumping to conclusions at each latest input. What provides this balance is a modern estimation algorithm that accepts each data fragment as it appears from a nav sensor, immediately weighing it in accordance with its ability to shed light on every variable to be estimated. That ability is determined by accounting for all factors that influence how much or how little the data can reveal about each of those variables: Those factors include

- Instantaneous geometry (e.g., distance along a skewed line carries implications about more than one coordinate direction),
- Timing of each measurement (e.g., distance measurements separated by known time intervals carry implications about velocity as well as position), and
- Data accuracy, compared with the accuracy of estimates existing before measurement.

Only when all these factors are taken into account are accuracy and flexibility as well as versatility maximized. To approach the ramifications gradually, consider a helicopter hovering at constant altitude, which is to be determined on the basis of repeated altimeter observations. After setting the initial *a posteriori* estimate to the first measurement \hat{Y}_1 , an *a priori* estimate $\hat{x}_2^{(-)}$ is predicted for the second measurement and that estimate is refined by a second observation,

$$\hat{x}_2^{(-)} = \hat{x}_1^{(+)}; \quad \hat{x}_2^{(+)} = \hat{x}_2^{(-)} + \frac{1}{2}z_2, \quad z_2 \triangleq \hat{Y}_2 - \hat{x}_2^{(-)} \quad (14.1)$$

and a third observation,

$$\hat{x}_3^{(-)} = \hat{x}_2^{(+)}; \quad \hat{x}_3^{(+)} = \hat{x}_3^{(-)} + \frac{1}{3}z_3, \quad z_3 \triangleq \hat{Y}_3 - \hat{x}_3^{(-)} \quad (14.2)$$

and then a fourth observation,

$$\hat{x}_4^{(-)} = \hat{x}_3^{(+)}; \quad \hat{x}_4^{(+)} = \hat{x}_4^{(-)} + \frac{1}{4}z_4, \quad z_4 \triangleq \hat{Y}_4 - \hat{x}_4^{(-)} \quad (14.3)$$

which now clarifies the general expression for the m^{th} observation,

$$\hat{x}_m^{(-)} = \hat{x}_{m-1}^{(+)}; \quad \hat{x}_m^{(+)} = \hat{x}_m^{(-)} + \frac{1}{m}z_m, \quad z_m \triangleq \hat{Y}_m - \hat{x}_m^{(-)} \quad (14.4)$$

which can be rewritten as

$$\hat{x}_m^{(+)} = \frac{m-1}{m}\hat{x}_m^{(-)} + \frac{1}{m}\hat{Y}_m, \quad m > 0 \quad (14.5)$$

Substitution of $m = 1$ into this equation produces the previously mentioned condition that the first *a posteriori* estimate is equal to the first measurement; substitution of $m = 2$, combined with that condition, yields a second *a posteriori* estimate equal to the average of the first two measurements. Continuation with $m = 3, 4, \dots$ yields the general result that, after m measurements, estimated altitude is simply the average of all measurements.

This establishes an equivalence between the *recursive* estimation formulation expressed in (14.1)–(14.5) and the *block* estimate that would have resulted from averaging all data together in one step. Since that average is widely known to be optimum when all observations are statistically equally accurate, the recursion shown here must then be optimum under that condition. For measurement errors that are sequentially independent random samples with zero mean and variance R , it is well known that the mean squared estimation error $P_m^{(+)}$ after averaging m measurements is just R/m . That is the variance of the *a posteriori* estimate (just after inclusion of the last observation); for the *a priori* estimate the variance $P_m^{(-)}$ is $R/(m-1)$. It is instructive to express the last equation above as a blended sum of old and new data, weighted by factors

$$\frac{R}{P_m^{(-)} + R} \equiv \frac{R/P_m^{(-)}}{1 + R/P_m^{(-)}} = \frac{m-1}{m} \quad (14.6)$$

and

$$\frac{P_m^{(-)}}{P_m^{(-)} + R} = \frac{1}{m} \quad (14.7)$$

respectively; weights depend on variances, giving primary influence to information having lower mean squared error. This concept, signified by the left-hand sides of the last two equations, is extendable to more general conditions than the restrictive (uniform variance) case considered thus far. We are now prepared to address more challenging tasks.

As a first extension, let the sequence of altimeter measurements provide repetitive refinements of estimates for both altitude x_1 and vertical velocity x_2 . The general expression for the m^{th} observation now takes a more inclusive form

$$\hat{x}_m^{(-)} = \Phi_m \hat{x}_{m-1}^{(+)}; \quad \hat{x}_m^{(+)} = \hat{x}_m^{(-)} + \mathbf{W}_m z_m, \quad z_m \triangleq \hat{Y}_m - \hat{x}_{1,m}^{(-)}, \quad \mathbf{x}_m \triangleq \begin{bmatrix} x_{m,1} \\ x_{m,2} \end{bmatrix} \quad (14.8)$$

The method accommodates estimation of multiple unknowns, wherein the status of a system is expressed in terms of a *state vector* (“state”) \mathbf{x} , in this case a 2×1 vector containing two *state variables* (“states”); superscripts and subscripts continue to have the same meaning as in the introductory example, but for these states the conventions $m,1$ and $m,2$ are used for altitude and vertical velocity, respectively, at time t_m . For this dynamic case the *a priori* estimate at time t_m is not simply the previous *a posteriori* estimate; that previous state must be premultiplied by the *transition matrix*,

$$\Phi_m = \begin{bmatrix} 1 & t_m - t_{m-1} \\ 0 & 1 \end{bmatrix} \quad (14.9)$$

which performs a time extrapolation. Unlike the static situation, elapsed time now matters since imperfectly perceived velocity enlarges altitude uncertainty between observations — and position measurements separated by known time intervals carry implicit velocity information (thus enabling vector estimates to be obtained from scalar data in this case). Weighting applied to each measurement is influenced by three factors:

- A sensitivity matrix \mathbf{H}_m whose (i, j) element is the partial derivative of the i^{th} component of the m^{th} measured data vector to the j^{th} state variable. In this scalar measurement case \mathbf{H}_m is a 1×2 matrix $[1 \ 0]$ for all values of m .
- A covariance matrix \mathbf{P}_m of error in state estimate at time t_m [the i^{th} diagonal element = mean squared error in estimating the i^{th} state variable and, off the diagonal, $P_{ij} = P_{ji} = \sqrt{P_{ii}P_{jj}} \times$ (correlation coefficient between i^{th} and j^{th} state variable uncertainty)].
- A covariance matrix \mathbf{R}_m of measurement errors at time t_m (in this scalar measurement case \mathbf{R}_m is a 1×1 “matrix,” i.e., a scalar variance R_m).

Although formation of \mathbf{H}_m and \mathbf{R}_m follows directly from their definitions, \mathbf{P}_m changes with time (e.g., recall the effect of velocity error on position error) and with measurement events (because estimation errors fall when information is added). In this “continuous-discrete” approach, uncertainty is decremented at the discrete measurement events

$$\mathbf{P}_m^{(+)} = \mathbf{P}_m^{(-)} - \mathbf{W}_m \mathbf{H}_m \mathbf{P}_m^{(-)} \quad (14.10)$$

and, between events, dynamic behavior follows a continuous model of the form

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{E} \quad (14.11)$$

where \mathbf{E} acts as a forcing function to maintain positive definiteness of \mathbf{P} (thereby providing stability and effectively controlling the remembrance duration—the “data window” denoted herein by T —for the estimator) while \mathbf{A} defines dynamic behavior of the state to be estimated ($\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ and $\dot{\Phi} = \mathbf{A}\Phi$). In the example at hand,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (14.12)$$

Given \mathbf{H}_m , \mathbf{R}_m , and $\mathbf{P}_m^{(-)}$ the optimal (Kalman) weighting matrix is

$$\mathbf{W}_m = \mathbf{P}_m^{(-)} \mathbf{H}_m^T (\mathbf{H}_m \mathbf{P}_m^{(-)} \mathbf{H}_m^T + R_m)^{-1} \quad (14.13)$$

which for a scalar measurement produces a vector \mathbf{W}_m as the above inversion simplifies to division by a scalar (which becomes the variance R_m added to P_{11} in this example):

$$\mathbf{W}_m = \mathbf{P}_m^{(-)} \mathbf{H}_m^T / (\mathbf{H}_m \mathbf{P}_m^{(-)} \mathbf{H}_m^T + R_m) \quad (14.14)$$

The preceding (hovering helicopter) example is now recognized as a special case of this vertical nav formulation. To progress further, horizontal navigation addresses another matter, i.e., location uncertainty in more than one direction—with measurements affected by more than one of the unknowns (e.g., lines of position [LOPs] skewed off a cardinal direction such as North or East; [Figure 14.1](#)). In the classic “compass-and-dividers” approach, dead reckoning would be used to plot a running accumulation of position increments until the advent of a fix from two intersecting straight or curved LOPs. The position

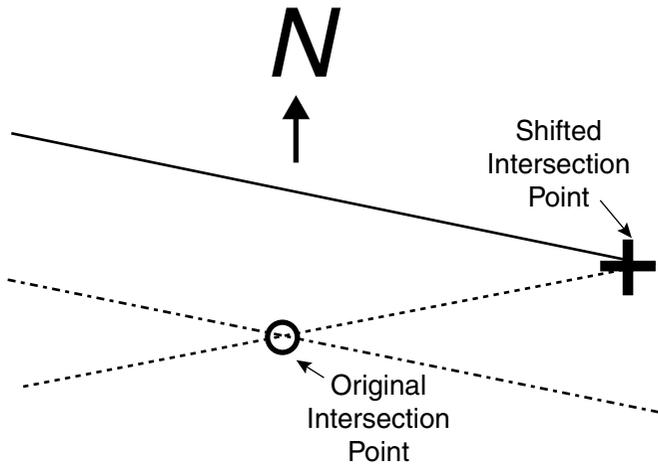


FIGURE 14.1 Nonorthogonal LOPs.

would then be reinitialized at that fixed position, from whence dead reckoning would continue until the next fix. For integrated nav we fundamentally alter that procedure as follows:

- In the reinitialization, data imperfections are taken into account. As already discussed, Kalman weighting (Equations 14.13 and 14.14) is based on accuracy of the dead-reckoning extrapolation as well as the variance of each measurement *and* its sensitivity to each state variable. An optimal balance is provided between old and new information, and the optimality inherently applies to updating of every state variable (e.g., to velocity estimates as well as position, even when only position is observed directly).
- Fixes can be incomplete. In this example, one of the intersecting LOPs may be lost. An optimal update is still provided by the partial fix data, weighted by \mathbf{W}_m of (14.14).

Implications of these two alterations can be exemplified by Figure 14.1, depicting a pair of LOPs representing partial fixes, not necessarily synchronous. Each scalar measurement allows the entire state vector to be optimally updated with weighting from (14.14) in the relation,

$$\hat{\mathbf{x}}_m^{(+)} = \hat{\mathbf{x}}_m^{(-)} + \mathbf{W}_m z_m \quad (14.15)$$

where z_m is the *predicted residual* formed by subtracting the predicted measurement from the value observed at time t_m and acceptance-tested to edit out wild data points;

$$z_m = y_m + \epsilon = Y_m - \hat{Y}_m^{(-)} + \epsilon = \hat{Y}_m - \hat{Y}_m^{(-)}; \quad \hat{Y}_m^{(-)} = Y(\hat{\mathbf{x}}_m^{(-)}) \quad (14.16)$$

The measurement function $Y(\mathbf{x})$ is typically a simple analytical expression (such as that for distance from a designated point, the difference between distances from two specified station locations, GPS pseudo-range or carrier phase difference, etc.). Its partial derivative with respect to each position state is obtained by simple calculus; other components of \mathbf{H}_m (e.g., sensitivity to velocity states) are zero, in which case updating of those states occurs due to dynamics from off-diagonal elements of \mathbf{P} in the product $\mathbf{P}_m^{(-)} \mathbf{H}_m^T \cdot R_m$ — whether constant or varying (e.g., with signal strength) — is treated as a known quantity; if not accurately known, a conservative upper bound can be used. The same is true for the covariance matrix \mathbf{P}_0 of error in state estimate at the time of initiating the estimation process — after which the changes are tracked by (14.10) at each measurement event, and by (14.11) between measurements — thus \mathbf{P} is always available for Equations 14.13 and 14.14.

It is crucial to note that the updates are *not* obtained in the form of newly measured coordinates, as they would have been for the classical “compass-and-dividers” approach. Just as old navigators knew how to use partial information, a properly implemented modern estimator would not forfeit that capability. The example just shown provides the best updates, even with no dependably precise way of obtaining a point of intersection when motion occurs between measurements. Furthermore, even with a valid intersection from synchronized observations, the North coordinate of the intersection in Figure 14.1 would be more credible than the East. To show this, consider the consequence of a measurement error effectively raising the dashed LOP to the solid curve as shown; the North coordinate of the new intersection point “+” exceeds that of point “O” — but by less than the East-West coordinate shift.

Unequal sensitivity to different directions is automatically taken into account via \mathbf{H}_m —just as the dynamics of \mathbf{P} will automatically provide velocity updating without explicitly forming velocity in terms of sequential changes in measurements — and just as individual values of R_m inherently account for measurement accuracy variations.

Theoretically then, usage of Kalman weighting unburdens the designer while ensuring optimum performance; no other weighing could provide lower mean squared error in the estimated value of any state. Practically, the fulfillment of this promise is realized by observing additional guidelines, some of which apply “across the board” (e.g., usage of algorithms that preserve numerical stability) while others are application dependent.

Now that a highly versatile foundation has been defined for general usage, the way is prepared for describing some specific applications. The versatility just mentioned is exhibited in the examples that follow. Attention is purposely drawn to the standard process cycle; models of dynamics and measurements are sufficient to define the operation.

14.3 Applications

Various operations will now be described, using the unified form to represent the state dynamics* with repetitive instantaneous refresh via discrete or discretized observations (fixes, whether full or partial). Finite space necessitates some limitations in scope here. First, all updates will be from position-dependent measurements (e.g., Doppler can be used as a source of continuous dead-reckoning data but is not considered herein for the discrete fixes). In addition, all nav reference coordinate frames under consideration will be locally level. In addition to the familiar North-East-Down (NED) and East-North-Up (ENU) frames, this includes any Wander Azimuth frame (which deviates from the geographic by only an azimuth rotation about the local vertical). Although these reference frames are not inertial (thus the velocity vector is not exactly the time integral of total acceleration as expressed in a nav frame), known kinematical adjustments will not be described in any depth here. This necessitates restricting the aforementioned data window T to intervals no greater than a tenth of the 84-minute Schuler period. The limitation is not very severe when considering the amount of measured data used by most modern avionics applications within a few minutes duration.

Farrell¹ is cited here for expansion of conditions addressed, INS characterization, broader error modeling, increased analytical development, the physical basis for that analysis, and myriad practical “*do* s and *don't* s” for applying estimation in each individual operation.

14.3.1 Position and Velocity along a Line

The vertical nav case shown earlier can be extended to the case of time-varying velocity; with accurately (not necessarily exactly) known vertical acceleration Z_v ,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ Z_v \end{bmatrix} \quad (14.17)$$

*A word of explanation is in order: For classical physics the term *dynamics* is reserved for the relation between forces and translational acceleration, or torques and rotational acceleration — while *kinematics* describes the relation between acceleration, velocity, and position. In the estimation field, all continuous time-variation of the state is lumped together in the term *dynamics*.

which allows interpretation in various ways. With a positive upward convention (as in the ENU reference, for example), x_1 can represent altitude above any datum while x_2 is upward velocity; a positive downward convention (NED reference) is also accommodated by simple reinterpretation. In any case, the above equation correctly characterizes actual vertical position and velocity (with true values for Z_V and all x s), and likewise characterizes *estimated* vertical position and velocity (denoted by circumflexes over Z_V and all x s). Therefore, by subtraction, it also characterizes *uncertainty in* vertical position and velocity (i.e., error in the estimate, with each circumflex replaced by a tilde \sim). That explains the role of this expression in two separate operations:

- Extrapolation of the *a posteriori* estimate (just after inclusion of the last observation) to the time of the next measurement, to obtain an *a priori* estimate of the state vector — which is used to predict the measurement’s value. If a transition matrix can readily be formed (e.g., Equation 14.9 in the example at hand), it is sometimes, but not always, used for that extrapolation.
- Propagation of the covariance matrix from time t_{m-1} to t_m via (14.11) initialized at the *a posteriori* value $\mathbf{P}_{m-1}^{(+)}$ and ending with the *a priori* value $\mathbf{P}_m^{(-)}$. Again, an alternate form using (14.9) is an option.

After these two steps, the cycle at time t_m is completed by forming gain from (14.14), predicted residual from (14.16), update via (14.15), and decrement by (14.10).

The operation just described can be visualized in a generic pictorial representation. Velocity data in a dead-reckoning (DR) accumulation of position increments predicts the value of each measurement. The difference z between the prediction and the observed fix (symbolically shown as a discrete event depicted by the momentary closing of a switch) is weighted by position gain W_{pos} and velocity gain W_{vel} for the update. Corrected values, used for operation thereafter, constitute the basis for further subsequent corrections.

For determination of altitude and vertical velocity, the measurement prediction block in Figure 14.2 is replaced by a direct connection; altimeter fixes are compared vs. the repeatedly reinitialized accumulation of products (time increment) \times (vertical velocity). In a proper implementation of Figure 14.2 time history of *a posteriori* position tracks the truth; RMS position error remains near $\sqrt{P_{11}}$. At the first measurement, arbitrarily large initial uncertainty falls toward sensor tolerance — and promptly begins rising at a rate dictated by $\sqrt{P_{22}}$. A second measurement produces another descent followed by another climb, but now at gentler slope, due to implicit velocity information gained from repeated position observations within a known time interval. With enough fix data the process approaches a quasi-static condition with $\sqrt{P_{11}}$ maintained at levels near RMS sensor error.

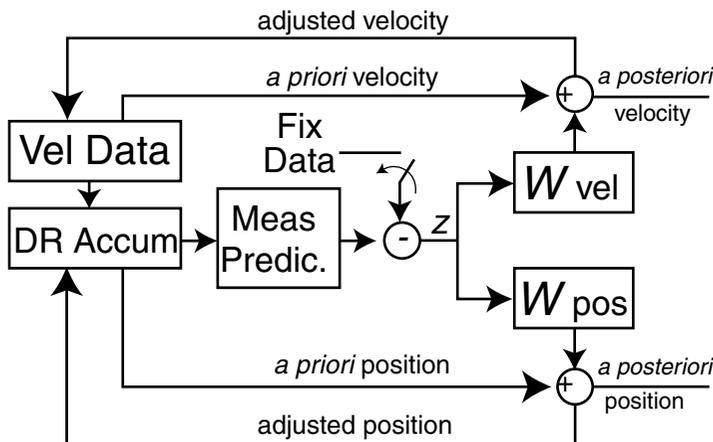


FIGURE 14.2 Position and velocity estimation.

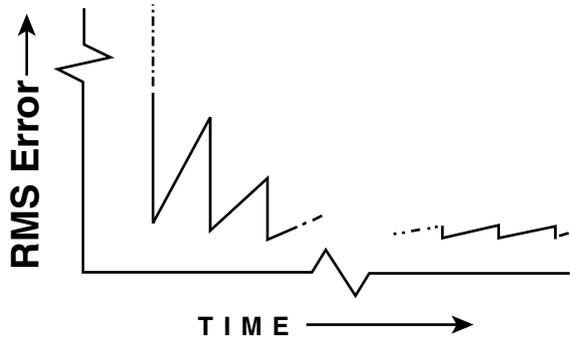


FIGURE 14.3 Time history of accuracy.

Extensive caveats, ramifications, etc. could be raised at this point; some of the more obvious ones will be mentioned here.

- In analogy with the static example, the *left* side of (14.7), with $P_{m11}^{(-)}$ substituted for $P_m^{(-)}$, implies high initial weighting followed by lighter weights as measurements accumulate. If fixes are from sensors with varying tolerance, the entire approach remains applicable; only parameter values change. The effect in Figure 14.3 would be a smaller step decrement, and less reduction in slope, when RMS fix error is larger.
- Vertical velocity can be an accumulation of products, involving instantaneous vertical acceleration which comes from data containing an accelerometer offset driven by a randomly varying error, e.g., having spectral density in conformance to \mathbf{E} of (14.11). With this offset represented as a third state, another branch would be added to Figure 14.2 and an augmented form of (14.17) could define dynamics in instantaneous altitude, vertical velocity, and vertical acceleration (instead of a constant bias component, extension to exponential correlation is another common alternative);

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} \quad (14.18)$$

Rather than ramping between fixes, position uncertainty then curves upward faster than the linear rate in Figure 14.3; curvature starts to decrease after the *third* fix. It takes longer to reach quasi-static condition, and closeness of “steady-state” $\sqrt{\overline{P}_{11}}$ to RMS sensor error depends on measurement scheduling density within a data window.

- (14.12) and Figure 14.2 can also represent position and velocity estimation along another direction, e.g., North or East — or both, as developed in the next section.

14.3.2 Position and Velocity in Three-Dimensional Space

For brevity, only a succinct description is given here. First consider excursion over a meridian with position x_1 expressed as a product [latitude (*Lat*) increment] \times [total radius of curvature ($R_M + \text{altitude}$)],

$$R_M = \frac{a_E(1 - e_E^2)}{[1 - e_E^2 \sin^2(Lat)]^{3/2}}; \quad a_E = 6378137 \text{ m}; \quad e_E^2 = (2 - f)f, \quad f = \frac{1}{298.25722} \quad (14.19)$$

so that, for usage of \mathbf{A} in (14.12), x_2 is associated with North component V_N of velocity. North position fixes could be obtained by observing the altitude angle of Polaris (appropriately corrected for slight

deviation off the North Pole). To use the formulation for travel in the East direction, the curvature radius is $(R_p + h)$,

$$R_p = a_E / \sqrt{1 - e_E^2 \sin^2(Lat)}; \quad h = \text{altitude} \quad (14.20)$$

and, while the latitude rate is $V_N / (R_M + h)$, the longitude rate is $V_E \sec(Lat) / (R_p + h)$.

Even for limited distance excursions within a data window, these spheroidal expressions would be used in kinematic state extrapolation, while our short-term ground rule allows a simplified (“flat-Earth” Cartesian) model to be used as the basis for matrix extrapolation in (14.11). The reason lies with *very* different sensitivities in Equations 14.15 and 14.16. The former is significantly less critical; a change δW would modify the *a posteriori* estimate by only the second-order product $z_m \delta \mathbf{W}_m$. By way of contrast, small variations in an anticipated measurement (from seemingly minor model approximations) can produce an unduly large deviation in the residual — a small difference of large quantities.

Thus, for accuracy of *additive state vector* adjustments (such as *velocity* \times *time* products in dynamic propagation), Equations 14.19 and 14.20 properly account for path curvature and for changes in direction of the nav axes as the path progresses. At the poles, the well-known singularity in $\{\sec(Lat)\}$ of course necessitates a modified expression (e.g., Earth-centered vector).

In applying (14.12) to all three directions, a basic decision must be made at the outset. Where practical, it is desirable for axes to remain separated, which produces three uncoupled two-state estimators. An example of this form is radar tracking at long range — long enough so that, within a data window duration, the line-of-sight (LOS) direction remains substantially fixed (i.e., nonrotating). If all three axes are monitored at similar data rates and accuracies, experience has shown that even a fully coupled six-state estimator has position error ellipsoid axes aligned near the sensor’s range/azimuth/elevation directions. In that case, little is lost by ignoring coupling across sensor reference axes — hence the triad of uncoupled two-state estimators, all in conformance to (14.12). To resolve vectors along cardinal directions at any time, all that is needed is the direction cosine matrix transformation between nav and sensor axes, which is always available.

When the conditions mentioned above do not hold, the reasoning needs to be revisited. If LOS direction rotates (which happens at short range), or if all three axes are *not* monitored at similar data rates, decoupling may or may not be acceptable; in any case it is suboptimal. If one axis (or a pair of axes) is *unmonitored* a fully coupled six-state estimator can dramatically outperform the uncoupled triad. In that case, although (14.12) represents uncoupled dynamics for each axis, coupling comes from multiple changing projections in measurement sensitivity \mathbf{H} as the sensor sight-line direction rotates.

Even the coupled formulation has a simple dynamic model in partitioned form; for a relative position vector \mathbf{R} and velocity \mathbf{V} driven by perturbing acceleration \mathbf{e} ,

$$\begin{bmatrix} \dot{\mathbf{R}} \\ \dot{\mathbf{V}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ \mathbf{V} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{e} \end{bmatrix} \quad (14.21)$$

where \mathbf{I} and $\mathbf{0}$ are null and identity partitions. The next section extends these concepts.

14.3.3 Position, Velocity, and Acceleration of a Tracked Object

In this chapter it has been repeatedly observed that velocity can be inferred from position-dependent measurements separated by known time intervals. In fact, a velocity *history* can be inferred. As a further generalization of methods just shown, the position reference need not be stationary. In the example now to be described, the origin will move with a supersonic jet carrying a radar and INS. Furthermore, the object whose state is being estimated can be external, with motions that are independent of the platform carrying the sensors that provide all the measurements.

For tracking, first consider the uncoupled case already described, wherein each of three separate estimator channels corresponds to a sensor reference axis direction and each channel has three kinematically related states, representing that directional component of relative (sensor-to-tracked-object) position, relative velocity, and total (not relative) acceleration of the tracked object.* The expression used to propagate state estimates between measurements in a channel conforms to standard kinematics, i.e.,

$$\begin{bmatrix} \hat{\mathbf{x}}_{m1}^{(-)} \\ \hat{\mathbf{x}}_{m2}^{(-)} \\ \hat{\mathbf{x}}_{m3}^{(-)} \end{bmatrix} = \begin{bmatrix} 1 & t_m - t_{m-1} & \frac{1}{2}(t_m - t_{m-1})^2 \\ 0 & 1 & t_m - t_{m-1} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{m-1,1}^{(+)} \\ \hat{\mathbf{x}}_{m-1,2}^{(+)} \\ \hat{\mathbf{x}}_{m-1,3}^{(+)} \end{bmatrix} - \begin{bmatrix} \frac{1}{2}(t_m - t_{m-1})q_m \\ q_m \\ 0 \end{bmatrix} \quad (14.22)$$

where q_m denotes the component, along the sensor channel direction, of the change in INS velocity during $(t_m - t_{m-1})$. In each channel, \mathbf{E} of (14.11) has only one nonzero value, a spectral density related to data window and measurement error variance σ^2 by

$$E_{33} = (20\sigma^2/T^5)/g^2 \quad (\text{g/sec})^2/\text{Hz} \quad (14.23)$$

To change this to a fully coupled 9-state formulation, partition the 9×1 state vector into three 3×1 vectors \mathbf{R} for relative position, \mathbf{V}_r for relative velocity, and \mathbf{Z}_T for the tracked object's total acceleration — all expressed in the INS reference coordinate frame. The partitioned state transition matrix is then constructed by replacing each diagonal element in (14.22) by a 3×3 identity matrix \mathbf{I}_{33} , each zero by a 3×3 null matrix, and multiplying each above-diagonal element by \mathbf{I}_{33} . Consider this transition matrix to propagate covariances as expressed in *sensor reference axes*, so that parameters applicable to a sensing channel are used in (14.23) for each measurement.

Usage of different coordinate frames for states (e.g., geographic in the example used here) and \mathbf{P} (sensor axes) must of course be taken into account in characterizing the estimation process. An orthogonal triad $\mathbf{I}_b \mathbf{J}_b \mathbf{K}_b$ conforms to directions of sensor sight-line \mathbf{I}_b , its elevation axis \mathbf{J}_b in the normal plane, and the azimuth axis $\mathbf{I}_b \times \mathbf{J}_b$ normal to both. The instantaneous direction cosine matrix $\mathbf{T}_{b/A}$ will be known (from the sensor pointing control subsystem) at each measurement time. By combination with the transformation $\mathbf{T}_{A/G}$ from geographic to airframe coordinates (obtained from INS data), the transformation from geographic to sensor coordinates is

$$\mathbf{T}_{b/G} = \mathbf{T}_{b/A} \mathbf{T}_{A/G} \quad (14.24)$$

which is used to resolve position states along $\mathbf{I}_b \mathbf{J}_b \mathbf{K}_b$:

$$\frac{1}{|\mathbf{R}|} \mathbf{T}_{b/G} \mathbf{R} = \begin{bmatrix} 1 \\ p_A \\ -p_E \end{bmatrix} \quad (14.25)$$

where p_A and p_E — small fractions of a radian — are departures above and to the right, respectively, of the *a priori* estimated position from the sensor sight-line (which due to imperfect control does not look exactly where the tracked object is anticipated at t_m).

*Usage of relative acceleration states would have sacrificed detailed knowledge of INS velocity history, characterizing ownship acceleration instead with the random model used for the tracked object. To avoid that unnecessary performance degradation the dynamic model used here, in contrast to (14.18), has a forcing function with nonzero mean.

For application of (14.16), p_A and p_E are recognized in the role of *a priori* estimated measurements — adjusting the “dot-off-the-crosshairs” azimuth (“AZ”) and elevation (“EL”) observations so that a full three-dimensional fix (range, AZ, EL) in this operation would be

$$\begin{bmatrix} y_R \\ y_{AZ} \\ y_{EL} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{|\mathbf{R}|} & 0 \\ 0 & 0 & -\frac{1}{|\mathbf{R}|} \end{bmatrix} \mathbf{T}_{b/G} \mathbf{R} - \begin{bmatrix} 0 \\ p_A \\ -p_E \end{bmatrix} \quad (14.26)$$

Since \mathbf{R} contains the first three states, its matrix coefficient in (14.26) provides the three nonzero elements of \mathbf{H} ; e.g., for scalar position observables, these are comprised of

- The top row of $\mathbf{T}_{b/G}$ for range measurements,
- The middle row of $\mathbf{T}_{b/G}$ divided by scalar range for azimuth measurements,
- The bottom row of $\mathbf{T}_{b/G}$ divided by scalar range $\times (-1)$, for elevation measurements.

By treating scalar range coefficients as well as the direction cosines as known quantities in this approach, *both* the dynamics *and* the observables are essentially linear in the state. This has produced success in nearly all applications within the experience of this writer. The sole need for extension arose when distances and accuracies of range data were extreme (the cosine of the angle between the sensor sight-line and range vector could not be set at unity). Other than that case, the top row of $\mathbf{T}_{b/G}$ suffices for relative position states, and also for relative velocity states when credible Doppler measurements are available.

A more thorough discourse would include a host of additional material, including radar and optical sensing considerations, sensor stabilization — with its imperfections isolated from tracking, error budgets, kinematical correction for gradual rotation of the acceleration vector, extension to multiple track files, sensor fusion, myriad disadvantages of alternative tracking estimator formulations, etc. The ramifications are too vast for inclusion here.

14.3.4 Position, Velocity, and Attitude in Three-Dimensional Space (INS Aiding)

In the preceding section, involving determination of velocity history from position measurement sequences, dynamic velocity variations were expressed in terms of an acceleration vector. For nav (as opposed to tracking of an external object) with high dynamics, the history of velocity is often tied to the angular orientation of an INS. In straight-and-level Northbound flight, for example, an unknown tilt ψ_N about the North axis would produce a fictitious ramping in the indicated East velocity V_E ; in the short-term this effect will be indistinguishable from a bias n_{aE} in the indicated lateral component (here, East) of accelerometer output. More generally, velocity *vector* error will have a rate

$$\dot{\mathbf{v}} = \boldsymbol{\psi} \times \mathbf{A} + \mathbf{n}_a = -\mathbf{A} \times \boldsymbol{\psi} + \mathbf{n}_a \quad (14.27)$$

where bold symbols (\mathbf{v}, \mathbf{n}) contain the geographic components equal to corresponding scalars denoted by italicized quantities (v, n) and \mathbf{A} represents the vector, also expressed in geographic coordinates, of the total nongravitational acceleration experienced by the IMU. Combined with the intrinsic kinematical relation between \mathbf{v} and a position vector error \mathbf{r} , in a naw frame rotating at $\tilde{\omega}$ rad/sec, the 9-state dynamics with a time-invariant misorientation $\boldsymbol{\psi}$ can be expressed via 3×3 matrix partitions,

$$\begin{bmatrix} \dot{r} \\ \dot{v} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} -\tilde{\omega} \times & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (-\mathbf{A} \times) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} r \\ v \\ \psi \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{n}_a \\ \mathbf{e} \end{bmatrix} \quad (14.28)$$

which lends itself to numerous straightforward interpretations. For brevity, these will simply be listed here:

- For strapdown systems, it is appropriate to replace vectors such as \mathbf{A} and \mathbf{n}_a by vectors initially expressed in vehicle coordinates and transformed into geographic coordinates, so that parameters and coefficients will appear in the form received.
- Although both \mathbf{n}_a and \mathbf{e} appear as forcing functions, the latter drives the highest-order state and thus exercises dominant control over the data window.
- If \mathbf{n}_a and \mathbf{e} contain both bias and time-varying random (noisy) components, (14.28) is easily reexpressible in augmented form, wherein the biases can be estimated along with the corrections for estimated position, velocity, and orientation. Especially for accelerometer bias elements, however, observability is often limited; therefore the usage of augmented formulations should be adopted judiciously. In fact, the number of states should in many cases be *reduced*, as in the next two examples:
 - In the absence of appreciable sustained horizontal acceleration, the azimuth element of mis-orientation is significantly less observable than the tilt components. In some operations this suggests replacing (14.28) with an eight-state version obtained by omitting the ninth state and deleting the last row and column of the matrix.
 - When the last *three* states are omitted — while the last three rows and columns of the matrix are deleted — the result is the fully coupled three-dimensional position and velocity estimator (14.21).

The options just described can be regarded as different modes of the standard cyclic process already described, with operations defined by dynamics and measurement models. Any discrete observation could be used with (14.28) or an alternate form just named, constituting a mode subject to restrictions that were adopted here for brevity (position-dependent observables only, with distances much smaller than Earth radius).

At this point, expressions could be given for measurements as functions of the states and their sensitivities to those state variables: (14.26) provides this for range and angle data; it is now appropriate to discuss GPS information in an integrated nav context.

14.3.5 Individual GPS Measurements as Observables

The explosive growth of navigation applications within the past decade has been largely attributed to GPS. Never before has there been nav data source of such high accuracy, reachable from any location on the Earth's surface at any time. Elsewhere in this book the reader has been shown how GPS data can be used to

- Solve for 3D position and user clock offset with pseudo-range observations received simultaneously from each of four space vehicles (SVs),
- Use local differential GPS corrections that combine, for each individual SV, compensation for propagation delays plus SV clock and ephemeris error,
- Compensate via wide-area augmentation which, though not as accurate as local, is valid for much greater separation distances between the user and reference station,
- Use differencing techniques with multiple SVs as well as multiple receivers to counteract the effects of the errors mentioned and of user clock offsets,
- Apply these methods to carrier phase as well as to pseudo-range so that, once the cycle count ambiguities are resolved, results can be accurate to within a fraction of the L-band wavelength.

Immediately we make a definite departure from custom here; each scalar GPS observable will call for direct application of (14.15). To emphasize this, results will first be described for instances of sparse measurement scheduling. Initial runs were made with real SV data, taken before the first activation of selective availability (SA) degradations, collected from a receiver at a known stationary location but spanning intervals of several hours. Even with that duration consumed for the minimum required measurements, accuracies of 1 or 2 m were obtained: not surprising for GPS with good geometry and no SA.

The results just mentioned, while not considered remarkable, affirm the point that full fixes are not at all necessary with GPS. They also open the door for drawing dependable conclusions when the same algorithms are driven by simulated data containing errors from random number generators. A high-speed aircraft simulation was run under various conditions, always with no more than one pseudo-range observation every 6 sec. (and furthermore with some gaps even in that slow data rate). Since the results are again unremarkable, only a brief synopsis suffices here:

- Estimates converged as soon as the measurements accumulated were sufficient to produce a nav solution (e.g., two asynchronous measurements from each of three noncoplanar SVs for a vehicle moving in three dimensions, with known clock state, or four SVs with all states initially unknown).
- Initial errors tended to wash out; accuracies of the estimates just mentioned were determined by measurement error levels amplified by geometry.
- Velocity errors tended toward levels proportional to a ratio $(\text{RMS measurement error})/(T)$, where T here represents time elapsed since the first measurement on a course leg, or the data window — whichever is smaller. The former definition of the denominator produced a transient at the onset and when speed or direction changed.
- Doppler data reduced the transient, and INS velocity aiding minimized or removed it.
- Extreme initial errors interfered with these behavioral patterns somewhat — readily traceable to usage of imprecise direction cosines — but the effects could be countered by reinitialization of estimates with *a posteriori* values and recycling the measurements.

These results mirror familiar real-world experience (including actual measurement processing by this author); they are used here to emphasize adequacy of partial fixes at low rates, which many operational systems fail to exploit.

Although the approach just described is well known (i.e., in complete conformance to the usual Kalman filter updating cycle) and the performance unsurprising, the last comment is significant. There are numerous applications wherein SV sight-lines are often obscured by terrain, foliage, buildings, or structure of the vehicle carrying the GPS receiver. In addition, there can be SV outages (whether from planned maintenance or unexpected failures), intermittently strong interference or weak signals, unfavorable multipath geometry in certain SV sight-line directions, etc., and these problems can arise in critical situations.

At the time of this writing there are widespread opportunities, prompted by genuine need, to replace loose (cascaded) configurations by tightly coupled (integrated) configurations. Accentuating the benefit is the bilateral nature of tight integration. As the tracking loops (code loop and, where activated, carrier phase track) contribute to the estimator, the updated state enhances ability to maintain stable loop operation. For a properly integrated GPS/INS, this enhancement occurs even with narrow bandwidth in the presence of rapid change. Loop response need not follow the dynamics, only the error in perceived dynamics.

It is also noted that the results just described are achievable under various conditions and can be scaled over a wide accuracy range. Sensitivity \mathbf{H} of an individual SV observation contains the SV-to-receiver unit vector; when satellite observations are differenced, the sensitivity \mathbf{H} contains the difference of two SV-to-receiver unit vectors. Measurements may be pseudo-ranges with SA ($\sigma > 30$ m), pseudo-ranges without SA ($\sigma < 10$ m, typically), differentially corrected pseudo-ranges ($\sigma = 1$ or 2 m), or carrier phase (with ambiguities resolved, σ at 1 cm or less). In all cases, attainable performance is determined by σ and the span of \mathbf{H} for each course leg. An analogous situation holds for other nav aids when used with the standard updating procedure presented herein.

14.4 Conclusion

Principles of nav system integration have been described, within the limits of space. Inevitably, some restrictions in scope were adopted; those wishing to pursue the topic in greater depth may consult the sources which follow.

References

1. Farrell, J. L., *Integrated Aircraft Navigation*, Academic Press, New York, 1976. (Now available in paperback only; 800/628-0885 or 410/647-6165.)
2. Bierman, *Factorized Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.
3. Institute of Navigation Redbooks (reprints of selected GPS papers); Alexandria, VA, 703/683-7101.
4. Brown and Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, 1996.
5. Kayton and Fried (Eds.), *Avionics Navigation Systems*, John Wiley & Sons, New York, 1997.

Further Information

1. Journal and Conference Proceedings from the Institute of Navigation, Alexandria, VA.
2. Tutorials from Conferences sponsored by the Institute of Navigation (Alexandria, VA) and the Position Location And Navigation Symposium (PLANS) of the Institute of Electrical and Electronic Engineers (IEEE).
3. Transactions of the Institute of Electrical and Electronic Engineers (IEEE) Aerospace and Electronic Systems Society (AES).