ELE8088: Control & Estimation Theory **QUB, 2021**

Handout 15: Beyond the Kalman Filter

Lecturer: Pantelis Sopasakis *Date:* ______

Topics: Nonlinearities ◦ Extended Kalman Filter (EKF) ◦ Moving Horizon Estimation (MHE).

Last update: January 31, 2022 at 02:01:46.

15.1 Randomness and Nonlinearities

Given the pdf (or pmf) of a random variable X, we can determine its expectation and variance. However, the expectation and variance of functions of X, i.e., random variables $Y = f(X)$, are typically difficult to determine. Here we will present two approximation methods: linearisation and Monte Carlo simulations.

15.1.1 Linearisation

Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$, which can be written as follows

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \tag{15.1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are smooth functions. The **Jacobian matrix** of f is a function $Jf : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ given by

$$Jf(x) = \begin{bmatrix} \nabla f_1(x)^\intercal \\ \nabla f_2(x)^\intercal \\ \vdots \\ \nabla f_m(x)^\intercal \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{15.2}$$

Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable function. Let X be a $\mathbb{R}^n$-valued random variable. Define $Y = f(X)$ and take $\overline{X} := \mathbb{E}[X]$, $\Sigma_X := \mathrm{Var}[X]$. Then,

$$Y = f(X) \approx f(\overline{X}) + Jf(\overline{X})(X - \overline{X}), \tag{15.3}$$

so the expectation and variance of Y can be approximated by

$$\mathbb{E}[Y] \approx \overline{Y}_{\mathrm{lin}} := f(\overline{X}) \tag{15.4a}$$

$$\mathrm{Var}[Y] \approx \Sigma_{Y,\mathrm{lin}} := Jf(\overline{X})\Sigma_X Jf(\overline{X})^\intercal. \tag{15.4b}$$

Roughly speaking, this approximation is good when f is "not too nonlinear"[1] and Σ_X is not too large.

15.1.2 Monte Carlo approximation

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and X be a $\mathbb{R}^n$-valued random variable. Define $Y = f(X)$. The expectation and variance of Y can be approximated by

$$\overline{Y}_{\text{mc}} = \tfrac{1}{N} \sum_{i=1}^{N} f(x^i), \tag{15.5a}$$

$$\Sigma_{Y,\text{mc}} = \tfrac{1}{N} \sum_{i=1}^{N} (f(x^i) - \overline{Y}_{\text{mc}})(f(x^i) - \overline{Y}_{\text{mc}})^\intercal, \tag{15.5b}$$

respectively, where $(x^i)_{i=1}^N$ are independent samples from the distribution of X. Although for any finite N, the above estimators of the expectation and the variance may be biased, they (almost surely) converge to the true values as $N \to \infty$[2].

15.1.3 Comparison

Consider the function $f(x) = e^x$ and $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, the linearisation-based estimation of the expected value and variance of $Y = f(X)$ are

$$\overline{Y}_{\text{lin}} = e^{\mu}, \tag{15.6a}$$

$$\Sigma_{Y,\text{lin}} = e^{2\mu}\sigma^2. \tag{15.6b}$$

It is known that[3]

$$\mathbb{E}[Y] = e^{\mu+\sigma^2/2}, \tag{15.7a}$$

$$\text{Var}[Y] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}. \tag{15.7b}$$

[1] in the sense that the higher-order terms of Taylor's theorem that we omitted can be ignored.

[2] provided f is measurable and has a finite variance; this is a consequence of Kolmogorov's strong law of large numbers.

[3] **Exercise 1 (☕☕).** Use LotUS to derive these results.

Note: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = e^X$ follows the *log-normal* distribution with parameters μ and σ, which has a known pdf, mean and variance.

Let $X \sim \mathcal{N}(0.5, 0.01)$; then the expectation and variance of $Y = e^X$ are

	Correct	Monte Carlo			Linearisation
		$N = 10^2$	$N = 10^4$	$N = 10^6$	
Expectation	1.6570	1.6613	1.6564	1.6569	1.6487
Variance	0.0276	0.0330	0.0280	0.0276	0.0272

and for $X \sim \mathcal{N}(0.5, 0.15)$

	Correct	Monte Carlo			Linearisation
		$N = 10^2$	$N = 10^4$	$N = 10^6$	
Expectation	2.1170	2.9074	2.0926	2.1179	1.6487
Variance	2.2237	3.0378	2.7821	2.9119	1.3591

Note that Monte Carlo is an unbiased estimator of $\mathbb{E}[Y]$, i.e., $\mathbb{E}\left[\overline{Y}_{\text{mc}}\right] = \mathbb{E}[Y]$, but it may require an extremely high number of samples to produce a good estimate. On the other hand, linearisation works well when the variance is small. The linearisation method leads to the extended Kalman Filter, while Monte Carlo principles are used in the Particle Filter.

15.2 Extended Kalman Filter

15.2.1 Extended Kalman Filter

Consider the following nonlinear system

$$x_{t+1} = f(x_t, w_t), \tag{15.8a}$$

$$y_t = h(x_t, v_t), \tag{15.8b}$$

where x_0, $(w_t)_t$ and $(v_t)_t$ are independent, $(w_t)_t$ and $(v_t)_t$ have zero mean and covariance matrices Q_t and R_t respectively. Even if w and v are Gaussian, x_t and y_t may not be (and typically are not).

The objective is to determine $\hat{x}_{t+1|t}$, $\hat{x}_{t|t}$ and the corresponding covariance matrices.

EKF uses the linearisation technique. It can work well, especially when variance is low or the system is approximately linear. It is a very popular in practice (especially in navigation). It is not an optimal estimator; may even diverge.

How it works: use the linearisation method to produce approximate measurement and time update steps.

For the *measurement update* step, we start with the standard initialisation: $\hat{x}_{0|-1} = \tilde{x}_0$ and $\Sigma_{0|-1} = P_0$ and we linearise the output function at $x = \hat{x}_{t|t-1}$:

$$C = J_x h(\hat{x}_{t|t-1}, 0), \tag{15.9a}$$

$$\widetilde{R} = J_v h(\hat{x}_{t|t-1}, 0)\, R\, J_v h(\hat{x}_{t|t-1}, 0)^\intercal \tag{15.9b}$$

The measurement update is

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \Sigma_{t|t-1} C^\intercal (C \Sigma_{t|t-1} C^\intercal + \widetilde{R})^{-1} (y_t - h(x_{t|t-1}, 0)) \tag{15.10a}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} C^\intercal (C \Sigma_{t|t-1} C^\intercal + \widetilde{R})^{-1} C \Sigma_{t|t-1} \tag{15.10b}$$

Then, for the time update step, we linearise the dynamics at $x = \hat{x}_{t|t}$

$$A = J_x f(\hat{x}_{t|t}, 0), \tag{15.11a}$$

$$\widetilde{Q} = J_w f(\hat{x}_{t|t}, 0)\, Q\, J_w f(\hat{x}_{t|t}, 0)^\intercal \tag{15.11b}$$

The time update is

$$\hat{x}_{t+1|t} = f(\hat{x}_{t|t}, 0), \tag{15.12a}$$

$$\Sigma_{t+1|t} = A\Sigma_{t|t}A^\intercal + \widetilde{Q}. \tag{15.12b}$$

15.3 EKF Application

Let $r_t, u_t, a_t \in \mathbb{R}^2$ be the (planar) position, velocity and acceleration of a vehicle with dynamics

$$r_{t+1} = r_t + 0.2u_t, \tag{15.13a}$$

$$u_{t+1} = u_t + 0.2a_t, \tag{15.13b}$$

$$a_{t+1} = \Phi a_t + w_t, \tag{15.13c}$$

where $w_t \sim \mathcal{N}(0_2, 0.2I_2)$, and

$$\Phi = \begin{bmatrix} 0.50 & 0.87 \\ -0.87 & 0.48 \end{bmatrix}. \tag{15.14}$$

Three beacons are positioned at $r^{(1)} = (3, 2)$, $r^{(2)} = (2, -3)$ and $r^{(3)} = (-5, 3)$; a sensor can measure the distances to these points — in particular $y_{i,t} = \|r - r^{(i)}\|_2 + v_{i,t},\ i \in \mathbb{N}_{[1,3]}$, where $v_{i,t} \sim \mathcal{N}(0, 4)$.

We have a system with state vector $\boldsymbol{x}_t = (r_t, u_t, a_t)$ and output y_t. The dynamics is linear

$$\boldsymbol{x}_{t+1} = \begin{bmatrix} I_2 & 0.2I_2 & \\ & I_2 & 0.2I_2 \\ & & \Phi \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} 0_{4\times 2} \\ I_2 \end{bmatrix} w_t,$$

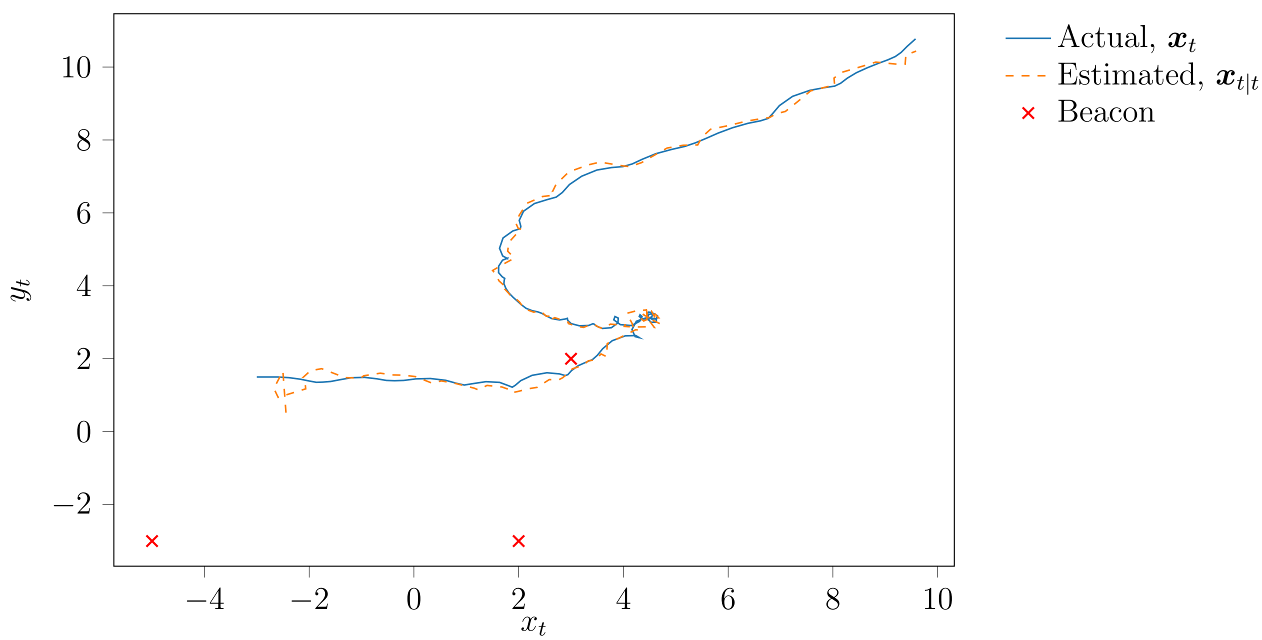
but the output equation is nonlinear

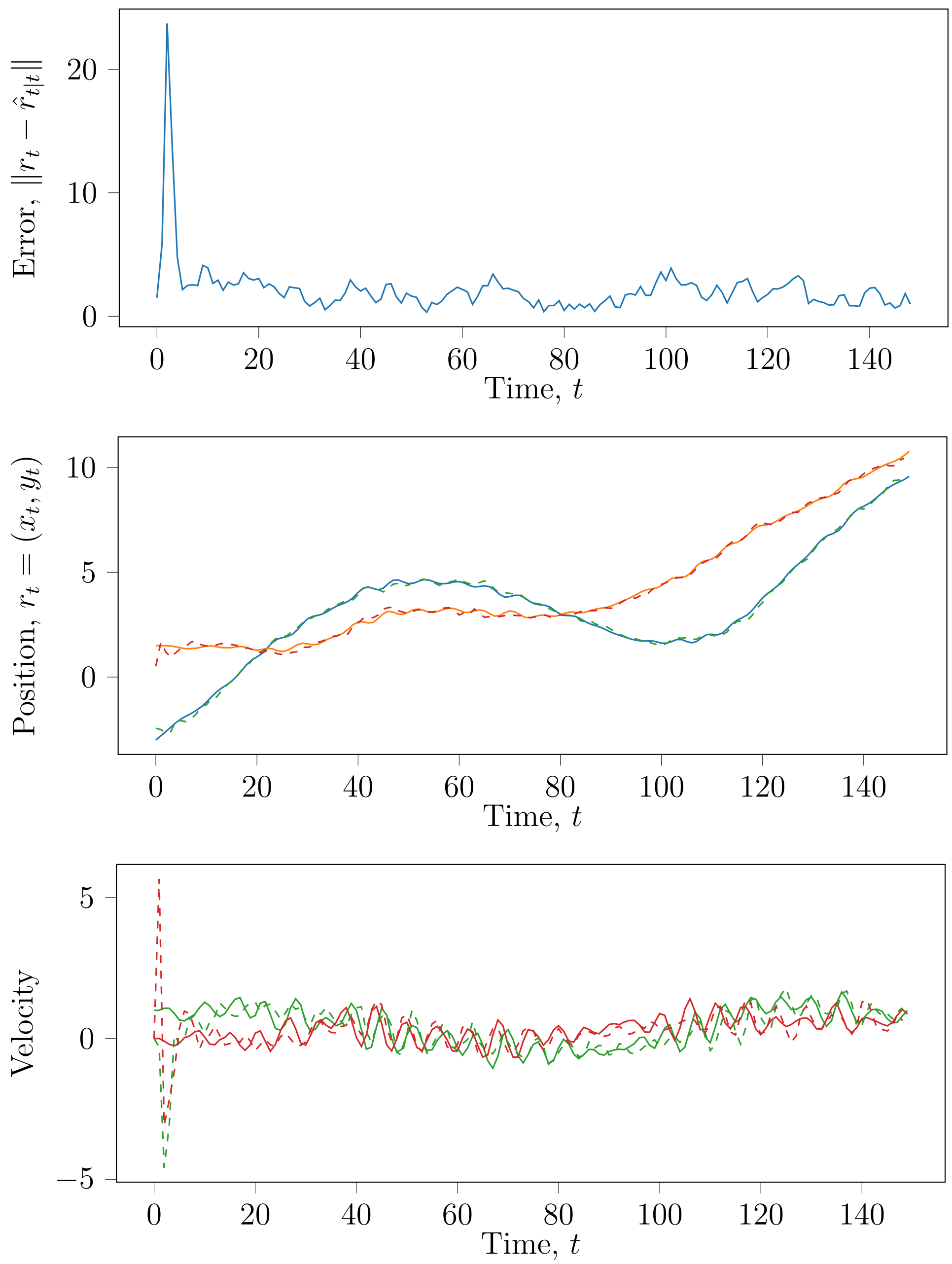
$$y_t = h(\boldsymbol{x}_t, v_t),$$

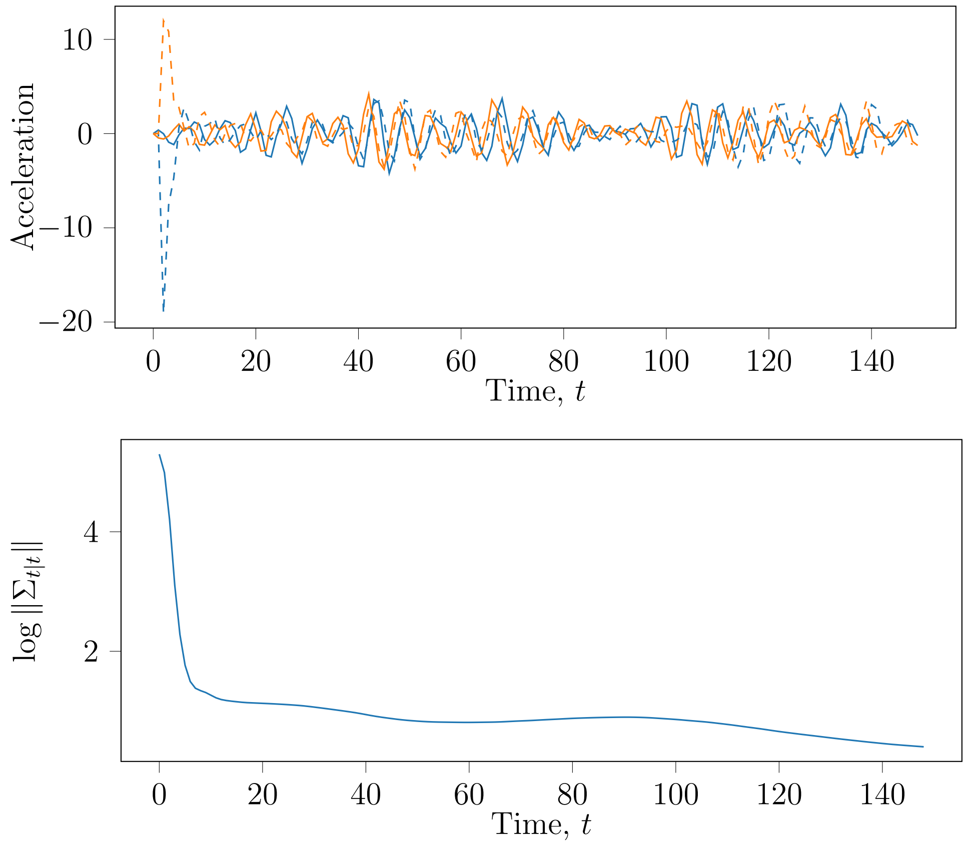
The gradient of $\|x\|$ is $\nabla\|x\| = x/\|x\|$, for $x \neq 0$. Let us denote this function by π. Then,

$$J_x h(x, v) = \begin{bmatrix} \pi(r_t - r^{(1)})^\intercal & \\ \pi(r_t - r^{(2)})^\intercal & 0_{3\times 4} \\ \pi(r_t - r^{(3)})^\intercal & \end{bmatrix},$$

and $J_v h(x, v) = I_3$. Some simulation results are shown below. You can play with the Python code on **Google Colab**. Some simulation results are shown below.




Error, $\|r_t - \hat{r}_{t|t}\|$
0
10
20
0 20 40 60 80 100 120 140
Time, t
Position, $r_t = (x_t, y_t)$
0
5
10
0 20 40 60 80 100 120 140
Time, t
Velocity
−5
0
5
0 20 40 60 80 100 120 140
Time, t


Acceleration
10
0
−10
−20
0
20
40
60
80
100
120
140
Time, t
$\log \|\Sigma_{t|t}\|$
4
2
0
20
40
60
80
100
120
140
Time, t

15.4 Moving Horizon Estimation

Consider the nonlinear dynamical system

$$x_{t+1} = f(x_t, w_t), \tag{15.15a}$$

$$y_t = h(x_t) + v_t, \tag{15.15b}$$

where x_0, w_t and v_t are temporally and mutually uncorrelated, w_t and v_t follow known distributions (not necessarily normal)[4]

$$p_{w_t}(w) \propto \exp[-\ell_w(w)], \tag{15.16a}$$

$$p_{v_t}(v) \propto \exp[-\ell_v(v)], \tag{15.16b}$$

$$p_{x_0}(x_0) \propto \exp[-\ell_{x_0}(x_0)]. \tag{15.16c}$$

The disturbance terms can be constrained — in particular we assume that $w_t \in W$ and $v_t \in V$ (nonempty, closed), Constrained states: $x_t \in X$. The extended Kalman filter cannot take into account constraints.

In order to describe the constraints we assume that $\ell_w(w) = \infty$, for $w \notin W$, $\ell_v(v) =$ for $v \notin V$, and $\ell_{x_0}(x_0) = \infty$, for $x_0 \notin X$, therefore $\mathbf{dom}\, \ell_w = W$, $\mathbf{dom}\, \ell_v = V$, $\mathbf{dom}\, \ell_{x_0} = X$.

Our objective is to estimate $x_{0:T} = (x_0, \ldots, x_T)$ and $w_{0:T-1} = (w_0, \ldots, w_{T-1})$ from measurements $y_{0:T-1} = (y_0, \ldots, y_{T-1})$ using a MAP approach. The posterior distribution is

$$p(x_{0:T}, w_{0:T-1} \mid y_{0:T-1}) \propto p(y_{0:T-1} \mid x_{0:T}, w_{0:T-1}) p(x_{0:T}, w_{0:T-1}). \tag{15.17}$$

We can show that

$$p(x_{0:T}, w_{0:T-1}) = \begin{cases} 0, & \text{if not } x_{t+1} = f(x_t, w_t) \text{ or } x_{t+1} \notin X \\ p_{x_0}(x_0) \prod_{t=0}^{N-1} p_{w_t}(w_t), & \text{otherwise} \end{cases} \tag{15.18}$$

This can be written as

$$p(x_{0:T}, w_{0:T-1}) = p_{x_0}(x_0) \prod_{t=0}^{N-1} p_{w_t}(w_t) \mathbb{1}_{\{0\}}(x_{t+1} - f(x_t, w_t)) \mathbb{1}_X(x_{t+1}), \tag{15.19}$$

[4] Since $w \in W$, it is $p_w(w) = 0$ for $w \notin W$.

where $\mathbb{1}_A(s) = 1$ if $s \in A$ and $\mathbb{1}_A(s) = 0$ otherwise. Furthermore,

$$p(y_{0:T-1} \mid x_{0:T}, w_{0:T-1}) = \prod_{t=0}^{T-1} p(y_t \mid x_t) = \prod_{t=0}^{T-1} p_{v_t}(y_t - h(x_t)). \tag{15.20}$$

With the convention $-\ln 0 = \infty$, we have $-\ln \mathbb{1}_A(s) = \delta_A(s)$, define the function

$$\begin{aligned} L(x_{0:T}, w_{0:T-1} \mid y_{0:T-1}) = & -\ln p(x_{0:T}, w_{0:T-1} \mid y_{0:T-1}) \\ &= \ell_{x_0}(x_0) + \sum_{t=0}^{T-1} \ell_{w_t}(w_t) + \ell_{v_t}(y_t - h(x_t)) \\ &\quad + \delta_{\{0\}}(x_{t+1} - f(x_t, w_t)) + \delta_X(x_{t+1}). \end{aligned} \tag{15.21}$$

The MAP estimate of $x_{0:T}$ and $w_{0:T-1}$ given $y_{0:T-1}$ can be determined by solving

$$\operatorname*{minimise}_{x_{0:T}, w_{0:T-1}} L(x_{0:T}, w_{0:T-1} \mid y_{0:T-1}), \tag{15.22}$$

which is equivalently written as

$$\operatorname*{minimise}_{x_{0:T}, w_{0:T-1}} \ell_{x_0}(x_0) + \sum_{t=0}^{T-1} \ell_{w_t}(w_t) + \ell_{v_t}(y_t - h(x_t)), \tag{15.23a}$$

$$\text{subject to: } x_{t+1} = f(x_t, w_t),\ t \in \mathbb{N}_{[0,T-1]}, \tag{15.23b}$$

$$x_t \in X,\ t \in \mathbb{N}_{[0,T]}. \tag{15.23c}$$

$$\begin{aligned} \mathbb{P}_T^{\text{map}}(y_{0:T-1}) :\ & \operatorname*{minimise}_{x_{0:T}, w_{0:T-1}} \ell_{x_0}(x_0) + \sum_{t=0}^{T-1} \ell_{w_t}(w_t) + \ell_{v_t}(y_t - h(x_t)), \\ & \text{subject to: } x_{t+1} = f(x_t, w_t),\ t \in \mathbb{N}_{[0,T-1]}, \\ & x_t \in X,\ t \in \mathbb{N}_{[0,T]}. \end{aligned}$$

At time T, having observed $y_{0:T-1}$ we can solve the above problem to estimate $x_{0:T}$ and $w_{0:T-1}$.

This problem can be solved explicitly if f is linear and w_t, v_t and x_0 are independent and normally distributed. However, as time goes on and we accumulate more measurements, the size of the optimisation problem grows unbounded. We will use (forward) DP to cast this problem as a fixed-horizon optimal control problem (a problem with horizon N that does not grow).

15.4.1 Forward DP

Let us apply the *forward* dynamic programming procedure as we did in Handout 14.

$$\begin{aligned}
\widehat{V}_T^\star &= \min_{x_{0:T}, w_{0:T-1}} \left\{ \ell_{x_0}(x_0) + \sum_{t=0}^{T-1} \underbrace{\ell_{w_t}(w_t) + \ell_{v_t}(y_t - h(x_t))}_{\ell_t(x_t, w_t)} \;\middle|\; \begin{array}{l} x_{t+1} = f(x_t, w_t), \\ t \in \mathbb{N}_{[0,T-1]} \end{array} \right\} \\
&= \min_{x_{0:T}, w_{0:T-1}} \left\{ \ell_{x_0}(x_0) + \sum_{t=0}^{T-1} \ell_t(x_t, w_t) \;\middle|\; \begin{array}{l} x_{t+1} = f(x_t, w_t), \\ t \in \mathbb{N}_{[0,T-1]} \end{array} \right\} \\
&= \min_{x_{1:T}, w_{1:T-1}} \left\{ \underbrace{\min_{x_0, w_0} \left\{ \ell_{x_0}(x_0) + \ell_0(x_0, w_0) \mid x_1 = f(x_0, w_0) \right\}}_{V_1^\star(x_1)} \right. \\
&\qquad\qquad \left. + \sum_{t=1}^{T-1} \ell_t(x_t, w_t) \;\middle|\; \begin{array}{l} x_{t+1} = f(x_t, w_t), \\ t \in \mathbb{N}_{[1,T-1]} \end{array} \right\} \\
&= \min_{x_{1:T}, w_{1:T-1}} \left\{ V_1^\star(x_1) + \sum_{t=1}^{T-1} \ell_t(x_t, w_t) \;\middle|\; \begin{array}{l} x_{t+1} = f(x_t, w_t), \\ t \in \mathbb{N}_{[1,T-1]} \end{array} \right\},
\end{aligned} \tag{15.24}$$

and we can keep applying this procedure forwards. The *forward* dynamic programming procedure can be written concisely as

$$V_0^\star(x_0) = \ell_{x_0}(x_0), \tag{15.25a}$$

$$V_{t+1}^\star(x_{t+1}) = \min_{x_0, w_0} \left\{ V_t^\star(x_t) + \ell_t(x_t, w_t) \mid x_{t+1} = f(x_t, w_t) \right\}, \tag{15.25b}$$

and $\widehat{V}_T^\star = \min_{x_T} V_T^\star(x_T)$.[5] This allows us to write the MAP estimation problem as

$$\underset{x_{T-N:T}, w_{T-N:T-1}}{\text{minimise}} \; V_{T-N}^\star(x_{T-N}) + \sum_{t=T-N}^{T-1} \ell_t(x_t, w_t), \tag{15.26a}$$

$$\text{subject to: } x_{t+1} = f(x_t, w_t),\, t \in \mathbb{N}_{[T-N,T-1]}, \tag{15.26b}$$

$$x_t \in X,\, t \in \mathbb{N}_{[T-N,T]}. \tag{15.26c}$$

This problem uses a window of measurements of fixed length N. However, the *arrival cost* $V_{T-N}^\star$ is very difficult to compute.

... instead, we shall use a different *prior weighting* function $\Gamma_{T-N}(x)$ and solve

$$\underset{x_{T-N:T}, w_{T-N:T-1}}{\text{minimise}} \; \Gamma_{T-N}(x_{T-N}) + \sum_{t=T-N}^{T-1} \ell_t(x_t, w_t), \tag{15.27a}$$

$$\text{subject to: } x_{t+1} = f(x_t, w_t),\, t \in \mathbb{N}_{[T-N,T-1]}, \tag{15.27b}$$

$$x_t \in X,\, t \in \mathbb{N}_{[T-N,T]}. \tag{15.27c}$$

One possible choice for the prior weighting is $\Gamma_{T-N}(x) = 0$. If N is sufficiently large and some additional assumptions are satisfied, this works: the estimation error goes to zero if the disturbances go to zero.

15.5 EKF vs MHE

Example taken from: E.L. Haseltine & J.B. Rawlings, A Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation, Ind.Eng.Chem.Res. 2005, 44(8):2451-60.

Consider the following chemical reaction take takes place in gaseous phase

$$2A \longrightarrow B,$$

with rate coefficient $k = 0.16$ and reaction rate $r = kP_A^2$, where P_A and P_B are the *partial pressures* of A and B respectively. The state is the vector $x = [P_A\ P_B]^\intercal$ and the system

[5] Note that $V_t^\star$ corresponds to $-\ln p(x_t \mid y_{0:t-1})$.

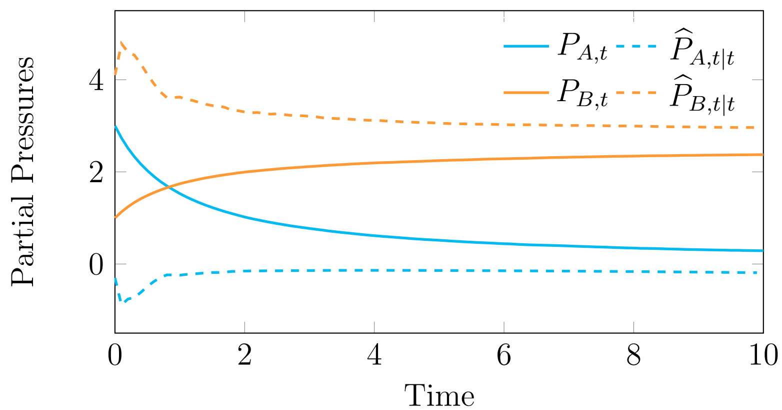
dynamics is

$$x_{t+1} = \underbrace{\begin{bmatrix} \frac{x_{t,1}}{2k\Delta t x_{t,1}+1} \\ x_{t,2} + \frac{k\Delta t x_{t,1}^2}{2k\Delta t x_{t,1}+1} \end{bmatrix}}_{F(x_t)} + w_t,$$

where $\Delta t = 0.1$ is the sampling time. We can measure the total pressure, that is

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + v_t,$$

where $v_t \sim \mathcal{N}(0, 0.1^2)$ ($R = 0.1^2$) and $w_t \sim \mathcal{N}(0, 0.001^2 I_2)$ ($Q = 0.001^2 I_2$). Suppose that $x_0 = [3\ 1]^\intercal$ while our prior knowledge about x_0 is $\bar{x}_0 = [0.1\ 4.5]^\intercal$ and $P_0 = 6^2 I_2$.



We see that the estimates are of rather poor quality and $\widehat{P}_{B,t|t}$ gives negative pressure values, which do not make sense.

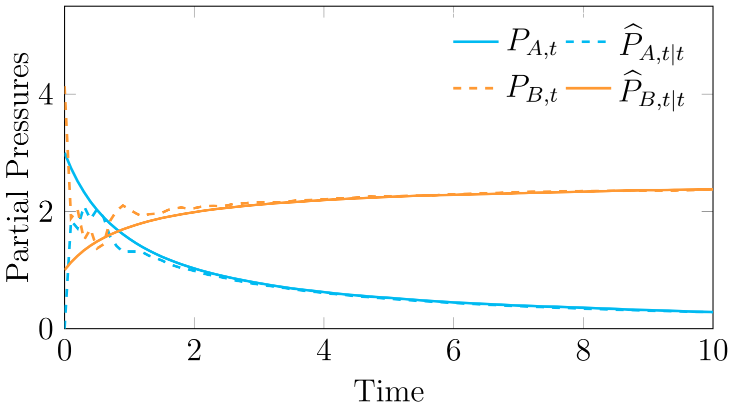
Next, we formulate the following FIE-MHE problem that is solved at time $T-1$ using all measurements $y_{0:T-1}$

$$\mathbb{P}_T(y_{0:T-1}) : \underset{x_{0:T}, w_{0:T-1}, v_{0:T-1}}{\text{minimise}} \ \tfrac{1}{2}\|x_0 - \bar{x}_0\|^2_{P_0^{-1}} + \tfrac{1}{2}\sum_{t=0}^{T-1} \|w_t\|^2_{Q^{-1}} + \|v_t\|^2_{R^{-1}}, \tag{15.28a}$$

$$\text{subject to: } x_{t+1} = F(x_t) + w_t,\ t \in \mathbb{N}_{[0,T-1]}, \tag{15.28b}$$

$$y_t = [1\ 1]x_t + v_t,\ t \in \mathbb{N}_{[0,T-1]}, \tag{15.28c}$$

$$x_t \geq 0,\ t \in \mathbb{N}_{[0,T]}. \tag{15.28d}$$



$P_{A,t}$
$\widehat{P}_{A,t|t}$
$P_{B,t}$
$\widehat{P}_{B,t|t}$
Partial Pressures
Time
0
2
4
6
8
10