Estimation and Measurement Algorithms

Surya P. N. Singh, spns@uq.edu.au, ELEC 3004/7312 – Digital Linear Systems: Signals & Control!

I. Overview

Estimation can be considered as an algorithm that combines *a priori* system knowledge to infer information from noisy observations. To some extent, the principle of estimation is to act as a means for reducing data to measurements, a task that is central to science and engineering. However, estimation implies more than just data processing, and includes computation to reduce error. Hence *optimal estimation*is the process of finding an estimate that minimizes the error in accordance with a specific optimality criterion.

The estimation problem need not only be considered as a single sensor measuring a single system. Rather, one of the most prominent applications of optimal estimation methods is their role in multisensor systems [13]. In this case, the job of the estimator is to distill a single optimal measure from data obtained by complementary, and potentially redundant, sensors, a familiar example of which is shown in Fig. 1.

Fig. 1. **What time is it?** In this case the multisensor fusion problem is depicted by multiple clocks each giving different readings. The goal of estimation is to combine system knowledge to give the "best" final measure for the values of times and humidity in the room.

Intuitively, given no additional information, all one can do is to trust the measurement(s). However, additional information is often available. Estimation methods attempt to provide a means for handling these cases. In other words, it addresses the question of "how to distill multiple points into one (more meaningful) result?". To some extent, the answer is prosaic and includes techniques such as "averaging" or "maximum/minimum point selection." Though these are common forms of estimation, they are not necessarily optimal results.

In the case of dynamic legged locomotion measurement, estimation is key as several forms of *a priori* knowledge are available, and include: locomotion models, governing dynamics, and sensor operational characteristics. Further, it is essential as the abrupt and discontinuous dynamics associated with the footfalls and leg thrusts corrupt the measurements, which if used uncorrected would lead to instability.

A. Definitions

Although fundamentally the same, the subject of estimation is treated with some subtle differences in the various communities that consider this problem (e.g., mechanics, signal processing, and economics). Thus, for clarity, some remarks are made on the terminology that is used to describe the estimation process.

In Newtonian mechanics, an object is free-floating and acted on by a series of forces [11]. A *static* system is the case when the total net force on an object is zero, an occurrence that is more prevalent than the definition suggests. A *dynamic* system, by contrast, is one in which the net force is not zero.

The process of estimation is based on the concept of *states*, which may be defined as the set of specific physical quantities that determine the system and its progression in time (absent external excitation(s)). [4].

The variables that make up the state, the *state variables*, are obtained from the governing mechanics. The variables chosen are not necessarily unique, but the number of variables needed to specify the state, the *system order*, is unique as this is based on the number of degrees of freedom of the underlying mechanics [4]. That is, the system may be described by different sets of variables, but each set will have the same number of variables. In which case, the choice of which set to employ is left as a design option; although the mechanics involved may suggest an obvious set.

Employing the mathematical formulation of a *state-space* allows for the state variables to be written in vector notation (i.e., a *state vector*), typically as $\mathbf{x} = [x_1, x_2, \dots x_n]$. To insure that the number of differential equations is the same as the system order, this approach requires that the differential equations are first-order with respect to time. Thus yielding the general state-space form of differential equation as $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, t)$, where $\mathbf{u}$ is a vector specifying control inputs. This has the result that higher-order differential equations need to be modeled with cascading differential sets (i.e., $\frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dx}{dt}\right)$). A *linear dynamic system* is one in which the general dynamic system function may be represented as a linear system (i.e., the function $f(\mathbf{x}, \mathbf{u}, t)$ is linear). Examples of such systems include a lumped mass acted upon by a linear spring and pendulums having small amplitudes.

When combined with linear algebra techniques, this provides a significant mathematical and notational convenience. For a state vector of dimension n, $\mathbf{x}$, and a control input of dimension m, $\mathbf{u}$, the output of dimension l, $\mathbf{y}$, may be represented in canonical vector-matrix differential equation form as:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{1}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \tag{2}$$

where $\mathbf{u}$ may be represented as $\mathbf{u} = -G\mathbf{x}$ (with $-G$ being a gain matrix having a minus sign for negative feedback) and the matrices A (dimension $n \times n$), B (dimension $n \times m$), C (dimension $l \times n$), and D (dimension $l \times m$) come from the linearized system model. Also, if the system is *time invariant*, this is further simplified as $\mathbf{A}$ and $\mathbf{B}$ are constant giving $H(s)$.

$$H(s) = C(sI - A)^{-1}B \tag{3}$$

The difficulty, and principal advantage, of this approach is that the equations of motion need to be in the form of differential equations. While this complicates the modeling, the state space approach is generally more physically intuitive, yields a result that may be applied directly, and importantly allows for statistical descriptions [13].

In this sense, estimation may be considered as calculating the state vector from a series of observations corrupted by noise. In particular, *filtering* is the resultant estimate of the state-vector for a current time based on all past measurements. This compares with *prediction* which is the estimate of the state vector at some future time and *smoothing* for an estimate at some prior time [13].

For completeness, it should be noted that these definitions are generally couched in a controls viewpoint and vary slightly from those in other domains. For instance, in the computational sciences [17] the definition of state is expanded to include **all** aspects of the environment and system operation; and, it is further categorized as being static or dynamic on the basis of whether these variables change over a given interval. This contrasts with the aforementioned dynamics definition as it is primarily time-based as opposed to being force-based. The difference in definitions is illustrated by an example of a vehicle moving with constant velocity, which is static in the former case and dynamic in the latter.

B. Background

Measurement errors are inevitable. While some early methods for estimation may be traced back to Galileo , formal estimation methods, in particular the method of least squares, were introduced circa 1800 and are generally attributed to Gauss [8]. In addition to least squares, Gauss also introduced the notion of a *normal distribution* [13]. Typically expressed as $\mathcal{N}(x; \mu, \sigma)$, it is a key parametric function that allows for the description of a data set by its first and second moments, which are the statistical values of the mean (μ) and the standard deviation (σ). This is expressed for one dimensional case (i.e., with the scalar, x representing the data) as:

$$\mathcal{N}(x; \mu, \sigma) = \frac{1}{\left(\sqrt{2\pi}\right)\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{4}$$

and, for the multivariate case (i.e., with the vectors $\mathbf{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ for the multidimensional data, means, and covariance matrix respectively) as:

$$\begin{aligned} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})\right) \\ \mathbf{x} &= [x_1, x_2, \ldots, x_n] \\ \boldsymbol{\mu} &= [\mu_1, \mu_2, \ldots, \mu_n] \\ \boldsymbol{\Sigma} &= \begin{bmatrix} {\sigma_1}^2 & \cdots & \sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \sigma_n\sigma_1 & \cdots & {\sigma_n}^2 \end{bmatrix} \end{aligned} \tag{5}$$

where the vector $\mathbf{x}$ is multidimensional data taken for dimensions 1 to n, the vector $\boldsymbol{\mu}$ are the most likely average values for each dimension, and the covariance matrix, $\boldsymbol{\Sigma}$, is a positive semidefinite symmetric matrix holding the variances or piecewise correlations between all dimensional variables. In certain cases, it may also be consider as a matrix expression of a variance tensor.

In addition to providing a compact (two parameter) space for representing distributions, it also provides a first means for encoding stochastic concepts, a consideration that is increasingly tractable with rapid digital computing. When encoding probabilities the normal distribution belongs to the class of functions known as *probability density functions* that describe the likelihood of an event through a continuum. Just as probabilities are bounded to the set $[0, 1]$, the distribution function, $p(\mathbf{x})$, is bounded such that $\int p(\mathbf{x})dx = 1$.

Estimation in a probabilistic context was later treated by maximum likelihood methods in which the result is the event(s) which occurred with maximum probability given the observed data [13]. This framework allows for a probabilistic treatment of the data, but not the system. This concern was treated with the inclusion of random process theory to yield the Wiener filter (in the continuous case) and the Chapman-Kolmogorov equation (for the discrete case) [14]. However, the Wiener filter operates in the frequency domain and uses correlation functions and impulse responses that limit it to stationary systems (though this was later extended to non-stationary regimes). The work of Kalman [9], Bucy [10], Swerling [16], and others led to the treatment of the estimation problem in state-space and, in particular, in discrete-time, resulting in an optimal estimator for linear dynamic systems with white noise, *the Kalman filter*.

Concurrently, a more complete mathematical treatment of probabilities was being developed. This study considered the process of prediction (and hence also estimation), but primarily in the form of inference from PDFs as compared to plant models and their corresponding differential equations.

A popular, and central, tool for solving this inference is Bayes theorem [14]:

$$p(\mathbf{x}|\mathbf{Z}) = \frac{p(\mathbf{Z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{Z})} \tag{6}$$

where $\mathbf{x}$ is a vector of mutually exclusive events and $\mathbf{Z}$ is vector of measurements made for each event. In the trivial case with no errors, the events and their measurements are identical giving that $p(\mathbf{x}|\mathbf{Z}) = p(\mathbf{Z}|\mathbf{x})$. This approach offers the advantage of conveniently representing both the estimated system and the observed data fully by probabilistic models [13]. This is especially beneficial with modern computing techniques as this provides an on-line mechanism for conditioning of the base probabilities (i.e., updating these values to account for new information), thus allowing for the creation of models through numerical data processing. Implicitly, though, the conditioning process assumes a *Markov chain*; that is, future progression from a given current value is not a function of prior values (or in other words retroactively influenced by the path to the current point). When used to find the most probable outcome given a measurement this process is referred to as *Bayesian estimation*.

Modern estimation methods in this domain have extended these concepts well beyond the normal function. Popular variants include: multiple model techniques such as the multi-modal filter in [3], the Assumed Density Filter (ADF) which uses moment matching to calculate true mean and variance [17]; the Unscented Kalman Filter (UKF) which accounts for non-linear environments by basing its estimate of future states on multiple points around the mean [7]; and the Particle Filter [14] which makes several predictions at each discrete step in a Monte-Carlo manner. As with the normal methods, these routines do best when the underlying model faithfully represents the dynamics. Thus a significant challenge in estimation remains application of the correct models at the correct times as too complex a model requires excessive resources, while too simple an approach is doomed by modeling errors.

II. Linear Dynamic State Estimation Review

The state-space form is an equivalent representation of scalar second order ordinary differential equations (ODE), and thus is especially convenient for modeling linear dynamic systems. The goal of estimation in this space is to recover the state-vector from noisy measurements and is performed using methods ranging from averaging (i.e., least-squares) to full non-parametric simulations (i.e., Monte Carlo techniques).

A. Least-squares

The typical approach to extracting signals from noisy data is the method of least-squares. For polynomial linear dynamics systems, this may be considered as curve-fitting of a polynomial against time; that is, finding the coefficients p_1 and p_0 for the equation $\hat{x} = p_1 t + p_0$ from a data series z_i collected at t_i having n points (and with $n \geq 2$ for this linear case). The solution of this may be found with linear algebra techniques such as matrix inversion (as shown in equation 7) or singular value decomposition (SVD) [19].

$$\begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} n & \sum_1^n t_i \\ \sum_1^n t_i & \sum_1^n t_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_1^n z_i \\ \sum_1^n t_i z_i \end{bmatrix} \tag{7}$$

One issue with the approach as presented is that a its does not work well for linear ordinary differential equations whose solutions are not readily expressed as linear polynomials, such as a simple sinusoidal resonator given by $\ddot{x} = -\omega^2 x$. A principal issue with this procedure, however, is that although the data may be random and noisy, calculation of the estimate is deterministic. Least squares has no means for accounting for variance of the source signal or measurements. With enough points from a linear polynomial data set, the estimate will converge to the governing linear function, but its estimate will be in error during short time intervals. Although a batch method is shown, recursive least squares methods are available and are given in [19] and [15].

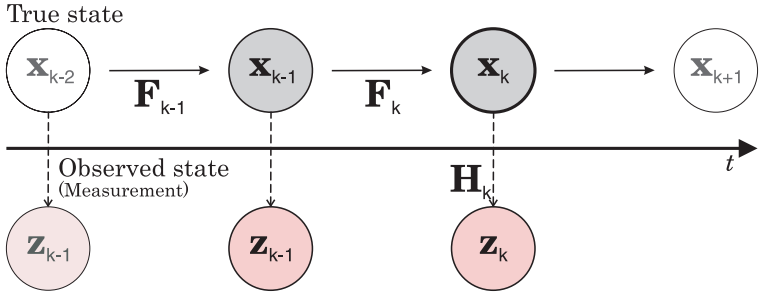

Fig. 2. General description of the estimation problem viewed as reconstructing the true state from partial, noisy observations

B. Observers

As shown in Figure 2, the linear estimation problem may be considered to be a mode in which observations are made from a hidden true system. A classical tool for considering and analyzing this is the notion of an observer.

The observer framework treats the measurement as the output of the "process." Using equation 2 as a basis, this gives a linear measurement model as $Vy = \mathbf{Cx}$. Methods based on this notion provide a means for characterizing the variables that are visible to a system by estimating those variables that are not directly measured with specific rates of asymptotic convergence of the error given no noise [4]. Such estimation of the unmeasured variables might actually be preferable to directly measuring them given the relative strength or ease of particular sensing modalities. However, care must be taken to ensure that the portion of the state space under consideration is indeed observable. That is, that system variables are properly feedback. For example, a system comprising of a mass and a force is unobservable in position if the *only* measurement is the acceleration. This is because, even with the double integration, an initial position error will last indefinitely. Details of the convergence and design of observers is beyond the scope of this section and detailed in several references such as [12] and Chapter 7 of [4].

The standard observer model presented does not explicitly account for noise effects. While it provides a means for designing convergence of the measurement model to the system model, an optimality criteria is not specified. Achievement of this requires the introduction of concepts from random process representation and optimal control. These subjects are briefly surveyed here as they are extensively reviewed in [18] and [4] amongst others. As it will turn out, the linear least squares optimal result in a perfectly linear regime is *the Kalman filter*.

C. Random process representation

A convenient function for random processes is the normal distribution as it is both unimodal (i.e., has a single most common value) and simply specified by two parameters. Estimation with these functions is convenient and also referred to as Gaussian filters [17] or Estimation with Random Process [4]. Another advantage of the Gaussian representation is that it does not require massive quantities of data to define the model with confidence. This is convenient in highly dynamic environments where a statistical model may need to be reconditioned regularly; that is, the dynamic conditions lead to the model expiring rapidly.

The noise characteristics may be simplified in certain cases, such as when the process is stationary (i.e., measurements are not dependent of the absolute value of the time when they were taken) or when the experiments are ergodic (i.e., one sample represents the entire process). Most importantly for estimation is the consideration that measurement noise may be modeled solely as being white noise. This implies that the noise has uniform power through a spectrum, a uniform covariance, and a zero mean. As the frequency spectrum for this function is uniform, it models white (broad-spectrum) noise well, but does poorly for shot-noise and other deterministic effects.

Considering the random process as being additive only, allows for the noise to be solely represented by a vector added to the state space description. Thus, we introduce two random process vectors: v for the output noise and w for the systematic uncertainty[1]. Adding this to the standard state-space form (Eq. 1 and 2) gives:

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} + w \tag{8}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du} + v \tag{9}$$

The transformation of these equations from the control domain to the estimation domain is treated by taking advantage of the duality between these two areas. As elaborated in the next section, the measurement of estimation is analogous to the output of controls; that is, the result, $\mathbf{y}$, in Eq. 9 represents the measured data.

[1]The choice of notation for the random processes is **not** standard in the literature. For instance, the notation adopted (with v associated with the output or measurement) follows the style of [6], [19], [13], [5]. References [4], [14] switch this so that w is associated with the output and v with the system dynamics. In [17] the vectors δ_t and ε_t are used for the discrete measurement and system noise respectively. Finally, the initial Kalman paper [9] treats the noise with the control vector $\mathbf{u}$ and thus provides no definitive standard.

D. Estimation Control Duality

As indicated above, an interesting result of linear systems is that for each observable dynamic system in a filtering problem, there exists a corresponding dynamic system for the control problem [13]. The interdependency present is hardly surprising given that both estimation and control share the same plant. This similarity can be extended more formally in the description of a duality between the estimation and controls processes in which we consider the ideal estimator to be one that processes plant observations to return the current state, as compared to the ideal controller which processes the current state to return a desired plant outcome.

	Estimation		**Control**
Model:	$\dot{\mathbf{x}} = \mathbf{F}\mathbf{x}$ (discrete: $\mathbf{x} = \mathbf{F}_k\mathbf{x}$)	$\leftrightarrow$	$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, $\mathbf{A} = \mathbf{F}^{\mathsf{T}}$
Regulates:	P (covariance)	$\leftrightarrow$	M (performance matrix)
Minimized function:	Q (or GQG^{T})	$\leftrightarrow$	V
Optimal Gain:	K	$\leftrightarrow$	G
Completeness law:	Observability	$\leftrightarrow$	Controllability

TABLE I
THE DUALITY BETWEEN ESTIMATION AND CONTROL

The advantage of this duality is that it allows the use of optimal control concepts in the determination of an optimal observer (sometimes referred to as an optimal estimator). Optimal control laws work by minimizing the deviations of the state from the desired outcome (or origin) as measured by a performance integral (i.e., through the minimization of an integral cost function). This problem is extensively treated in the controls literature, where it is noted that the solution of a general integral equation is not generally solvable except for special cases. One of these cases is in the linear dynamic regime with a quadratic integral function for the error. The rationale for this may be considered to be akin to that of the least squares criterion. Also while a quadratic criterion may well not be the design objective, it has an amenable analytic solution that translates well to the discrete domain. Typically, these problems are constructed as regulators or controllers whose objective is to converge to a zero value.

III. KALMAN FILTER

Based on the previous section, the Kalman filter (KF) may be viewed as a linear estimation technique having a least-squares optimal observer gain that returns the current state mean and variance values. A complete derivation of the KF in this sense is given in [13] and with examples in [19]. The advantage of this approach is that simplifies the modeling, especially from the dynamic equations of motion perspective taken in the modeling chapter. It also directly ties into an optimal control strategy.

It should be noted that the optimal observer perspective is but one view of the Kalman filter. Given the approximate and uncertain nature of the models, they may be replaced with statistical representations of the state. Under a maximum-likelihood philosophy, the KF uses statistical functions for the measurement and model equations for the system prediction. Another complimentary, and arguably more flexible, statistical approach is to derive this filter from a Bayesian, perspective as introduced in [6]. A review of the KF from this perspective is given in [1] with extensions to particle filters in [14] and applications primarily in robotics given in [17]. The power of this perspective is the ability to use data to derive the models and to adapt these models via probabilistic conditioning.

A. Notation

The following notation is introduced to describe variables associated with the operation and development of the KF:

$\mathbf{x}$: the state vector

$\mathbf{x}_{A|B}$: the state of $\mathbf{x}$ at time A based on data taken up to time B

$\hat{\mathbf{x}}$: estimate of the true state vector

$\mathbf{F}$: system dynamics matrix in continuous time (equivalent to $\mathbf{A}$ in Eq. 1)

$\mathbf{G}$: system control matrix relating deterministic input, $\mathbf{u}$, to the state (equivalent to $\mathbf{B}$ in Eq. 1)

$\mathbf{H}$: measurement matrix in continuous time (equivalent to $\mathbf{C}$ in Eq. 2)

$\mathbf{F}_i$: system model in **discrete** time at $t = t_i$

$\mathbf{H}_i$: measurement model in **discrete** time at $t = t_i$

$\mathbf{P}_i$: estimate covariance in **discrete** time at $t = t_i$

$\mathbf{w}$: process uncertainty (noise) vector (of type $\mathcal{N}(0, s)$)

$\mathbf{Q}$: process noise matrix, $\mathbf{Q} = E\left[ww^{\mathsf{T}}\right]$

$\mathbf{Q}_i$: $\mathbf{Q}$ in discrete time at $t = t_i$

$\mathbf{v}$: measurement noise vectors (of type $\mathcal{N}(0, \sigma)$)

$\mathbf{R}_i$: the measurement variance matrix, $\mathbf{R} = E\left[vv^{\mathsf{T}}\right]$, in discrete time at $t = t_i$

Thus, the state-space description for the system is (Eq. 8):

$$\dot{\mathbf{x}} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{u} + \mathbf{w} \tag{10}$$

and, recalling that the measurement in estimation is the equivalent of an output in the control case, the measurement becomes :

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v} \tag{11}$$

For clarity, an illustration of the estimation process with the above notion is given in Fig. 3.

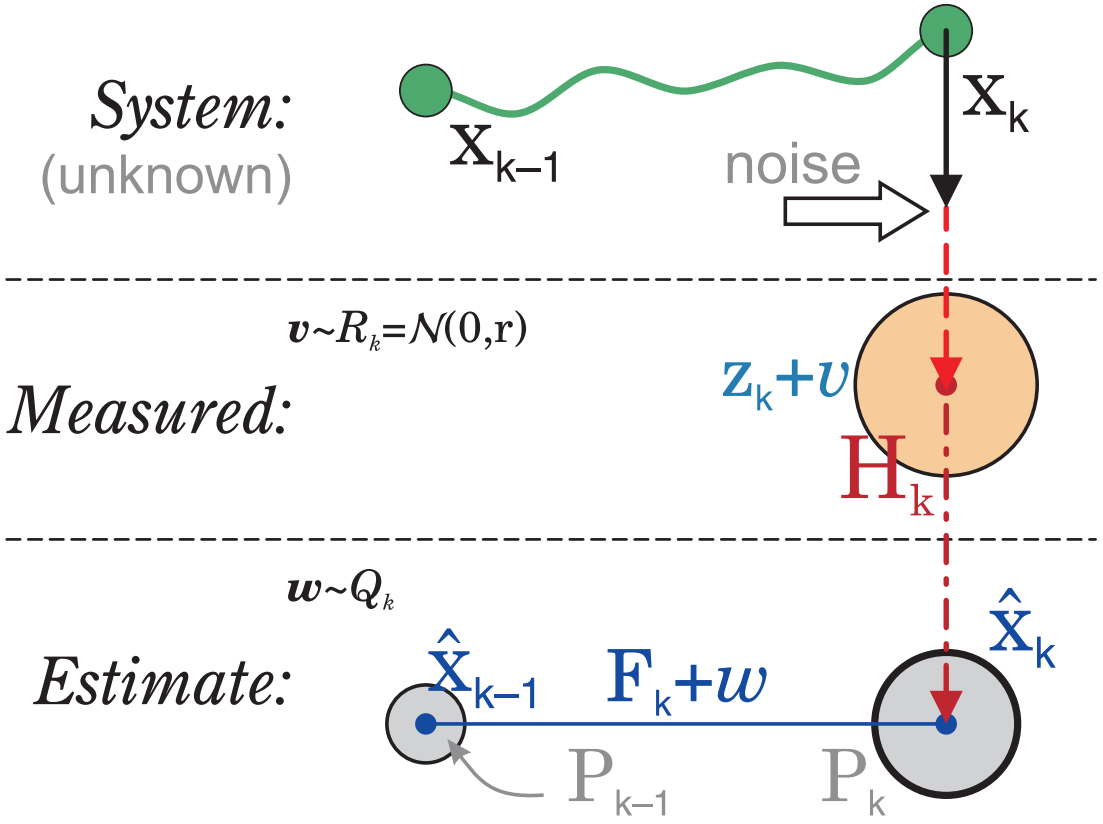

Fig. 3. The estimation process as in terms of the above notation. Here the state moves along some unknown trajectory and its partial measurements are obscured by unknown noise. The KF combines knowledge in the form of a linear model, $\mathbf{x}$, and its covariance, $\mathbf{P}$ to give a final improved estimate

B. Assumptions

Under several (ideal) assumptions, the KF provides a least-squares optimal solution. These assumptions include linear system dynamics and a mutually independent, zero-mean Gaussian (white) noise. The derivation of this is detailed in [13] and demonstrated generally in [19] and with presentation specific to mobile robots in [2].

In order to remain a Gaussian linear process, it is imperative that the future states of the filter remain Gaussian. This occurs when a Gaussian is operated on by either a linear function or another Gaussian function. This can be proved to occur given that the following assumptions hold [14]:

1) The system function ($\mathbf{F}_{k-1}$) and the measurement function ($\mathbf{H}_k$) are both known linear functions.
2) The random noise distributions ($\mathbf{w}_{k-1}$ and $\mathbf{v}_k$) are Gaussian densities of known parameters. This implies that they are mutually independent, zero-mean white Gaussian functions with covariances Q_{k-1} and R_k.
3) The process and measurment noise terms are uncorrelated (i.e., $E[w(i)v(j)] = 0 \forall i, j$)

C. Discretization

As the governing dynamics are often considered in continuous space, it is necessary to transform them to discrete space so as to be properly applied with the discrete KF. While a continuous form of the KF exists [19], the discrete form is ideal as the measurement stream and digital processing are discrete.

Starting with Eq. 10 and 11, discretization of these relations is performed for a known sampling time t_s via the fundamental matrix, $\Phi(t)$ such that $\mathbf{F}_k = \Phi(t_s)$, with:

$$\Phi(t) = \mathscr{L}^{-1}\left[(\mathbf{sI} - \mathbf{F})^{-1}\right] \tag{12}$$

where $\mathscr{L}^{-1}$ represents the inverse Laplace transform. As inverse transforms are not always available, an approximate value for $\Phi(t)$ may be found by Taylor series expansion via:

$$\begin{aligned} \Phi(t_s) &= \boldsymbol{e}^{\mathbf{F}t_s} = \mathbf{I} + \mathbf{F}t_s + \frac{(\mathbf{F}t_s)}{2} + \dots \\ \Phi(t_s) &\approx \mathbf{I} + \mathbf{F}t_s \end{aligned} \tag{13}$$

For the system model, the process uncertainty and gain measurement have to similarly be discretized to get the process and measurement noise matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ respectively. This is performed by:

$$\mathbf{Q}_k = \int_0^{t_s} \Phi\left(t\right)\mathbf{Q}\Phi^{\mathsf{T}}\left(t\right)dt \tag{14}$$

$$\mathbf{G}_k = \int_0^{t_s} \Phi\left(t\right)\mathbf{G}dt \tag{15}$$

The discretization of the measurement noise is much simpler since the measurement is inherently discrete, thus all that is required is the covariance between each state variable, i, and each sensed variable, j, giving:

$$\mathbf{R}_k[i,j] = {\sigma_{ij}}^2 \tag{16}$$

it should be noted that in several cases that the variables are completely independent with the result that the $\mathbf{R}$ matrix becomes diagonal. In these cases, the system noise matrix, $\mathbf{Q}$, maybe approximated in a similar manner as also being diagonal with:

$$\mathbf{Q}_k[i,j] \approx \begin{cases} t_s \cdot {s_{ij}}^2 \in i = j \\ 0 \in i \neq j \end{cases} \tag{17}$$

an example of such a simplification is seen in the kinematic case of the Kalman Filter (detailed in section III-E).

D. Description of the Discrete Filter

As mentioned, the recursive KF may be considered as an online discrete observer that essentially acts to average measurements against dynamic linear system model predictions in a weighted manner to give an optimal estimate in least-squares manner given the aforementioned assumptions. The process is sketched in Fig. 4.

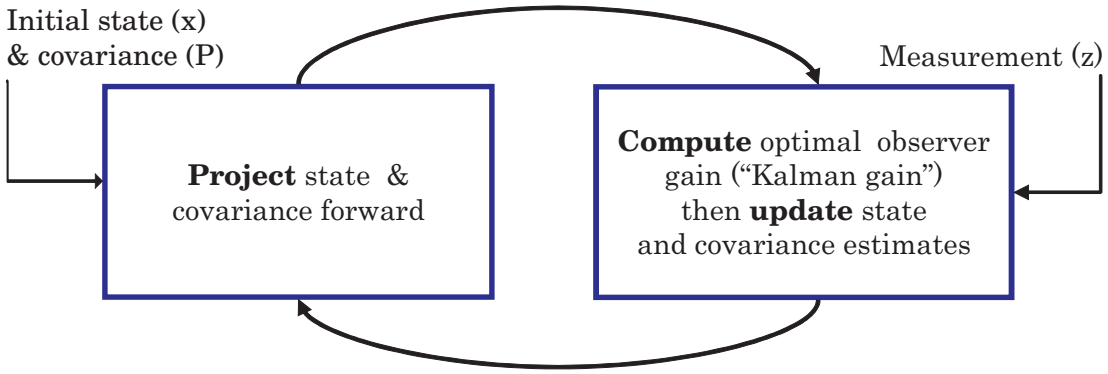

Fig. 4. The recursive KF process operates on discrete measurements by predicting future state actions and then updating this with the measurement to give the final estimate.

The Kalman Gain, $\mathbf{K}_k$ scales the update between the measurement and the prediction and is derived via the Riccati equations [19]. Thus the final state estimate, $\hat{\mathbf{x}}_{k|k}$, is found recursively as follows:

$$\text{Prediction: } \hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}, \quad \text{(state prediction)} \tag{18}$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}{\mathbf{F}^{\mathsf{T}}}_{k-1}, \quad \text{(covariance prediction)} \tag{19}$$

$$\text{Kalman Gain: } \mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}}[\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}} + \mathbf{R}_k]^{-1}, \tag{20}$$

$$\text{Update: } \mathbf{P}_{k|k} = [\mathbf{I} - \mathbf{K}_k\mathbf{H}]\mathbf{P}_{k|k-1}, \quad \text{(covariance update)} \tag{21}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}) \quad \text{(state update)} \tag{22}$$

These equations verify intuition about the KF as an optimal observer. Eq. 22 shows the Kalman gain acts to weight measurement data from the *innovation*, or the difference between the actual measurement and forward predicted state, to the final estimate relative to the system and measurement covariances. That is, if $\mathbf{Q}_k$ is small the estimator will treat the system model as being excellent. Second, Eq. 21 shows the estimate covariance is reduced by additional information, even if that information has large uncertainty (i.e., $\mathbf{P}_{k|k} \propto (\mathbf{I} - \mathbf{R}_k^{-1})\mathbf{Q}_{k-1}$). In this way the KF operates so as to minimize the variance of the error in the estimate and in do doing acts as the (linear least squares) optimal estimator for the linear estimation problem.

E. Kinematic Kalman Filter

A particular case of the KF occurs when estimating a system with no deterministic inputs and whose differential equations yield polynomial results. This results in a simplification of the system dynamics, $\mathbf{F}$, and fundamental matrix. In practice, this serves to represent the kinematic relations between the state variables for a rigid body, hence the "kinematic" termonology.

This gives the $\mathbf{F}$ matrix as:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{23}$$

The discrete fundamental matrix, $\mathbf{F_k}$, may be found using Eq. 12 or 13. Such cases are not uncommon. For example, a static body (or a body for which no dynamic information is available) in an inertial navigation application can track position by integrating signals and adding them to the initial value. For a free body in space, the accelerations would be constant, unless it is restrained to motion on the ground plane (i.e. always normal to the gravity vector) in which case it would be zero. Such a system may be represented with a second order state space of $\mathbf{x} = [y\, \dot{y}]$. This results in the following polynomial state space formulation for the position tracking problem:

$$\begin{bmatrix} \dot{y} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ \dot{y} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ g \end{bmatrix} \tag{24}$$

This results in the state space form of equations with $\dot{x_1} = x_2$ and $x_2 = u$, whose solution is given by:

$$\mathbf{x}(t) = c \exp(\mathbf{A}t) \tag{25}$$

where c is an integration constant. This result gives the expected polynomial Newtonian equations:

$$\ddot{y} = -g \tag{26}$$

$$\dot{y} = gt + p_1 \tag{27}$$

$$y = p_0 + p_1 t + \tfrac{gt^2}{2} \tag{28}$$

where g represents gravity and the constants p_0 and p_1 are from initial conditions. A more interesting example in estimation is when noise and acceleration bias are added. If a liberal definition of the state space is employed, it may be expanded to being three dimensional (i.e., $\mathbf{x} = [y\, \dot{y}\, \mathrm{BIAS}]$. In a similar manner to the equations above, this gives an expanded state space, the result of which may be represented by the following $\mathbf{F}$ and $\mathbf{F_k}$ matrices.

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{29}$$

$$\mathbf{F_k} = \begin{bmatrix} 0 & t_s & \frac{{t_s}^2}{2} \\ 0 & 0 & t_s \\ 0 & 0 & 0 \end{bmatrix} \tag{30}$$

which gives the optimal observed estimate for the position as

$$\hat{\mathbf{x}}_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1}) \tag{31}$$

whose simplified result in the case of no measurement error matches the expected result from mechanics given in Eq. 28.

For small discrete time intervals, this case also allows for a significant simplification of system noise. Consider the case where the only variable with systematic noise is unique to each state (i.e., no cross coupling). This would give a simple $\mathbf{Q}$ of

$$\mathbf{Q} = \begin{bmatrix} s_1^2 & 0 & 0 \\ 0 & s_2^2 & 0 \\ 0 & 0 & s_3^2 \end{bmatrix} \tag{32}$$

where the variances, s_1^2, s_2^2, s_3^2 would typically be found from experimental data (i.e., tuned). The approximate result from Eq. 17 is very similar to the exact result solved by Eq. 14 for a time interval t_s is given by:

$$\mathbf{Q}_k = \begin{bmatrix} s_1^2 t_s + \frac{s_2^2 {t_s}^3}{3} + \frac{s_3^2 {t_s}^5}{20} & \frac{4s_2^2 {t_s}^2 + s_3^2 {t_s}^4}{8} & \frac{s_3^2 {t_s}^3}{6} \\ \frac{4s_2^2 {t_s}^2 + s_3^2 {t_s}^4}{8} & s_2^2 t_s + \frac{s_3^2 {t_s}^3}{3} & \frac{s_3^2 {t_s}^2}{2} \\ \frac{s_3^2 {t_s}^3}{6} & \frac{s_3^2 {t_s}^2}{2} & s_3^2 t_s \end{bmatrix} \approx \begin{bmatrix} s_1^2 t_s & 0 & 0 \\ 0 & s_2^2 t_s & 0 \\ 0 & 0 & s_3^2 t_s \end{bmatrix} \tag{33}$$

A polynomial system model with zero system noise carries no more information than a least squares calculation. As might be expected, the polynomial KF with infinite initial conditions gives an identical result to that of a recursive least squares filter [19].

IV. Practical Considerations

Linear Gaussian estimators, such as the Kalman filter, have the advantage of being simple and efficient. Their simplicity derives from both the limited number of parameters needed to describe a state and the linear transition between states. In particular, each dimension of the state is entirely modeled by only **two** parameters: the mean (μ) and the covariance (σ). The linearity allows for representation and computation by linear algebra, a boon when being implemented on a computer.

To generalize, it is important to keep the uncertainty of the state vector small when applying the EKF.

1) Matrix sizes: For ease of implementation, equation 18 is annotated below with the dimensions of the key terms where n is the dimension of the state-space (i.e., number of variables are in the state-vector), m is the measurement dimension (i.e., number of variables being sensed), and j is the dimension of the control vector (i.e., the number of variables being controlled):

$$\begin{aligned} \underbrace{\hat{\mathbf{x}}_{k|k-1}}_{n\times 1} &= \underbrace{\mathbf{F}_{k-1}}_{n\times n}\hat{\mathbf{x}}_{k-1|k-1} + \underbrace{\mathbf{G}_{k-1}}_{n\times j}\underbrace{\mathbf{u}_{k-1}}_{j\times 1} \\ \underbrace{\mathbf{P}_{k|k-1}}_{n\times n} &= \underbrace{\mathbf{Q}_{k-1}}_{n\times n} + \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}{\mathbf{F}^{\mathsf{T}}}_{k-1} \\ \underbrace{\mathbf{K}_k}_{n\times m} &= \mathbf{P}_{k|k-1}\underbrace{\mathbf{H}^{\mathsf{T}}}_{n\times m}\underbrace{\left[\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}} + \mathbf{R}_k\right]^{-1}}_{m\times m} \\ \mathbf{P}_{k|k} &= [\mathbf{I} - \mathbf{K}_k\mathbf{H}]\mathbf{P}_{k|k-1} \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\underbrace{\mathbf{z}_k}_{m\times 1} - \underbrace{\mathbf{H}}_{m\times n}\hat{\mathbf{x}}_{k|k-1} - \mathbf{H}\mathbf{G}_k\mathbf{u}_{k-1}) \end{aligned} \tag{34}$$

A. Efficiency

Considering that a motivating interest for the estimation process is the synthesis of an optimal control strategy, the time required by the estimation process is a practical concern. While a minimal delay is ideal, a pertinent design consideration is not only the minimal estimation time, but also the variation in estimator performance for a marginal increase in time. In other words, does the improved estimate compensate for the cost to the controller arising from an increased delay.

An intuitive explanation for this result is that since the method cycles through dynamical models instead of computing (or rather linearising) them anew throughout the process, it is able to reuse prior results and better able to take advantage of the rhythmical structure present.

This compares favourably to the traditional EKF techniques which are on the order of $\mathbf{O}\left(\mathbf{k^{2.4}} + \mathbf{n^2}\right)$. Where k is the dimension of the state space and n is the dimension of the measurement space.

A second, related efficiency metric, that is popular in system design, is the memory (or resource) requirements of this estimator.

It should be noted that a small memory footprint provides an engineering benefit by allowing for compiler optimization by having the necessary data reside in high-speed cache memories.

V. Summary

The previous sections have demonstrated that the recursive EKF estimate is optimal **if** the *high restrictive* assumptions hold. That is to say, that no algorithm can do better in a least squares sense for this space.

REFERENCES

[1] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, New York, 2001.

[2] B. Barshan and H. F. Durrant-Whyte. Inertial navigation systems for mobile robots. *IEEE Transactions on Robotics and Automation*, 11(3):328–342, 1995.

[3] H.A.P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, 1988.

[4] B. Friedland. *Control system design: an introduction to state-space methods*. McGraw-Hill, New York, 1986.

[5] M. Grewal and A. Andrews. *Kalman Filtering : Theory And Practice Using MATLAB*. John Wiley, New York, 2nd edition, 2001.

[6] Y. Ho and R. Lee. A bayesian approach to problems in stochastic estimation and control. *Automatic Control, IEEE Transactions on*, 9(4):333–339, 1964.

[7] S. J. Julier and J. K. Uhlmann. New extension of the kalman filter to nonlinear systems. In Ivan Kadar, editor, *Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068, pages 182–193. SPIE, 1997.

[8] T. Kailath. A view of three decades of linear filtering theory. *IEEE Transactions on Information Theory*, 20(2):146–181, 1974.

[9] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Transactions of the ASME. Series D)*, 82:35–45, March 1960.

[10] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering (Transactions of the ASME. Series D)*, 83:95–107, December 1961.

[11] T. R. Kane, P. W. Likins, and D. A. Levinson. *Spacecraft dynamics*. McGraw-Hill, New York, 1983.

[12] D. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control,*, 16(6):596–602, 1971.

[13] Technical Staff of the Analytic Sciences Corporation. *Applied Optimal Estimation*. MIT Press, Cambridge, 1974. Ed: A. Gelb.

[14] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: Particle Filters for Tracking Applications*. Artech House, Boston, 2004.

[15] A. H. Sayed. *Fundamentals of Adaptive Filtering*. IEEE Press, New York, 2003.

[16] P. Swerling. First order error propagation in a stagewise smoothing procedure for satellite observations. *The Journal of the Astronautical Science*, 6(3):46–52, 1959.

[17] S. Thrun, W. Burgard, and D. Fox. *Probabilistic robotics*. MIT Press, Cambridge, MA, 2005.

[18] K. Watanabe. *Adaptive estimation and control*. Prentice Hall, Englewood Cliffs, NJ, 1991.

[19] P. Zarchan and H. Musoff. *Fundamentals of Kalman Filtering: A Practical Approach*. American Institute of Aeronautics and Astronautics, Reston, VA, 2000.