Lecture 5: Recursive Nonlinear Non-Gaussian Estimation, Part I

AER1513: State Estimation

Timothy D. Barfoot

University of Toronto

Outline

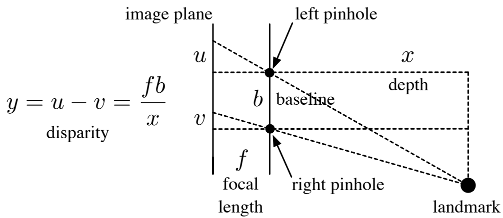
UNIVERSITY OF TORONTO

Motivation

- in the linear-Gaussian lectures, we discussed two main perspectives to estimation: full Bayesian and maximum a posteriori
- we saw that for linear motion and observation models driven by Gaussian noise, these two paradigms come to the same answer (i.e., the MAP point was the mean of the full Bayesian approach)
- this is because the full posterior is exactly Gaussian and therefore the mean and mode (i.e., maximum) are the same point
- this is not true once we move to nonlinear models, since the full Bayesian posterior is no longer Gaussian
- we'll look at the implications of this, starting with a simple intuitive example

UNIVERSITY OF TORONTO

Example: estimating landmark position



– a simple model of a stereo camera is

$$y = \frac{fb}{x} + n \tag{1}$$

- x, is the position of a landmark (in metres)
- y, is the disparity measurement (in pixels)
- f is the focal length (in pixels)
- b is the camera separation (in metres)
- n is the measurement noise (in pixels)

UNIVERSITY OF TORONTO

Example: estimating landmark position

– to perform Bayesian inference,

$$p(x|y) = \frac{p(y|x)p(x)}{\int_{-\infty}^{\infty} p(y|x)p(x)\,dx} \tag{2}$$

we require expressions for $p(y|x)$ and $p(x)$

– we assume that the measurement noise is zero-mean Gaussian, $n \sim \mathcal{N}(0, R)$, such that

$$p(y|x) = \mathcal{N}\left(\frac{fb}{x}, R\right) = \frac{1}{\sqrt{2\pi R}} \exp\left(-\frac{1}{2R}\left(y - \frac{fb}{x}\right)^2\right) \tag{3}$$

– we assume that the prior is Gaussian, where

$$p(x) = \mathcal{N}\left(\check{x}, \check{P}\right) = \frac{1}{\sqrt{2\pi\check{P}}} \exp\left(-\frac{1}{2\check{P}}\left(x - \check{x}\right)^2\right) \tag{4}$$

UNIVERSITY OF TORONTO

Comment on Bayesian approach

– the Bayesian framework provides an implied order of operations that we would like to make explicit:

assign prior $\rightarrow$ draw x_{true} $\rightarrow$ draw y_{meas} $\rightarrow$ compute posterior

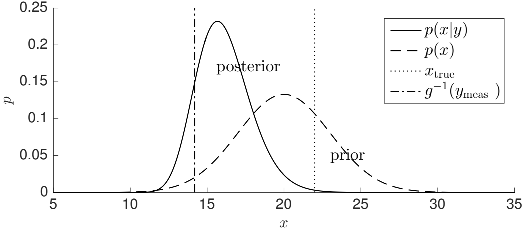
– the true state, x_{true} is drawn from the prior

– the measurement, y_{meas}, is generated by observing the true state through the camera model and adding noise

– the estimator then reconstructs the posterior from the measurement and prior, without knowing x_{true} (it's hidden)

– adherence to this process is necessary to ensure 'fair' comparison between state estimation algorithms

Example: estimating landmark position, using full Bayes



– picking the parameter values,

$$\check{x} = 20 \text{ [m]}, \quad \check{P} = 9 \text{ [m}^2\text{]}, \tag{5}$$

$$f = 400 \text{ [pixel]}, \quad b = 0.1 \text{ [m]}, \quad R = 0.09 \text{ [pixel}^2\text{]}$$

then the full Bayesian posterior (non-Gaussian, computed numerically) is shown above for one random draw with

$$x_{\text{true}} = 22 \text{ [m]}, \quad y_{\text{meas}} = \frac{fb}{x_{\text{true}}} + 1 \text{ [pixel]}$$

UNIVERSITY OF TORONTO

Example: estimating landmark position, using MAP

- in the MAP approach, we seek the state that maximizes the posterior:

$$\hat{x}_{\text{map}} = \arg\max_x \, p(x|y) \tag{6}$$

- equivalently, we can minimize the negative log likelihood:

$$\hat{x}_{\text{map}} = \arg\min_x \left(-\ln(p(x|y))\right) \tag{7}$$

  which can be easier when the PDFs involved are from the exponential family

- as we are seeking only the most likely state, we can use Bayes' rule to write

$$\hat{x}_{\text{map}} = \arg\min_x \left(-\ln(p(y|x)) - \ln(p(x))\right) \tag{8}$$

  where we drop $p(y)$ since it does not depend on x

Example: estimating landmark position, using MAP

- relating this back to the stereo camera example presented earlier, we can write

$$\hat{x}_{\text{map}} = \arg\min_{x} J(x) \tag{9}$$

  with

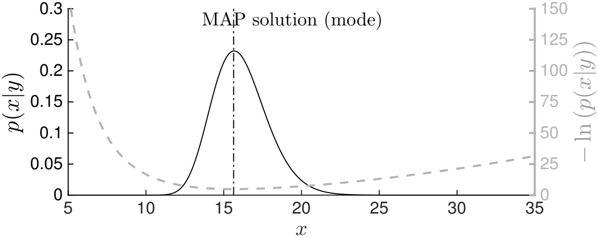
$$J(x) = \frac{1}{2R}\left(y - \frac{fb}{x}\right)^2 + \frac{1}{2\check{P}}\left(\check{x} - x\right)^2 \tag{10}$$

  where we have dropped any further normalization constants that do not depend on x

- we can then find $\hat{x}_{\text{map}}$ using any number of numerical optimization techniques

UNIVERSITY OF TORONTO

Example: estimating landmark position, using MAP



- the figure shows the posterior from the stereo camera example, $p(x|y)$, as well as the negative log-likelihood of the posterior, $-\ln(p(x|y)$
- we see that the MAP solution is simply the value of x that maximizes (or minimizes) the posterior (or its neg. log-likelihood)
- in other words, the MAP solution is the mode of the posterior, which is not generally the same as the mean

UNIVERSITY OF TORONTO

Are we lost with MAP?

- we often report the average performance of our estimators, $\hat{x}$, with respect to ground truth:

$$e_{\text{mean}}(\hat{x}) = E_{XN}[\hat{x} - x_{\text{true}}] \tag{11}$$

  where $E_{XN}[\cdot]$ is the expectation operator we are averaging over both the random draw of x_{true} from the prior, as well as the random draw of n from the measurement noise

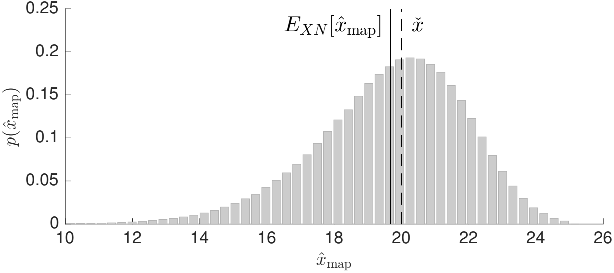
- since x_{true} is assumed to be independent of n, we have $E_{XN}[x_{\text{true}}] = E_X[x_{\text{true}}] = \check{x}$, and so

$$e_{\text{mean}}(\hat{x}) = E_{XN}[\hat{x}] - \check{x} \tag{12}$$

- you may be surprised to learn that under this performance measure, MAP estimation can be biased (i.e., $e_{\text{mean}}(\hat{x}_{\text{map}}) \neq 0$), for nonlinear measurement models

UNIVERSITY OF TORONTO

MAP is biased



- the figure shows the performance of our MAP estimator over 1,000,000 random trials of the stereo landmark estimator problem
- we see that $e_{\mathrm{mean}} \approx -33.0$ cm, which indicates a bias
- things get even worse in practice because here we've assumed we know the true prior from which the landmark position was drawn

UNIVERSITY OF TORONTO

Summary so far

- we've just seen that even in the simplest of estimation problems, the MAP approach does not find the mean of the full Bayesian posterior for nonlinear problems
- computing the full Bayesian posterior was easy in our 1D example, but in general it's intractable and we will need to make some approximations
- it will be important to understand the implications of the approximations we choose

For a non-Gaussian posterior, the mean and the mode are typically not the same point.

System

- we define our system using the following nonlinear, time-varying models:

$$
\begin{aligned}
\text{motion model:} \quad \mathbf{x}_k &= \mathbf{f}\left(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k\right), \qquad k = 1 \ldots K && \text{(13a)} \\
\text{observation model:} \quad \mathbf{y}_k &= \mathbf{g}\left(\mathbf{x}_k, \mathbf{n}_k\right), \qquad k = 0 \ldots K && \text{(13b)}
\end{aligned}
$$

  where k is again the discrete-time index and K its maximum
- the variables have the following meanings:

$$
\begin{aligned}
\text{system state}: &\quad \mathbf{x}_k \in \mathbb{R}^N \\
\text{initial state}: &\quad \mathbf{x}_0 \in \mathbb{R}^N \sim \mathcal{N}\left(\check{\mathbf{x}}_0, \check{\mathbf{P}}_0\right) \\
\text{input}: &\quad \mathbf{v}_k \in \mathbb{R}^N \\
\text{process noise}: &\quad \mathbf{w}_k \in \mathbb{R}^N \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_k\right) \\
\text{measurement}: &\quad \mathbf{y}_k \in \mathbb{R}^M \\
\text{measurement noise}: &\quad \mathbf{n}_k \in \mathbb{R}^M \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_k\right)
\end{aligned}
$$

Problem statement

– our state estimation problem is still the following:

Definition

The problem of state estimation is to come up with an estimate, $\hat{\mathbf{x}}_k$, of the true *state* of a system, at one or more timesteps, k, given knowledge of the initial state, $\check{\mathbf{x}}_0$, a sequence of measurements, $\mathbf{y}_{0:K,\text{meas}}$, a sequence of inputs, $\mathbf{v}_{1:K}$, as well as knowledge of the system's motion and observation models.

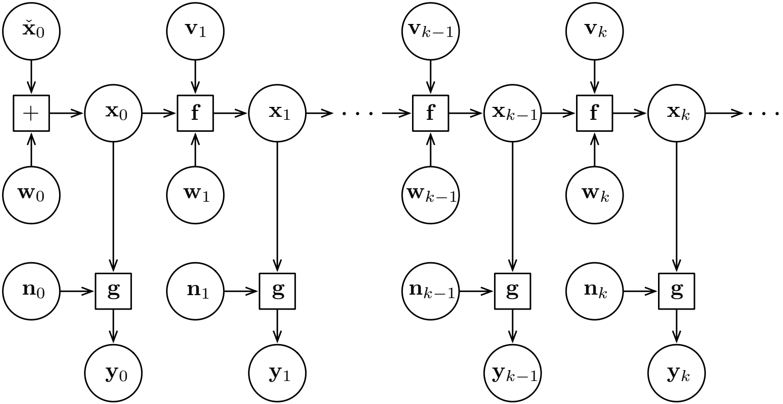
Markov property

A stochastic process has the Markov property if the conditional PDF of future states of the process, given the present state, depend only upon the present state, but not on any other past states, i.e., they are conditionally independent of these older states. Such a process is called Markovian or a Markov process.

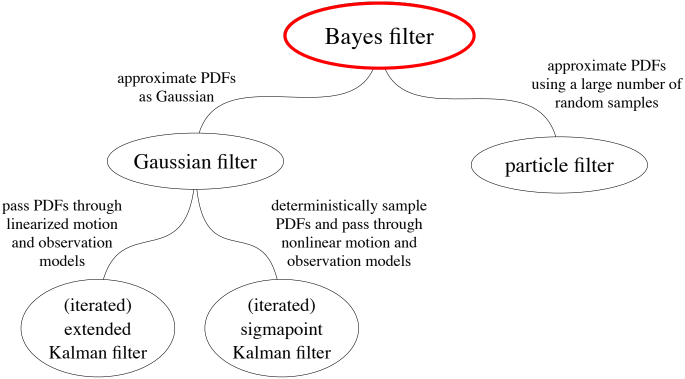
- our system is such a Markov process; once we know the value of $\mathbf{x}_{k-1}$, we do not need to know the value of any previous states in order to evolve the system forwards in time to compute $\mathbf{x}_k$
- this property was exploited fully in the section on linear-Gaussian estimation
- but what about nonlinear, non-Gaussian systems? Can we still have a recursive solution?

UNIVERSITY OF TORONTO

Markov property



Filter taxonomy



UNIVERSITY OF TORONTO

Bayes filter

- the Bayes filter seeks to come up with a entire PDF to represent the likelihood of the state, $\mathbf{x}_k$, using only measurements up to and including the current time
- using our notation from before, we want to compute

$$p(\mathbf{x}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k}) \tag{14}$$

- this is also sometimes called the belief function for $\mathbf{x}_k$
- we will not make any assumptions about the nature of the models or PDFs involved in our system (for now)

Bayes filter: step 1

- by employing the statistical independence of the measurements (given the state), we may factor out the latest measurement to have

$$p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k}) = \eta \, p(\mathbf{y}_k|\mathbf{x}_k) \, p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) \qquad (15)$$

  where we have employed Bayes' rule to reverse the dependence and η serves to preserve the axiom of total probability
- we'll turn our attention to the second factor next

UNIVERSITY OF TORONTO

Bayes filter: step 2

– turning our attention to the second factor, we introduce the hidden state, $\mathbf{x}_{k-1}$, and integrate over all possible values:

$$p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) = \int p(\mathbf{x}_k, \mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1})\, d\mathbf{x}_{k-1}$$
$$= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1})\, p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1})\, d\mathbf{x}_{k-1} \tag{16}$$

– the introduction of the hidden state can be viewed as the opposite of marginalization – this is a common trick in Bayesian inference

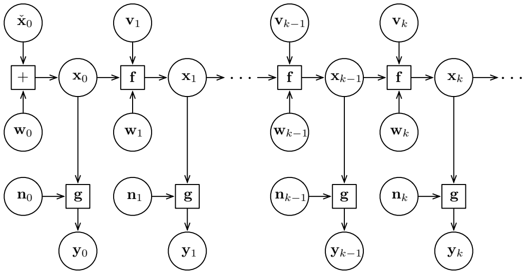
UNIVERSITY OF TORONTO

Bayes filter: step 3

- the next step is subtle and is the cause of many limitations in recursive estimation
- since our system enjoys the Markov property, we use said property (on the estimator) to say that

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{v}_k) \tag{17a}$$

$$p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) = p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1}) \tag{17b}$$



UNIVERSITY OF TORONTO

Bayes filter: step 4

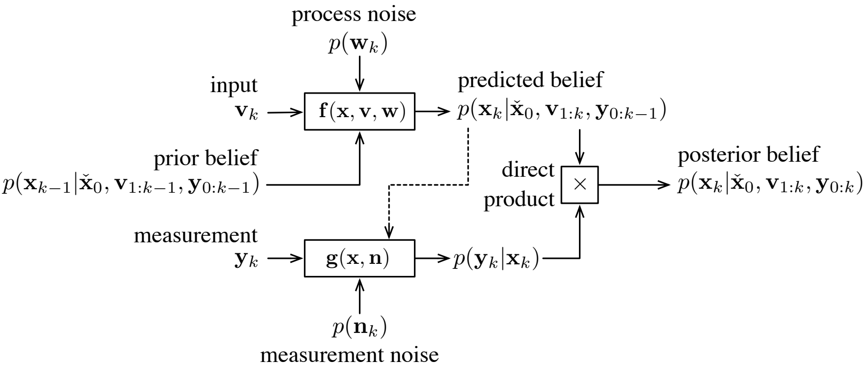
- substituting the step 3 simplifications into step 2 and that into step 1 we have the Bayes filter:

$$\underbrace{p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k})}_{\text{posterior belief}}$$
$$= \eta \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\substack{\text{observation} \\ \text{correction} \\ \text{using } \mathbf{g}(\cdot)}} \int \underbrace{p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{v}_k)}_{\substack{\text{motion} \\ \text{prediction} \\ \text{using } \mathbf{f}(\cdot)}} \underbrace{p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1})}_{\text{prior belief}} \, d\mathbf{x}_{k-1} \tag{18}$$

- we can see this takes on a predictor-corrector form, just like the KF

UNIVERSITY OF TORONTO

Bayes filter



Bayes filter commentary

– although exact, the Bayes filter is really nothing more than a mathematical artifact; it can never be implemented in practice, except for the linear-Gaussian case

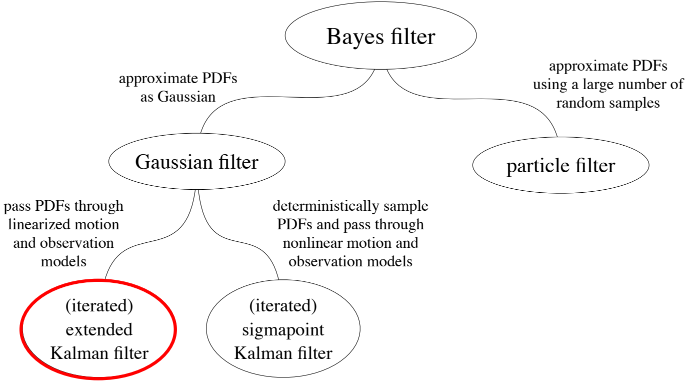
– there are two primary reasons for this, and as such we need to make appropriate approximations:

  (i) PDFs live in an infinite-dimensional space (as do all continuous functions) and as such an infinite amount of memory (i.e., infinite number of parameters) would be needed to completely represent the belief, $p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k})$

  (ii) the integral in the Bayes filter is computationally very expensive; it would require infinite computing resources to evaluate exactly

UNIVERSITY OF TORONTO

Filter taxonomy



UNIVERSITY OF TORONTO

Extended Kalman filter

- if the belief is constrained to be Gaussian, the noise Gaussian, and we linearize the motion and observation models in order to carry out the integral (and also the direct product) in the Bayes filter, we arrive at the famous extended Kalman filter (EKF)
- the EKF is called extended because it is the extension of the Kalman filter to nonlinear systems
- the EKF is still the mainstay of estimation and data fusion in many circles, and can often be effective for mildly nonlinear, non-Gaussian systems

EKF: approximation 1

– we first constrain our belief function for $\mathbf{x}_k$ to be Gaussian:

$$p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k\right) \tag{19}$$

where $\hat{\mathbf{x}}_k$ is the mean and $\hat{\mathbf{P}}_k$ the covariance

– next, we assume that the noise variables, $\mathbf{w}_k$ and $\mathbf{n}_k$ $(\forall k)$, are Gaussian as well:

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \tag{20a}$$

$$\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \tag{20b}$$

UNIVERSITY OF TORONTO

EKF: approximation 2

– next, we linearize the motion and observation models about the current state estimate mean:

$$\mathbf{f}\left(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k\right) \approx \check{\mathbf{x}}_k + \mathbf{F}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \mathbf{w}'_k \qquad (21a)$$

$$\mathbf{g}\left(\mathbf{x}_k, \mathbf{n}_k\right) \approx \check{\mathbf{y}}_k + \mathbf{G}_k\left(\mathbf{x}_k - \check{\mathbf{x}}_k\right) + \mathbf{n}'_k \qquad (21b)$$

where

$$\check{\mathbf{x}}_k = \mathbf{f}\left(\hat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \mathbf{0}\right), \quad \mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{x}_{k-1}}\right|_{\hat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \mathbf{0}}, \qquad (22a)$$

$$\mathbf{w}'_k = \left.\frac{\partial \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k}\right|_{\hat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \mathbf{0}} \mathbf{w}_k \qquad (22b)$$

and

$$\check{\mathbf{y}}_k = \mathbf{g}\left(\check{\mathbf{x}}_k, \mathbf{0}\right), \quad \mathbf{G}_k = \left.\frac{\partial \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k)}{\partial \mathbf{x}_k}\right|_{\check{\mathbf{x}}_k, \mathbf{0}}, \qquad (23a)$$

$$\mathbf{n}'_k = \left.\frac{\partial \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k)}{\partial \mathbf{n}_k}\right|_{\check{\mathbf{x}}_k, \mathbf{0}} \mathbf{n}_k \qquad (23b)$$

UNIVERSITY OF TORONTO

EKF: motion model analysis

– the statistical properties of the current state, $\mathbf{x}_k$, given the old state and latest input, are

$$\mathbf{x}_k \approx \check{\mathbf{x}}_k + \mathbf{F}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \mathbf{w}'_k \tag{24a}$$

$$E\left[\mathbf{x}_k\right] \approx \check{\mathbf{x}}_k + \mathbf{F}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \underbrace{E\left[\mathbf{w}'_k\right]}_{\mathbf{0}} \tag{24b}$$

$$E\left[\left(\mathbf{x}_k - E\left[\mathbf{x}_k\right]\right)\left(\mathbf{x}_k - E\left[\mathbf{x}_k\right]\right)^T\right] \approx \underbrace{E\left[\mathbf{w}'_k \mathbf{w}'^T_k\right]}_{\mathbf{Q}'_k} \tag{24c}$$

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{v}_k) \approx \mathcal{N}\left(\check{\mathbf{x}}_k + \mathbf{F}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right), \mathbf{Q}'_k\right) \tag{24d}$$

EKF: observation model analysis

– the statistical properties of the current measurement, $\mathbf{y}_k$, given the current state, are

$$\mathbf{y}_k \approx \check{\mathbf{y}}_k + \mathbf{G}_k \left(\mathbf{x}_k - \check{\mathbf{x}}_k\right) + \mathbf{n}'_k \tag{25a}$$

$$E\left[\mathbf{y}_k\right] \approx \check{\mathbf{y}}_k + \mathbf{G}_k \left(\mathbf{x}_k - \check{\mathbf{x}}_k\right) + \underbrace{E\left[\mathbf{n}'_k\right]}_{\mathbf{0}} \tag{25b}$$

$$E\left[\left(\mathbf{y}_k - E\left[\mathbf{y}_k\right]\right)\left(\mathbf{y}_k - E\left[\mathbf{y}_k\right]\right)^T\right] \approx \underbrace{E\left[\mathbf{n}'_k \mathbf{n}'^T_k\right]}_{\mathbf{R}'_k} \tag{25c}$$

$$p(\mathbf{y}_k|\mathbf{x}_k) \approx \mathcal{N}\left(\check{\mathbf{y}}_k + \mathbf{G}_k(\mathbf{x}_k - \check{\mathbf{x}}_k), \mathbf{R}'_k\right) \tag{25d}$$

UNIVERSITY OF TORONTO

EKF

- substituting in the motion and observation approximations, Bayes filter becomes

$$\underbrace{p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k})}_{\mathcal{N}(\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k)} = \eta \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\mathcal{N}(\check{\mathbf{y}}_k + \mathbf{G}_k(\mathbf{x}_k - \check{\mathbf{x}}_k), \mathbf{R}'_k)}$$
$$\times \int \underbrace{p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{v}_k)}_{\mathcal{N}(\check{\mathbf{x}}_k + \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}), \mathbf{Q}'_k)} \underbrace{p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1})}_{\mathcal{N}(\hat{\mathbf{x}}_{k-1}, \hat{\mathbf{P}}_{k-1})} d\mathbf{x}_{k-1} \tag{26}$$

UNIVERSITY OF TORONTO

EKF: predictor

– Using our identity for passing a Gaussian through a (stochastic) nonlinearity, we can see that the integral is also Gaussian:

$$\underbrace{p(\mathbf{x}_k|,\check{\mathbf{x}}_0,\mathbf{v}_{1:k},\mathbf{y}_{0:k})}_{\mathcal{N}(\hat{\mathbf{x}}_k,\hat{\mathbf{P}}_k)} = \eta \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\mathcal{N}(\check{\mathbf{y}}_k+\mathbf{G}_k(\mathbf{x}_k-\check{\mathbf{x}}_k),\mathbf{R}'_k)} \times \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{v}_k)\, p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0,\mathbf{v}_{1:k-1},\mathbf{y}_{0:k-1})\, d\mathbf{x}_{k-1}}_{\mathcal{N}(\check{\mathbf{x}}_k,\mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^T+\mathbf{Q}'_k)} \tag{27}$$

EKF: corrector

– we are faced with taking the direct product of two Gaussian PDFs, which we studied in the probability lecture:

$$\underbrace{p(\mathbf{x}_k|\check{\mathbf{x}}_0, , \mathbf{v}_{1:k}, \mathbf{y}_{0:k})}_{\mathcal{N}(\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k)}$$
$$= \underbrace{\eta\, p(\mathbf{y}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{v}_k)\, p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1})\, d\mathbf{x}_{k-1}}_{\mathcal{N}\left(\check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k), (\mathbf{1} - \mathbf{K}_k\mathbf{G}_k)(\mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}'_k)\right)} \tag{28}$$

where $\mathbf{K}_k$ is known as the Kalman gain matrix:

$$\mathbf{K}_k = \check{\mathbf{P}}_k \mathbf{G}_k^T \left(\mathbf{G}_k \check{\mathbf{P}}_k \mathbf{G}_k^T + \mathbf{R}'_k\right)^{-1} \tag{29}$$

UNIVERSITY OF TORONTO

EKF

- comparing left and right sides of our posterior expression above we have

$$
\begin{aligned}
&\text{predictor:} & \check{\mathbf{P}}_k &= \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_k' && \text{(30a)} \\
& & \check{\mathbf{x}}_k &= \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \mathbf{0}) && \text{(30b)} \\
&\text{Kalman gain:} & \mathbf{K}_k &= \check{\mathbf{P}}_k\mathbf{G}_k^T\left(\mathbf{G}_k\check{\mathbf{P}}_k\mathbf{G}_k^T + \mathbf{R}_k'\right)^{-1} && \text{(30c)} \\
&\text{corrector:} & \hat{\mathbf{P}}_k &= \left(\mathbf{1} - \mathbf{K}_k\mathbf{G}_k\right)\check{\mathbf{P}}_k && \text{(30d)} \\
& & \hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k\underbrace{\left(\mathbf{y}_k - \mathbf{g}(\check{\mathbf{x}}_k, \mathbf{0})\right)}_{\text{innovation}} && \text{(30e)}
\end{aligned}
$$

which are the classic recursive update equations for the EKF

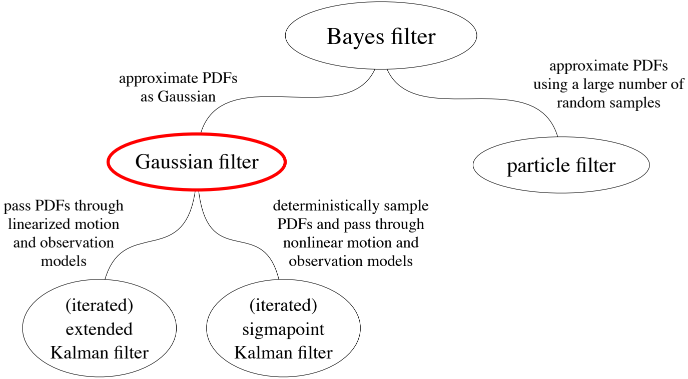
- they allow us to compute $\left\{\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k\right\}$ from $\left\{\hat{\mathbf{x}}_{k-1}, \hat{\mathbf{P}}_{k-1}\right\}$ using $\mathbf{v}_k$ and $\mathbf{y}_k$

EKF commentary

- we notice immediately similar structure to the KF for linear-Gaussian estimation
- there are two main differences:
  (i) the nonlinear motion and observation models are used to propagate the mean of our estimate
  (ii) there are Jacobians embedded in the $\mathbf{Q}_k'$ and $\mathbf{R}_k'$ covariances for the noise
- there is no proof that the EKF will work in general for a nonlinear system
- the main problem is that the operating point of the linearization is the mean of our estimate of the state, not the true state
- this can cause the EKF to diverge wildly in some cases, become biased or inconsistent

UNIVERSITY OF TORONTO

Filter taxonomy



UNIVERSITY OF TORONTO

Generalized Gaussian filter

- we can actually take a step back from the EKF and consider the class of all Gaussian filters
- we begin with a Gaussian prior at time $k-1$:

$$p(\mathbf{x}_{k-1}|\check{\mathbf{x}}_0, \mathbf{v}_{1:k-1}, \mathbf{y}_{0:k-1}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k-1}, \hat{\mathbf{P}}_{k-1}\right) \tag{31}$$

- we pass this forwards in time through the nonlinear motion model, $\mathbf{f}(\cdot)$, to propose a Gaussian prior at time k:

$$p(\mathbf{x}_k|\check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) = \mathcal{N}\left(\check{\mathbf{x}}_k, \check{\mathbf{P}}_k\right) \tag{32}$$

- this is the prediction step and incorporates the latest input, $\mathbf{v}_k$
- we'll decide how to actually do the computation later on

UNIVERSITY OF TORONTO

Generalized Gaussian filter: corrector

- for the correction step, we use the two-step inference approach and write a joint Gaussian for the state and measurement, at time k:

$$p(\mathbf{x}_k, \mathbf{y}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{x,k} \\ \boldsymbol{\mu}_{y,k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx,k} & \boldsymbol{\Sigma}_{xy,k} \\ \boldsymbol{\Sigma}_{yx,k} & \boldsymbol{\Sigma}_{yy,k} \end{bmatrix} \right) \quad (33)$$

- we then write the conditional Gaussian density for $\mathbf{x}_k$ (i.e., the posterior) directly as

$$\begin{aligned} &p(\mathbf{x}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k}) \\ &= \mathcal{N}\Big(\underbrace{\boldsymbol{\mu}_{x,k} + \boldsymbol{\Sigma}_{xy,k} \boldsymbol{\Sigma}_{yy,k}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}_{y,k})}_{\hat{\mathbf{x}}_k}, \underbrace{\boldsymbol{\Sigma}_{xx,k} - \boldsymbol{\Sigma}_{xy,k} \boldsymbol{\Sigma}_{yy,k}^{-1} \boldsymbol{\Sigma}_{yx,k}}_{\hat{\mathbf{P}}_k} \Big) \end{aligned} \quad (34)$$

  where we have defined $\hat{\mathbf{x}}_k$ as the mean and $\hat{\mathbf{P}}_k$ as the covariance

- the nonlinear observation model, $\mathbf{g}(\cdot)$, is used to compute $\boldsymbol{\mu}_{y,k}$

UNIVERSITY OF TORONTO

Generalized Gaussian filter: corrector

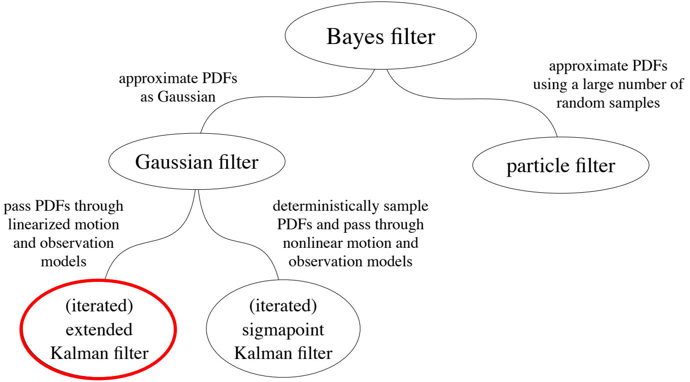
- from here, we can write the generalized Gaussian correction-step equations as

$$\begin{aligned} \mathbf{K}_k &= \mathbf{\Sigma}_{xy,k}\mathbf{\Sigma}_{yy,k}^{-1} && (35a) \\ \hat{\mathbf{P}}_k &= \check{\mathbf{P}}_k - \mathbf{K}_k\mathbf{\Sigma}_{xy,k}^T && (35b) \\ \hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \boldsymbol{\mu}_{y,k}\right) && (35c) \end{aligned}$$

  where we have let $\boldsymbol{\mu}_{x,k} = \check{\mathbf{x}}_k$, $\mathbf{\Sigma}_{xx,k} = \check{\mathbf{P}}_k$, and $\mathbf{K}_k$ is still known as the Kalman gain
- unfortunately, unless the motion and observation models are linear, we cannot compute all the remaining quantities required exactly: $\boldsymbol{\mu}_{y,k}$, $\mathbf{\Sigma}_{yy,k}$, and $\mathbf{\Sigma}_{xy,k}$
- we can make different approximations at this stage

UNIVERSITY OF TORONTO

Filter taxonomy



UNIVERSITY OF TORONTO

Iterated EKF corrector

- let's try linearization again, just for the corrector step
- our nonlinear measurement model is given by

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k) \tag{36}$$

- we linearize about an arbitrary operating point, $\mathbf{x}_{\mathrm{op},k}$:

$$\mathbf{g}\left(\mathbf{x}_k, \mathbf{n}_k\right) \approx \mathbf{y}_{\mathrm{op},k} + \mathbf{G}_k\left(\mathbf{x}_k - \mathbf{x}_{\mathrm{op},k}\right) + \mathbf{n}'_k \tag{37}$$

  where

$$\mathbf{y}_{\mathrm{op},k} = \mathbf{g}\left(\mathbf{x}_{\mathrm{op},k}, \mathbf{0}\right), \quad \mathbf{G}_k = \left.\frac{\partial \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_{\mathrm{op},k},\mathbf{0}}, \tag{38a}$$

$$\mathbf{n}'_k = \left.\frac{\partial \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k)}{\partial \mathbf{n}_k}\right|_{\mathbf{x}_{\mathrm{op},k},\mathbf{0}} \mathbf{n}_k \tag{38b}$$

- note, the observation model and Jacobians are evaluated at $\mathbf{x}_{\mathrm{op},k}$

UNIVERSITY OF TORONTO

IEKF

– using this linearized model, we can then express the joint density for the state and the measurement at time k as approximately Gaussian:

$$p(\mathbf{x}_k, \mathbf{y}_k | \check{\mathbf{x}}_0, \mathbf{v}_{1:k}, \mathbf{y}_{0:k-1}) \approx \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{x,k} \\ \boldsymbol{\mu}_{y,k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx,k} & \boldsymbol{\Sigma}_{xy,k} \\ \boldsymbol{\Sigma}_{yx,k} & \boldsymbol{\Sigma}_{yy,k} \end{bmatrix} \right)$$
$$= \mathcal{N}\left(\begin{bmatrix} \check{\mathbf{x}}_k \\ \mathbf{y}_{\mathrm{op},k} + \mathbf{G}_k(\check{\mathbf{x}}_k - \mathbf{x}_{\mathrm{op},k}) \end{bmatrix}, \begin{bmatrix} \check{\mathbf{P}}_k & \check{\mathbf{P}}_k \mathbf{G}_k^T \\ \mathbf{G}_k \check{\mathbf{P}}_k & \mathbf{G}_k \check{\mathbf{P}}_k \mathbf{G}_k^T + \mathbf{R}'_k \end{bmatrix} \right) \quad (39)$$

– this gives us all the unknown quantities that we need to substitute into the generalized Gaussian filter correction step:

$$\boldsymbol{\mu}_{y,k}, \quad \boldsymbol{\Sigma}_{yy,k}, \quad \boldsymbol{\Sigma}_{xy,k} \quad (40)$$

UNIVERSITY OF TORONTO

IEKF

- substituting in the generalized Gaussian filter correction step we have the IEKF corrector equations:

$$\begin{aligned} \mathbf{K}_k &= \check{\mathbf{P}}_k \mathbf{G}_k^T \left(\mathbf{G}_k \check{\mathbf{P}}_k \mathbf{G}_k^T + \mathbf{R}_k' \right)^{-1} && \text{(41a)} \\ \hat{\mathbf{P}}_k &= \left(\mathbf{1} - \mathbf{K}_k \mathbf{G}_k \right) \check{\mathbf{P}}_k && \text{(41b)} \\ \hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{y}_{\mathrm{op},k} - \mathbf{G}_k (\check{\mathbf{x}}_k - \mathbf{x}_{\mathrm{op},k}) \right) && \text{(41c)} \end{aligned}$$

- the operating point of the linearization differs from the EKF
- we iterate, each time setting the operating point to be the mean of the posterior at the last iteration:

$$\mathbf{x}_{\mathrm{op},k} \leftarrow \hat{\mathbf{x}}_k \qquad (42)$$

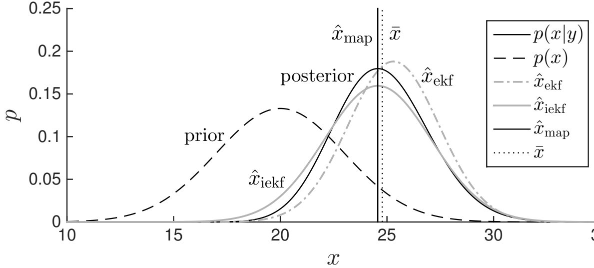
- this lets us linearize about better and better estimates
- at the first iteration we take $\mathbf{x}_{\mathrm{op},k} = \check{\mathbf{x}}_k$, same as the EKF

UNIVERSITY OF TORONTO

IEKF/EKF commentary

- a great question to ask at this point is, what is the relationship between the EKF/IEKF estimate and the full Bayesian posterior?
- it turns out that the IEKF estimate corresponds to a (local) maximum of the full posterior; in other words, it is a MAP estimate (just for the correction step, at just one timestep)
- this is pretty confusing since the 'mean' of the IEKF actually corresponds to the mode of the posterior
- on the other hand, since the EKF is not iterated, it can be very far from a local maximum; there is actually very little we can say about its relationship to the full posterior

IEKF/EKF



- going back to our stereo landmark example, we can visually verify that the IEKF does get to the maximum of the full posterior
- it's not clear what the EKF is doing; it does not correspond to the mean or the mode of the posterior
- in this version of the example, we used $x_{\text{true}} = 26$ [m] and $y_{\text{meas}} = \frac{fb}{x_{\text{true}}} - 0.6$ [pixel] to exaggerate the difference between the methods

UNIVERSITY OF TORONTO

Summary

- we began by investigating nonlinear recursive estimation techniques
- we started by deriving the very general Bayes filter, but argued that it was not possible to implement it exactly except in the linear-Gaussian case
- we then derived the famous EKF, which assumes the PDFs involved are Gaussian and that we can linearize the motion/observation models about the current state estimate
- we then derived a more general Gaussian filter and showed how to use linearization to arrive at the Iterated EKF, which improves on the corrector step of the one-shot EKF
- finally, we showed that the IEKF is really an MAP estimator (its mean is really the mode of the posterior!), while the EKF is not obviously related to the posterior
- next time we'll look at other recursive estimation techniques that make different approximations than linearization