Aalto University

STATE ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS

- Present the optimal discrete-time estimator.
- Discuss the difficulty in implementing it in practice due to memory requirements and computational requirements.
- Discuss numerical implementation of the optimal discrete-time estimator.
- Derive the suboptimal filter known as the extended Kalman Filter

ESTIMATION IN NONLINEAR STOCHASTIC SYSTEMS

The **general state space model** for discrete-time stochastic systems

$$x(k+1) = \mathbf{f}[k, x(k), u(k), v(k)]$$

$$z(k) = \mathbf{h}[k, x(k), w(k)]$$

At least one of the above two functions has to be nonlinear.

The noise sequences, white with known pdf and mutually independent.
The initial state, a known pdf and be independent of the noises.

The optimal nonlinear state estimator consists of the computation of the conditional pdf of the state x(k) given all the information available at time k: the prior information about the initial state, the intervening

inputs and the measurements through time k.

The Optimal Estimator

The **information set** available at k , growing with k

$$I^k = \{Z^k, U^{k-1}\}$$

For a stochastic system, an **information state** is a function of the available information set that completely summarizes the past of the system in a probabilistic sense.

As shown in the next subsection, the conditional pdf

$$p_k \triangleq p[x(k)|I^k]$$

satisfies this requirement if the two noise sequences (process and measurement) are *white and mutually independent*.

The **optimal estimator** consists of the *recursive functional relationship* between **the information states** $\mathbf{p_{k+1}}$ and $\mathbf{p_k}$, and is given by

$$\boxed{p_{k+1} = \frac{1}{c} p[z(k+1)|x(k+1)] \int p[x(k+1)|x(k), u(k)]\, p_k \, dx(k)}$$

Remarks

1. The implementation requires storage of a pdf, equivalent to an infinite-dimensional vector. The problem of carrying out the integration numerically.

2. In spite of these difficulties, the use of the **information state** has the following advantages:

    a. It can be approximated over a grid of points or by a piecewise analytical function.

    b. It yields directly the MMSE estimate of the current state — the conditional mean — and any other desired inf. (its conditional variance).

3. For **linear systems with Gaussian white noises**, the functional recursion becomes the Kalman Filter, the state's conditional mean and covariance define sufficient statistic

4. If a system is linear but the noises and/or the initial condition are not Gaussian, then, in general, there is no simple sufficient statistic and the recursion has to be used to obtain the optimal MMSE estimator.

Proof of the Recursion of the Conditional Density of the State

Bayes' formula

$$
\begin{aligned}
p_{k+1} &\triangleq p[x(k+1)|I^{k+1}] = p[x(k+1)|I^k, z(k+1), u(k)] \\
&= \frac{1}{c} p[z(k+1)|x(k+1), I^k, u(k)] p[x(k+1)|I^k, u(k)]
\end{aligned}
$$

If the measurement noise is white in the following sense: w(k + 1) conditioned on x(k + 1) has to be independent of w(j), j ≤ k and of v(j), j ≤ k, then

$$p[z(k+1)|x(k+1), I^k, u(k)] = p[z(k+1)|x(k+1)]$$

the state prediction pdf (a function, rather than the point prediction — the predicted value — as in the linear case)

$$p[x(k+1)|I^k, u(k)] = \int p[x(k+1)|x(k), I^k, u(k)] p[x(k)|I^k, u(k)] \, dx(k)$$

Chapman-Kolmogorov equation

Proof continues

the process noise sequence is white and independent of the measurement noises in the following sense: v(k) conditioned on x(k) has to be independent of v(j − 1), w(j), j ≤ k. The whiteness of v(k) is equivalent to requiring the state vector x(k) to be a Markov process.

$$p[x(k+1)|x(k), I^k, u(k)] = p[x(k+1)|x(k), u(k)]$$

state transition pdf

since the input u(k) enters the system after the realization of x(k) has occurred

$$p[x(k)|I^k, u(k)] = p[x(k)|I^k] \triangleq p_k$$

it follows

$$p[x(k+1)|I^k, u(k)] = \int p[x(k+1)|x(k), u(k)]p_k \, dx(k) = \phi[k+1, p_k, u(k)]$$

where φ is a transformation — an operator — that maps the function p_k into another function, namely, the state prediction pdf.

Proof continues

the **functional recursion for the information state** can be written with another transformation ψ that maps p_k into p_{k+1}

$$\boxed{p_{k+1} = \psi[k+1, p_k, z(k+1), u(k)]}$$

end of proof.

The Information State

p_k is an information state according to the definition of the previous subsection, i.e., that the pdf of any future state at j > k depends only on p_k and the intervening controls

$$U_k^{j-1} \triangleq \{u(i)\}_{i=k}^{j-1}$$

$$\begin{aligned} p[x(j)|I^k, U_k^{j-1}] &= \int p[x(j)|x(k), I^k, U_k^{j-1}] p[x(k)|I^k]\, dx(k) \\ &= \int p[x(j)|x(k), U_k^{j-1}]\, p_k\, dx(k) \\ &\triangleq \mu[j, p_k, U_k^{j-1}] \end{aligned}$$

where the whiteness of the process noise sequence and its independence from the measurement noises have been used

μ denotes the transformation that maps p_k and the known inputs into the pdf of a future state x(j) it follows that I_k **is summarized by** p_k.

Therefore, the whiteness and mutual independence of the two noise sequences is a sufficient condition for p_k **to be an information state**. It should be emphasized that the whiteness is the crucial assumption

The above conditions are equivalent to the requirement that x(k) be an incompletely observed Markov process — a Markov process observed partially and/or in the presence of noise.

If, for example, the process noise sequence is not white, it is obvious that p_k does not summarize the past data. In this case the vector x is not a state anymore and it has to be augmented.

This discussion points out the reason why the formulation of stochastic estimation and control problems is done with mutually independent white noise sequences.

Example of Linear vs. Nonlinear Estimation of a Parameter

Consider an unknown parameter x with **a prior pdf** uniform in the interval [x1, x2]

$$p(x) = \mathcal{U}[x; x_1, x_2] \triangleq \frac{1(x - x_1) - 1(x - x_2)}{x_2 - x_1}$$

1(·) denotes the unit step function.

A measurement

$$z = x + w$$

where w is independent of x and uniformly distributed within the interval [−a, a]

$$p(w) = \mathcal{U}[w; -a, a] = \frac{1(w + a) - 1(w - a)}{2a}$$

The Optimal MMSE Estimator

the optimal MMSE estimator (the conditional mean)

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)}$$

the **likelihood function** of x given z — **the pdf of z conditioned on x**

$$p(z|x) = \mathcal{U}[z; x-a, x+a] = \frac{1[z-(x-a)] - 1[z-(x+a)]}{2a}$$

the conditional pdf of x given z — **a posterior pdf**

$$\begin{aligned} p(x|z) &= \frac{\{1[x-(z-a)] - 1[x-(z+a)]\}[1(x-x_1) - 1(x-x_2)]}{2a(x_2-x_1)p(z)} \\ &= \mathcal{U}[x; (z-a) \vee x_1, (z+a) \wedge x_2] \end{aligned}$$

$$a \vee b \triangleq \max(a,b) \qquad a \wedge b \triangleq \min(a,b)$$

The conditional pdf is seen to be uniform in the interval, which is the intersection of the interval [x1, x2] from the prior and the interval [z−a, z+a] — the feasibility region of x given z.

The **MMSE estimator of x given** z (i.e., its **exact conditional mean**) is

$$\hat{x}^{\mathrm{MMSE}} = E[x|z] = \frac{(z-a) \vee x_1 + (z+a) \wedge x_2}{2}$$

Note that this estimator is a **nonlinear function** of the measurement z.

The **conditional variance of this estimator** is, for a given z

$$P_{xx|z} = \mathrm{var}(x|z) = \frac{[(z-a) \vee x_1 - (z+a) \wedge x_2]^2}{12}$$

Note that the **accuracy of the optimal estimate**, measured in terms of its **variance**, is **measurement-dependent** — it depends on the measurement z.

The LMMSE Estimator

The linear MMSE estimator of x in terms of z

$$\hat{x}^{\text{LMMSE}} = \bar{x} + P_{xz} P_{zz}^{-1} (z - \bar{z})$$

$$\bar{x} = E[x] = \frac{x_1 + x_2}{2}$$

$$\bar{z} = E[z] = \bar{x}$$

$$\begin{aligned} P_{xz} &= E[(x - \bar{x})(z - \bar{z})] = E[(x - \bar{x})(x - \bar{x} + w)] \\ &= E[(x - \bar{x})^2] = P_{xx} = \frac{(x_1 - x_2)^2}{12} \end{aligned}$$

$$\begin{aligned} P_{zz} &= E[(z - \bar{z})^2] = E[(x - \bar{x} + w)^2] \\ &= P_{xx} + P_{ww} = \frac{(x_1 - x_2)^2 + 4a^2}{12} \end{aligned}$$

continues

The MSE corresponding to the estimator

$$
\begin{aligned}
P_{xx}^{\text{LMMSE}} &= P_{xx} - \frac{P_{xz}^2}{P_{zz}} = \frac{P_{xx}P_{zz} - P_{xz}^2}{P_{zz}} \\
&= \frac{P_{xx}(P_{xx} + P_{ww}) - P_{xx}^2}{P_{xx} + P_{ww}} \\
&= \frac{P_{xx}P_{ww}}{P_{xx} + P_{ww}}
\end{aligned}
$$

Comparison of the Errors: Optimal vs. Best Linear

The comparison between the errors obtained from the optimal and best linear methods cannot be made between Covariances (or MSE) since the optimal case is a function of the observations.

It has to be made between averaged over all the possible measurements z.

Comparison of the Errors: Optimal vs. Best Linear

The average covariance associated with the optimal estimate is

$$P_{xx}^{\mathrm{OPT}} = E[P_{xx|z}] = \int P_{xx|z}\, p(z)\, dz$$

the comparison of these two estimators' variances for some numerical values. The interval over which x is uniformly distributed is x2 − x1 = 1, that is, its "prior" variance was 1/12 = 0.0833. The measurement ranged from very accurate for 0 < a << 1, to very inaccurate — practically noninformative — for a >>1.

a	P_{xx}^{LMMSE}	P_{xx}^{OPT}
0.05	8.25E-04	7.64E-04
0.1	3.2E-03	2.81E-03
0.2	11.49E-03	9.5371E-03
0.5	4.16E-02	3.5871E-02
0.8	5.99E-02	5.4693E-02
1	6.66E-02	6.2510E-02
2	7.84E-02	7.6387E-02
5	8.25E-02	8.1249E-02

Conclusion of the comparison

the benefit from the nonlinear estimation over the linear one is disappointingly modest in this case:

It ranges from negligible — 1% for the inaccurate measurement considered — to 6% for the accurate measurement; its maximum, which occurs in the midrange, is about 15% in variance.

No general conclusions can be drawn from the above toy/academic example — each problem requires its own evaluation. Nevertheless, it is an indication that in some problems the **benefit from nonlinear estimation can be quite limited.**

Estimation in Nonlinear Systems with Additive Noise

the system with dynamics

$$x(k+1) = f[k, x(k)] + v(k)$$

for simplicity, it is assumed that there is no input/control, and the noise is assumed additive and white, with pdf $p_{v(k)}[v(k)]$

The measurement

$$z(k) = h[k, x(k)] + w(k)$$

measurement noise is additive, white with pdf $p_{w(k)}[w(k)]$

and independent from the process noise

The initial state has the prior pdf p[x(0)] and is assumed to be independent from the two noise sequences

Prediction

$$p_k \triangleq p[x(k)|Z^k] \qquad \text{a grid with } n_k \text{ points } \{x(k)_j\}_{j=1}^{n_k}$$

the predicted pdf follows

$$p[x(k+1)|Z^k] = \int p[x(k+1)|x(k)]p[x(k)|Z^k]\,dx(k)$$

In view of the additivity of the process noise

$$p[x(k+1)|x(k)] = p_{v(k)}\{x(k+1) - f[k, x(k)]\}$$

$$p[x(k+1)|Z^k] = \int p_{v(k)}\{x(k+1) - f[k, x(k)]\}p[x(k)|Z^k]\,dx(k)$$

grid points for the state at $k+1$ are $\{x(k+1)_i\}_{i=1}^{n_{k+1}}$

$$p[x(k+1)_i|Z^k] = \sum_{j=1}^{n_k} \mu_j p_{v(k)}\{x(k+1)_i - f[k, x(k)_j]\}p[x(k)_j|Z^k]$$

where μ_j are the weighting coefficients of the numerical integration

Update

$$p[x(k+1)|Z^{k+1}] = \frac{1}{c} p[z(k+1)|x(k+1)]p[x(k+1)|Z^k]$$

The numerical implementation

$$p[x(k+1)_i|Z^{k+1}] = \frac{1}{c} p[z(k+1)|x(k+1)_i]p[x(k+1)_i|Z^k]$$

normalization constant

$$c = \int p[z(k+1)|x(k+1)]p[x(k+1)|Z^k]\, dx(k+1)$$

numerically

$$c = \sum_{i=1}^{n_{k+1}} \mu_i p[z(k+1)|x(k+1)_i]p[x(k+1)_i|Z^k]$$

Remarks

The number of grid points for a state of dimension n_x, with m points per state component, is

$$m^{n_x}$$

This **curse of dimensionality is the major stumbling block** in making numerical techniques practical due to the ensuing heavy storage and computational requirements.

Another issue is the selection of **the grid, which evolves in time**. The points have to cover a region in the state space outside which the probability mass is negligible.

Optimal Nonlinear Estimation — Summary

Given a system described by

1. **dynamic equation** perturbed by **white process noise** and
2. **measurement equation** perturbed by **white noise independent of the process noise**,

with at least **one of these equations nonlinear**, the **estimation of** the system's **state** consists then of the **calculation of its pdf conditioned on the entire available information**: the observations, the initial state information and the past inputs.

This conditional pdf has the property of being an information state, that is, it **summarizes probabilistically the past** of the system.
The **optimal nonlinear estimator** for such a discrete-time stochastic dynamic system consists of a **recursive functional relationship for the state's conditional pdf**. This conditional pdf then yields the **MMSE estimator for the state** — the **conditional mean of the state**.

This nonlinear estimator has to be used to obtain the conditional mean of the state unless the system is linear Gaussian: Both the dynamic and the measurement equations are linear and all the noises and the initial state are Gaussian.

If the system is **linear Gaussian**, this **functional recursion becomes** the recursion for the conditional mean and covariance, the **Kalman filter**.

Unlike the linear case, in the **nonlinear case**, the **accuracy of the estimator is measurement-dependent**. The numerical implementation of the nonlinear estimator on a **set of grid points in the state space** can be very **demanding computationally** — it suffers from the **curse of dimensionality**: The memory and computational requirements are exponential in the dimension of the state

THE EXTENDED KALMAN FILTER

- Very limited feasibility of the implementation of the optimal filter, the functional recursion, **suboptimal algorithms** are of interest.
- The recursive calculation of the sufficient statistic consisting of the **conditional mean and variance in the linear-Gaussian case** is the **simplest possible state estimation** filter.
- As indicated earlier, in the case of a **linear system with non-Gaussian random variables the same simple recursion yields an approximate mean and variance**.
- A **framework similar is desirable for a nonlinear system**. Such an estimator, called the **extended Kalman filter (EKF),** can be obtained by a series expansion of the nonlinear dynamics and of the measurement equations.

The (first-order) EKF is based on

- **Linearization** — first order series expansion — of the nonlinearities (in the dynamic and/or the measurement equation)
- **LMMSE estimation**

The second-order EKF relies on a second-order expansion, that is, it includes second-order correction terms.

Modeling Assumptions

for simplicity, it is assumed that there is no control, and the noise is assumed **additive, zero mean**, and **white**

$$\boxed{x(k+1) = f[k, x(k)] + v(k)}$$

$$E[v(k)] = 0$$

$$E[v(k)v(j)'] = Q(k)\delta_{kj}$$

the measurement noise is **additive, zero mean,** and **white**

$$\boxed{z(k) = h[k, x(k)] + w(k)}$$

$$E[w(k)] = 0$$

$$E[w(k)w(j)'] = R(k)\delta_{kj}$$

$\hat{x}(0|0)$ $P(0|0)$ The initial state is assumed to be **uncorrelated** with the two noise sequences.

The estimate at time k, an ***approximate conditional mean***

$$\hat{x}(k|k) \approx E[x(k)|Z^k]$$

and **the *associated covariance matrix*** $P(k|k)$

Strictly speaking, *P(k|k)* is the MSE matrix rather than the covariance matrix in view of the fact that the estimate is not the exact conditional mean.

Above implies that the estimation error is approximately zero mean.

Another assumption that will be made is that the third-order moments of the estimation error are approximately zero — this is exact in the case of zero-mean Gaussian random variables.

Derivation of the EKF, State Prediction

The vector Taylor series expansion,

$$\begin{aligned} x(k+1) &= f[k, \hat{x}(k|k)] + f_x(k)[x(k) - \hat{x}(k|k)] \\ &\quad + \frac{1}{2}\sum_{i=1}^{n_x} e_i [x(k) - \hat{x}(k|k)]' f_{xx}^i(k)[x(k) - \hat{x}(k|k)] \end{aligned}$$

$$f_x(k) \triangleq [\nabla_x f(k, x)']'|_{x=\hat{x}(k|k)} \triangleq \frac{\partial f}{\partial x}$$

+HOT

the Jacobian of the vector f, evaluated at the latest estimate of the state.

e_i is the ith n_x-dimensional Cartesian basis vector (ith component unity, the rest zero)

$$f_{xx}^i(k) \triangleq [\nabla_x \nabla_x' f^i(k, x)]|_{x=\hat{x}(k|k)} \triangleq \frac{\partial^2 f^i}{\partial x^2}$$

The Hessian of the ith component of f and HOT represents the higher-order terms

Note that, given the data Z_k, both the above Jacobian and the Hessian are deterministic quantities—only x(k) and v(k) are random variables

The **predicted state** to time k +1 from time k is obtained by taking the expectation conditioned on Z^k and neglecting the HOT

$$\hat{x}(k+1|k) = f[k, \hat{x}(k|k)] + \frac{1}{2}\sum_{i=1}^{n_x} e_i \, \mathrm{tr}[f_{xx}^i(k)P(k|k)]$$

The first order term is (approximately) zero mean and thus vanishes

The state prediction error

$$\tilde{x}(k+1|k) = f_x(k)\tilde{x}(k|k) + \frac{1}{2}\sum_{i=1}^{n_x} e_i\{\tilde{x}(k|k)' f_{xx}^i(k)\tilde{x}(k|k) - \mathrm{tr}[f_{xx}^i(k)P(k|k)]\} + v(k)$$

the HOT have already been dropped

Multiplying the above with its transpose and taking the expectation conditioned on Z_k yields **the state prediction covariance** (actually, the MSE matrix)

$$P(k+1|k) = f_x(k)P(k|k)f_x(k)' + \frac{1}{2}\sum_{i=1}^{n_x}\sum_{j=1}^{n_x} e_i e_j' \operatorname{tr}[f_{xx}^i(k)P(k|k)f_{xx}^j(k)P(k|k)] + Q(k)$$

The state prediction includes the **second-order "bias correction" term**.

This is dropped in the first order EKF.

The prediction covariance contains a **fourth-order "bias correction" term** obtained from the squaring of the second order term.

The first-order version is the same as in the linear filter — the Jacobian $f_x(k)$ plays the role of the transition matrix F(k).

Derivation of the EKF, Measurement Update

Similarly, the **predicted measurement** is, for the second-order filter

$$\hat{z}(k+1|k) = h[k+1, \hat{x}(k+1|k)] + \frac{1}{2}\sum_{i=1}^{n_z} e_i \, \mathrm{tr}[h^i_{xx}(k+1)P(k+1|k)]$$

e_i is the ith n_z-dimensional ith Cartesian basis vector

The **measurement prediction covariance** or **innovation covariance** or **residual covariance** — really MSE matrix —

$$S(k+1) = h_x(k+1)P(k+1|k)h_x(k+1)' + \frac{1}{2}\sum_{i=1}^{n_z}\sum_{j=1}^{n_z} e_i e_j' \mathrm{tr}[h^i_{xx}(k+1)P(k+1|k)h^j_{xx}(k+1)P(k+1|k)] + R(k+1)$$

the Jacobian of h $\quad h_x(k+1) \triangleq [\nabla_x h(k+1,x)']'\Big|_{x=\hat{x}(k+1|k)} \triangleq \frac{\partial h}{\partial x}$

the Hessian of its ith component

$$h^i_{xx}(k+1) \triangleq [\nabla_x \nabla'_x h^i(k+1,x)]\Big|_{x=\hat{x}(k+1|k)} \triangleq \frac{\partial^2 h^i}{\partial x^2}$$

Modifications for the First-Order EKF

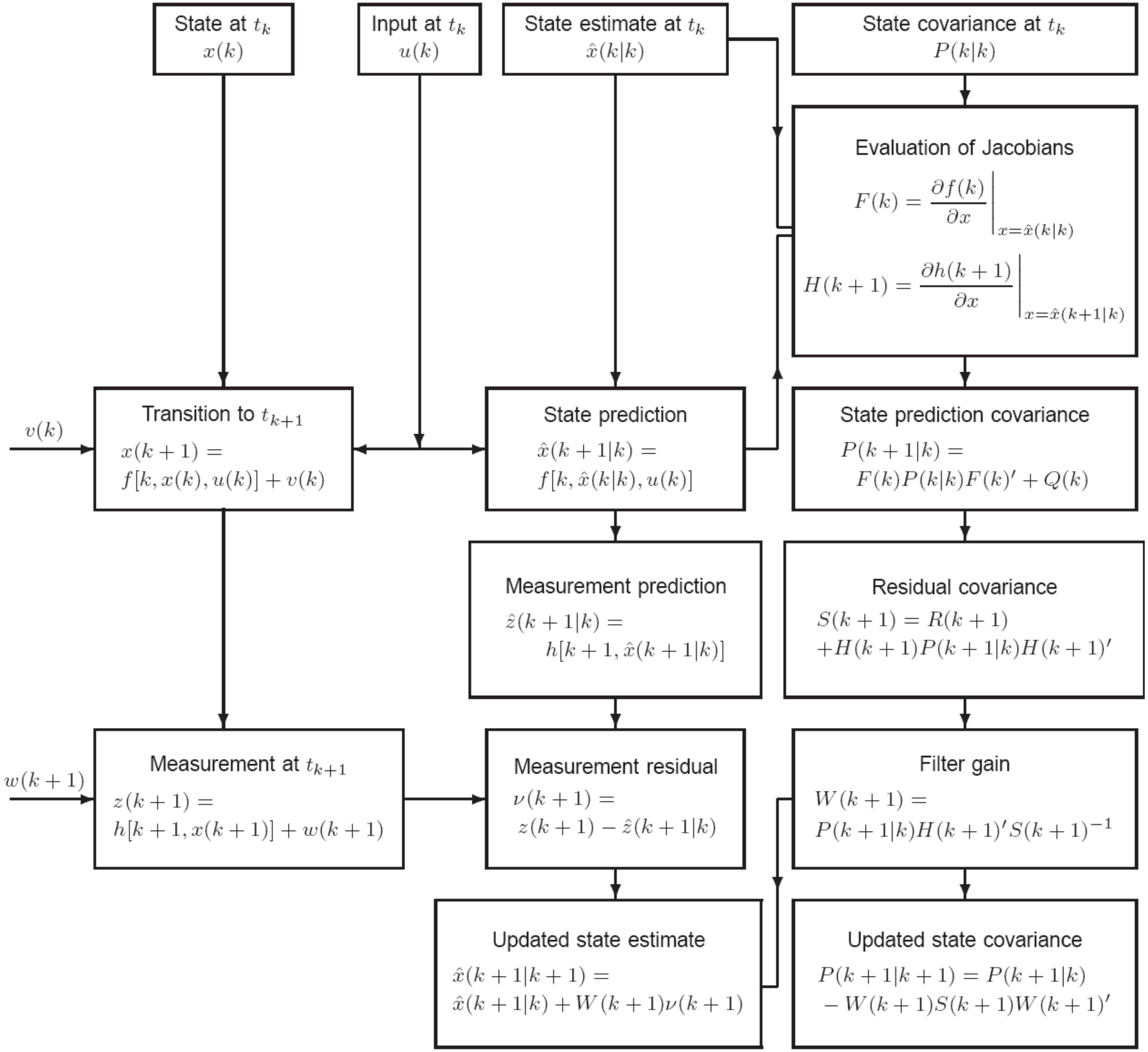
The second order “correction” term in and the corresponding fourth order term are dropped in the first-order EKF.

The first-order version s the same as in the linear filter — the Jacobian $h_x(k)$ plays the role of the measurement matrix H(k).

State Update

State Update the expression of the filter gain, the update equation for the state and its covariance are identical to those from the linear filter

Overview of the EKF Algorithm



Overview of the EKF Algorithm

- The **main difference** from the Kalman Filter is the **evaluation of the Jacobians** of the state transition and the measurement equations.
- Due to this, the **covariance computations** are not decoupled anymore from the state estimate calculations and **cannot be done offline**.
- The **linearization** (evaluation of the Jacobians) can be done, as indicated here, at the **latest state estimate for F** and **the predicted state for H**.
- Alternatively, it can be done along a nominal trajectory — a deterministic precomputed
- trajectory based on a certain scenario — which allows offline computation of the gain and covariance sequence.

A Cautionary Note about the EKF

The use of the **series expansion** in the state prediction and/or in the measurement prediction has **the potential** of introducing unmodeled errors that **violate some basic assumptions** about the prediction errors. :

1. The prediction errors are zero mean (unbiased).
2. The prediction errors have covariances equal to the ones computed by the algorithm.

In general, a nonlinear transformation will introduce **a bias and the covariance calculation based on a series expansion is not always accurate**.

There is **no guarantee that even the second-order terms can compensate for such errors**. Also, the fact that these expansions rely on Jacobians (and Hessians in the second-order case) that are evaluated at the estimated or predicted state rather than the exact state (which is unavailable) can cause errors.