Bioprocess modeling and estimation

An introduction to state estimation

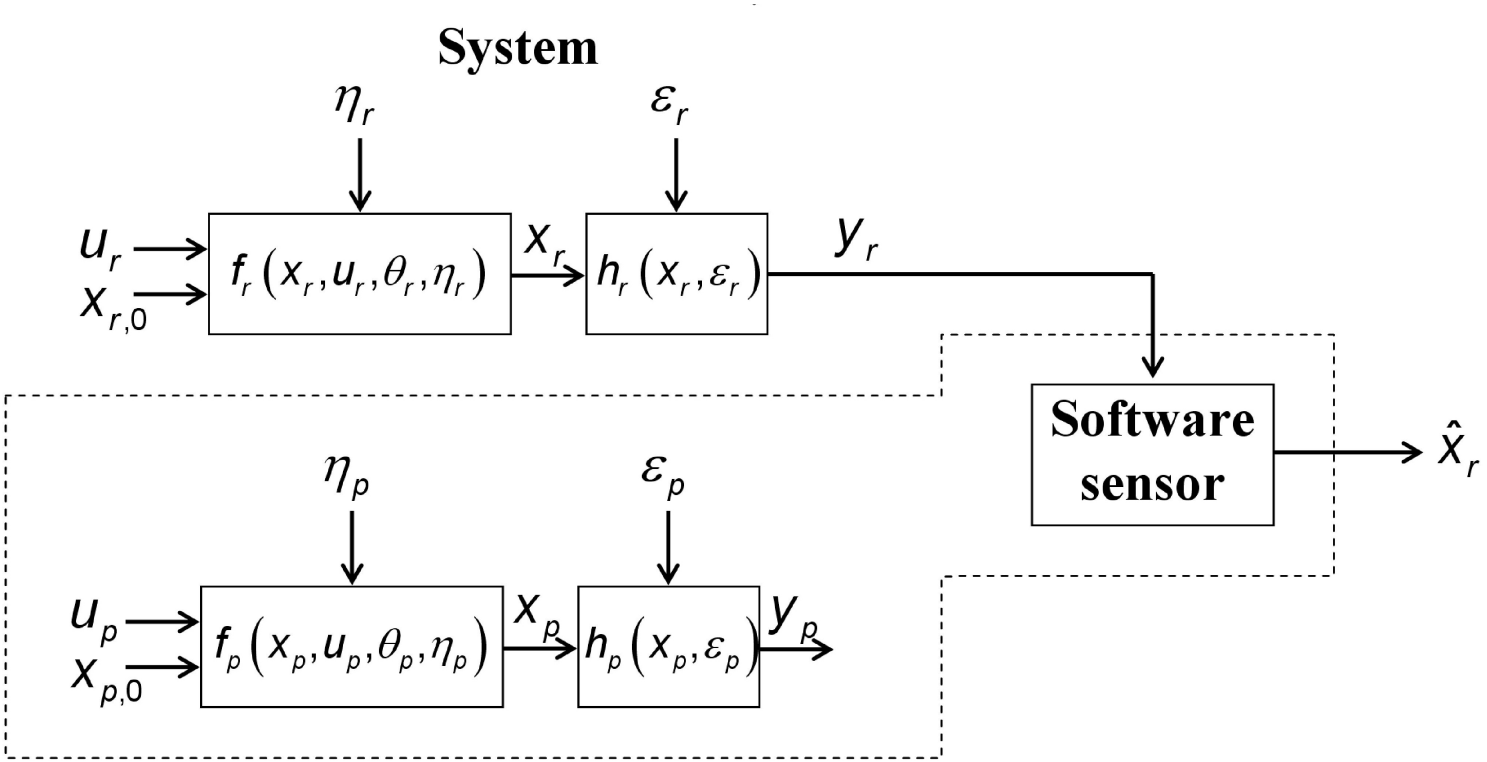
Alain Vande Wouwer[1] Philippe Bogaerts[2]
Special thanks to Guillaume Goffaux[1]

[1]Control Department
University of Mons

[2]3BIO
University of Brussels

Graduate School, March 2010

The state estimation problem



A link between parameter and state estimation

Back to parameter estimation ... linear scalar case

$$y(i) = \underline{x}^T(i)\underline{\theta}(i) + \varepsilon(i) \qquad i = [1, \cdots, M]$$

with time varying parameters expressed as an evolution model

$$\underline{\theta}(i+1) = A(i)\underline{\theta}(i) + \underline{\eta}(i)$$

with

- $E[\varepsilon(i)] = 0$ and $E[\varepsilon(i)\varepsilon(j)] = \sigma^2(i)\delta(i-j)$
- $E[\underline{\eta}(i)] = 0$ and $E[\underline{\eta}(i)\underline{\eta}^T(j)] = Q(i)\delta(i-j)$

As a starting point, we assume to know a posteriori estimates (i.e. taking account of the current measurement)

- $\hat{\underline{\theta}}(i|i) = \hat{\underline{\theta}}^+(i)$
- $P^+(i) = E[\tilde{\underline{\theta}}^+(i)\tilde{\underline{\theta}}^{+T}(i)]$

A link between parameter and state estimation

Based on this assumption, the first step is to build a prediction

$$\underline{\hat{\theta}}^{-}(i+1) = A(i)\underline{\hat{\theta}}^{+}(i)$$

which leads to an unbiased estimate ($E[\underline{\tilde{\theta}}^{-}(i+1)] = 0$) if

$$E[\underline{\tilde{\theta}}^{+}(i)] = 0$$

The covariance of the predictor is given by

$$\begin{aligned} P^{-}(i+1) &= E[\underline{\tilde{\theta}}^{-}(i+1)\underline{\tilde{\theta}}^{-T}(i+1)] \\ &= A(i)E[\underline{\tilde{\theta}}^{+}(i)\underline{\tilde{\theta}}^{+T}(i)]A^{T}(i) + E[\underline{\eta}(i)\underline{\eta}^{T}(i)] \\ &= A(i)P^{+}(i)A^{T}(i) + Q(i) \end{aligned}$$

A link between parameter and state estimation

The correction step is provided by the RLS

$$\underline{\hat{\theta}}^{+}(i+1) = \underline{\hat{\theta}}^{-}(i+1) + \frac{P^{-}(i+1)\underline{x}(i)}{\sigma^2(i)+\underline{x}^T(i)P^{-}(i+1)\underline{x}(i)}[y(i+1) - \hat{y}^{-}(i+1)]$$
$$P^{+}(i+1) = P^{-}(i+1) - \frac{P^{-}(i+1)\underline{x}(i)\underline{x}^T(i)P^{-}(i+1)}{\sigma^2(i)+\underline{x}^T(i)P^{-}(i+1)\underline{x}(i)}$$

A link between parameter and state estimation

If the problem is translated to state estimation ... $\underline{\theta}(i) \rightarrow \underline{x}(i)$ and $\underline{x}(i) \rightarrow C(i)$

$$\begin{aligned}\underline{x}(i+1) &= A(i)\underline{x}(i) + B(i)\underline{u}(i) + \underline{\eta}(i)\\ \underline{y}(i) &= C(i)\underline{x}(i) + \underline{\varepsilon}(i)\end{aligned}$$

The prediction is given by

$$\begin{aligned}\underline{\hat{x}}^{-}(i+1) &= A(i)\underline{\hat{x}}^{+}(i) + B(i)u(i)\\ P^{-}(i+1) &= A(i)P^{+}(i)A^{T}(i) + Q(i)\end{aligned}$$

and the correction by

$$\begin{aligned}\underline{\hat{x}}^{+}(i+1) &= \underline{\hat{x}}^{-}(i+1) + K(i+1)[y(i+1) - C(i+1)\hat{x}^{-}(i+1)]\\ P^{+}(i+1) &= P^{-}(i+1) - K(i+1)C(i+1)P^{-}(i+1)\end{aligned}$$

$$K(i+1) = P^{-}(i+1)C^{T}(i+1)\Big(R(i+1) + C(i+1)P^{-}(i+1)C^{T}(i+1)\Big)^{-1}$$

which is the Kalman filter we will address later on ...

Overview of the talk

- Observability
- The (extended) Kalman filter
- The asymptotic observer
- A hybrid Kalman - asymptotic observer
- The full and receding-horizon observer
- Interval observers

Indistinguishability

Consider a nonlinear dynamic system

$$\begin{aligned} \dot{\underline{x}}(t) &= \underline{f}(\underline{x}(t), \underline{u}(t), \underline{\theta}) \qquad \underline{x}(t_0) = \underline{x}_0 \\ \underline{y}(t) &= \underline{h}(\underline{x}(t)) \end{aligned}$$

The solution is given by $\chi_u\left(t_0, t, \underline{x}_0, \underline{\theta}\right)$

The pair of initial states $\underline{\mathrm{x}}_0^1$ *and* $\underline{\mathrm{x}}_0^2$ *is* ***indistinguishable****, if for every* $\underline{\mathrm{u}} \in \mathcal{U}$ *defined on* $t_0 \leq t \leq T$ *:*

$$\underline{\mathrm{h}}\left(\chi_u\left(t_0, t, \underline{\mathrm{x}}_0^1, \underline{\theta}\right)\right) = \underline{\mathrm{h}}\left(\chi_u\left(t_0, t, \underline{\mathrm{x}}_0^2, \underline{\theta}\right)\right)$$

$\mathcal{I}(\underline{x}_0)$ *denotes the set of indistinguishable states related to* $\underline{x}_0$

Observability

The system is **observable** if there are no indistinguishable pairs

$$\forall t_0,\ \forall \mathbf{x}_0^1, \mathbf{x}_0^2 \in \mathcal{X},\ \forall \underline{u} \in \mathcal{U},\ \exists T < \infty,\ \textit{such that}$$

$$\underline{\mathrm{h}}\left(\chi_u\left(t_0, t,\ \underline{\mathrm{x}}_0^1, \underline{\theta}\right)\right) = \underline{\mathrm{h}}\left(\chi_u\left(t_0, t, \underline{\mathrm{x}}_0^2, \underline{\theta}\right)\right) \qquad t_0 \leq t < T$$

$$\Rightarrow \quad \underline{\mathrm{x}}_0^1 = \underline{\mathrm{x}}_0^2$$

Observability and observers

Observability is related to the reconstruction of the system state $\underline{\hat{x}}(t)$, with the following minimum objectives:

1. an exact estimation for a perfect initialization : $\underline{\hat{x}}(t_0) = \underline{x}(t_0) \Rightarrow \underline{\hat{x}}(t) = \underline{x}(t)$,
2. an asymptotic estimation convergence : $\lim_{t\to\infty} \left\|\underline{\hat{x}}(t) - \underline{x}(t)\right\| = 0$.

Observability implies in addition that the convergence rate can be tuned (i.e. an exponential convergence rate).

Observability of linear time invariant (LTI) systems

$$\dot{\underline{x}}(t) = A\underline{x}(t) + B\underline{u}(t) \qquad \underline{x}(t_0) = \underline{x}_0$$
$$\underline{y}(t) = C\underline{x}(t)$$

Can we infer $\underline{x}(t)$ from $\underline{y}(t)$? Not directly, but we can get further information by differentiating m times the outputs:

$$\begin{aligned}
&\underline{y}(t) = C\underline{x}(t) \\
&\dot{\underline{y}}(t) = C\dot{\underline{x}}(t) = CA\underline{x}(t) + CB\underline{u}(t) \\
&\ddot{\underline{y}}(t) = CA\dot{\underline{x}}(t) + CB\dot{\underline{u}}(t) = CA^2\underline{x}(t) + CAB\underline{u}(t) + CB\dot{\underline{u}}(t) \\
&\vdots \\
&\underline{y}^{(m)}(t) = CA^m\underline{x}(t) + \textstyle\sum_{j=0}^{m-1} CA^{m-1-j}B\underline{u}^{(j)}(t)
\end{aligned}$$

These equations can be put in a more compact form:

$$\mathbf{Y}(t) = O\underline{x}(t) + T\mathbf{U}(t)$$

and the observability condition is $\mathrm{Rank}(O) = n_x$

Observability of nonlinear systems

The observability map is given by :

$$\underline{q}(\underline{x},\underline{u}) = \begin{bmatrix} \underline{q}_1(\underline{x},\underline{u}) \\ \vdots \\ \underline{q}_{n_y}(\underline{x},\underline{u}) \end{bmatrix}, \quad dim(\underline{q}(\underline{x},\underline{u})) = n_q = \sum_{j=1}^{n_y} n_{q_j} \geq n_x$$

$$\underline{q}_j(\underline{x},\underline{u}) = \begin{bmatrix} y_j(t) \\ \dot{y}_j(t) \\ \vdots \\ y^{(n_{q_j}-1)}(t) \end{bmatrix} = \begin{bmatrix} q_{j,1}(\underline{x},\underline{u}) \\ q_{j,2}(\underline{x},\underline{u}) \\ \vdots \\ q_{j,n_{q_j}}(\underline{x},\underline{u}) \end{bmatrix} = \begin{bmatrix} h_j(\underline{x}) \\ \frac{\partial q_{j,1}}{\partial \underline{x}} \underline{f}(\underline{x},\underline{u}) + \frac{\partial q_{j,1}}{\partial \underline{u}} \dot{\underline{u}}(t) \\ \vdots \\ \frac{\partial q_{j,n_{q_j}-1}}{\partial \underline{x}} \underline{f}(\underline{x},\underline{u}) + \frac{\partial q_{j,n_{q_j}-1}}{\partial \underline{u}} \dot{\underline{u}}(t) \end{bmatrix}$$

If a partition of the complete observability map is injective with respect to the state variables, then the system is observable. However, the analytical solution of the nonlinear system is difficult, or even impossible.

Observability of nonlinear systems

Local observability can be checked thanks to the inverse mapping theorem (local inverse)

$$O(\underline{x}, \underline{u}) = \frac{\partial \underline{q}(\underline{x}, \underline{u})}{\partial \underline{x}}$$

If the Jacobian is nonsingular, an inverse function exists in some neighborhood of $\underline{x}$ and local observability is proven.

$$\mathrm{rank}\,(O(\underline{x}, \underline{u})) = n_x \quad \text{or} \quad \det(O(\underline{x}, \underline{u})) \neq 0$$

Observability of nonlinear systems

The global observability analysis of nonlinear models can be simplified through the introduction of a canonical form [Zeitz, 84, 89; Gauthier-Kupka, 94]:

$$\dot{\underline{x}} = \begin{bmatrix} \dot{\underline{x}}^1 \\ \dots \\ \dot{\underline{x}}^i \\ \dots \\ \dot{\underline{x}}^{q-1} \\ \dot{\underline{x}}^q \end{bmatrix} = \begin{bmatrix} \underline{f}^1(\underline{x}^1, \underline{x}^2, \underline{u}) \\ \dots \\ \underline{f}^i(\underline{x}^1, \dots, \underline{x}^{i+1}, \underline{u}) \\ \dots \\ \underline{f}^{q-1}(\underline{x}^1, \dots, \underline{x}^q, \underline{u}) \\ \underline{f}^q(\underline{x}^1, \dots, \underline{x}^q, \underline{u}) \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} h_1(x_1^1) \\ h_2(x_1^1, x_2^1) \\ \dots \\ h_{n_1}(x_1^1, \dots, x_{n_1}^1) \end{bmatrix}$$

where

$$\underline{x}^T = \left[\underline{x}^1, \dots, \underline{x}^q\right], \quad \underline{f}^T = \left[\underline{f}^1, \dots, \underline{f}^q\right], \quad \underline{x}^{1,T} = \left[x_1^1, \dots, x_{n_1}^1\right], \quad \underline{h}^T = [h_1, \dots, h_{n_1}]$$

$$\forall i \in \{1, \dots, q\}, \quad \underline{x}^i \in \mathbb{R}^{n_i}, \quad n_1 \geq n_2 \geq \dots \geq n_q, \quad \sum_{1 \leq i \leq q} n_i = n_x$$

Observability of nonlinear systems

A system is said globally observable if

$$\forall j \in \{1, ..., n_1\} : \quad \frac{\partial h_j}{\partial x_j^1} \neq 0$$

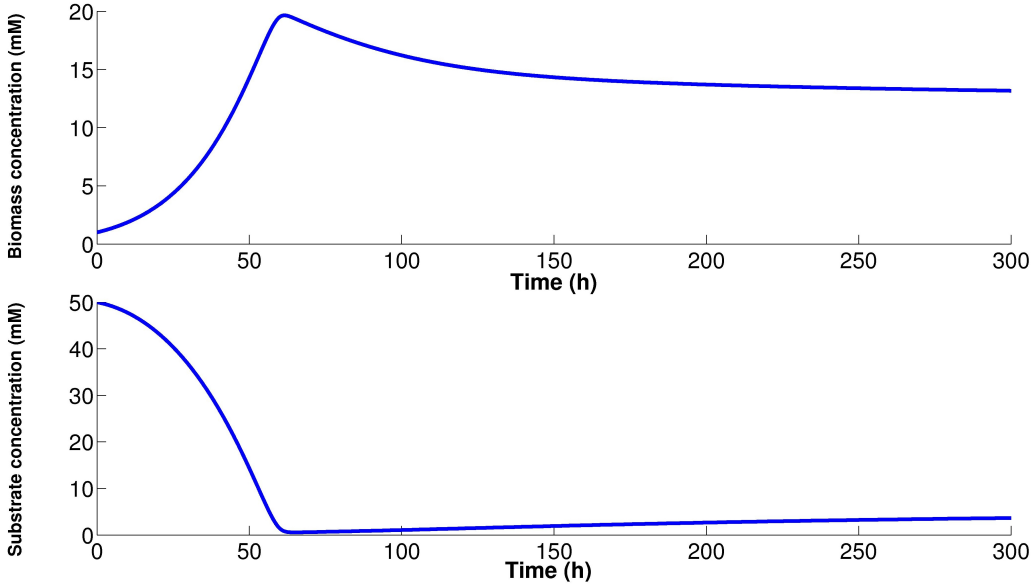
$$\forall i \in \{1, ..., q-1\}, \quad \forall (\underline{x}, \underline{u}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} : \quad \text{rank} \ \frac{\partial \underline{f}^i(\underline{x}, \underline{u})}{\partial \underline{x}^{i+1}} = n_{i+1}$$

The first conditions imply that the n_1 state variables can be inferred from the measurements, whereas the second ensure that any differences in the state trajectory can be detected in the measurements thanks to a pyramidal influence of the state subvector $\underline{x}^{i+1}$ on the evolution equations $\dot{\underline{x}}^i$.

Observability analysis: an example

$$k_S S \overset{\varphi}{\rightarrow}{}^{\wedge} X \qquad\qquad \varphi = \mu_m \frac{S}{S + K_S} X$$

$$\begin{aligned} \dot{X}(t) &= -D(t)X(t) + \varphi(X(t), S(t)) & X(t_0) = X_0 \\ \dot{S}(t) &= D(t)\left(S_{in}(t) - S(t)\right) - k_S \varphi(X(t), S(t)) & S(t_0) = S_0 \end{aligned}$$



Observability analysis: an example

Two cases are inspected according to the measured state variable :

As $n_y = 1$ and $n_x = 2$, a minimum size for the observability map is $n_q = 2$.

- $C = [1 \;\; 0]$ (biomass measurements) :

$$\begin{aligned} \underline{q}(\underline{x}, \underline{u}) &= \begin{bmatrix} X \\ \dot{X} \end{bmatrix} = \begin{bmatrix} X \\ -DX + \varphi(X, S) \end{bmatrix} \\ O(\underline{x}, \underline{u}) &= \begin{bmatrix} 1 & 0 \\ -D + \frac{\partial \varphi(X,S)}{\partial X} & \frac{\partial \varphi(X,S)}{\partial S} \end{bmatrix} \end{aligned}$$

with $\frac{\partial \varphi(X,S)}{\partial X} = \mu_m \frac{S}{S+K_S}$ and $\frac{\partial \varphi(X,S)}{\partial S} = \mu_m X \frac{K_S}{(S+K_S)^2}$

$$\Longrightarrow X \neq 0$$

Observability analysis: an example

- $C = [0 \;\; 1]$ (substrate measurements) :

$$\begin{aligned} \underline{q}(\underline{x}, \underline{u}) &= \begin{bmatrix} S \\ \dot{S} \end{bmatrix} = \begin{bmatrix} S \\ D\,(S_{in} - S) - k_S \varphi(X, S) \end{bmatrix} \\ O(\underline{x}, \underline{u}) &= \begin{bmatrix} 0 & 1 \\ -k_S \frac{\partial \varphi(X,S)}{\partial X} & -D - k_S \frac{\partial \varphi(X,S)}{\partial S} \end{bmatrix} \end{aligned}$$

with $\frac{\partial \varphi(X,S)}{\partial S} = \mu_m X \frac{K_S}{(S+K_S)^2}$ and $\frac{\partial \varphi(X,S)}{\partial X} = \mu_m \frac{S}{S+K_S}$

$$\Longrightarrow S \neq 0$$

Observability analysis: an example

These results are confirmed using a canonical form

$$\underline{\dot{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} f^1(x_1, x_2, \underline{u}) \\ f^2(x_1, x_2, \underline{u}) \end{bmatrix}, \quad y = h(x_1)$$

For instance, in the case of biomass measurements, $x_1 = X$ and $x_2 = S$:

$$\begin{aligned} \dot{x}_1(t) &= -D(t)x_1(t) + \varphi(x_1(t), x_2(t)) \\ \dot{x}_2(t) &= D(t)\left(S_{in}(t) - x_2(t)\right) - k_S\varphi(x_1(t), x_2(t)) \end{aligned}$$

$$\begin{aligned} x_1(t_0) &= X_0 \\ x_2(t_0) &= S_0 \end{aligned}$$

The Kalman filter

Continuous-time system equations and discrete-time measurements:

$$\begin{aligned} \dot{\underline{x}}(t) &= A(t)\underline{x}(t) + B(t)\underline{u}(t) + \underline{\eta}(t) \\ \underline{y}(t_k) &= C(t_k)\underline{x}(t_k) + \underline{\epsilon}(t_k) \end{aligned}$$

where $\underline{\eta}(t) \sim \mathcal{N}(0, R_\eta(t))$, $\underline{\epsilon}(t_k) \sim \mathcal{N}(0, R_\epsilon(t_k))$ and $\underline{x}(t_0) \sim \mathcal{N}(\underline{x}_0, P_0)$

The Kalman filter is the optimal Bayesian recursive filter, which allows the evaluation of the posterior probability density function:

$$\mathbb{P}(\underline{x}_k \,|\, Y_k) = \mathbb{P}(\underline{x}(t_k) \,|\, [\underline{y}(t_0), \ldots, \underline{y}(t_k)]) = \mathcal{N}(\underline{m}_{t_k|t_k}, P_{t_k|t_k})$$

from the prior density

$$\mathbb{P}(\underline{x}(t)|Y_{t_{k-1}}) = \mathcal{N}(\underline{m}_{t|t_{k-1}}, P_{t|t_{k-1}})$$

The Kalman filter

It requires an initialization

$$\underline{m}_{t_0|t_0} = \underline{x}_0 \qquad P_{t_0|t_0} = P_0$$

and proceeds in two steps: prediction + correction

- Continuous prediction step for $t_{k-1} \leq t \leq t_k$

$$\begin{aligned}
&\underline{\hat{x}}(t_{k-1}) = \underline{m}_{t_{k-1}|t_{k-1}} && P(t_{k-1}) = P_{t_{k-1}|t_{k-1}} \\
&\dot{\underline{\hat{x}}}(t) = A(t)\underline{\hat{x}}(t) + B(t)\underline{u}(t) && \dot{P}(t) = A(t)P(t) + P(t)A^T(t) + R_\eta(t \\
&\underline{m}_{t|t_{k-1}} = \underline{\hat{x}}(t) && P_{t|t_{k-1}} = P(t)
\end{aligned}$$

- Discrete correction step at $t = t_k$

$$\begin{aligned}
&\underline{m}_{t_k|t_{k-1}} = \underline{\hat{x}}(t_k) && P_{t_k|t_{k-1}} = P(t_k) \\
&K(t_k) = P_{t_k|t_{k-1}} C^T(t_k)(C(t_k)P_{t_{k|k-1}} C^T(t_k) + R_\epsilon(t_k))^{-1} \\
&\underline{m}_{t_k|t_k} = \underline{m}_{t_k|t_{k-1}} + K(t_k)(\underline{y}(t_k) - C(t_k)\underline{m}_{t_k|t_{k-1}}) \\
&P_{t_k|t_k} = P_{t_k|t_{k-1}} - K(t_k)C(t_k)P_{t_k|t_{k-1}}
\end{aligned}$$

The Kalman filter

- The initial state covariance matrix P_0 represents the confidence in the initial guess for the state vector $\underline{\hat{x}}_0$.
- The covariance matrix R_η describes the system noise. It is often "tuned" so as to keep some level of uncertainty on the state estimates preventing that the correction matrix $K(t_k)$ vanishes for $k \to \infty$, and so as to "adjust" the respective levels of confidence in the model and the measurements.
- The measurement error covariance matrix R_ϵ contains information on the measurement noise.
- The estimation error covariance matrix $P(t)$ can be used to assess the estimation accuracy and to compute confidence intervals for the estimators.

Extension to nonlinear systems: The extended Kalman filter

Continuous-time system equations and discrete-time measurements:

$$\begin{aligned} \dot{\underline{x}}(t) &= \underline{f}(\underline{x}(t), \underline{u}(t)) + \underline{\eta}(t) \\ \underline{y}(t_k) &= h(\underline{x}(t_k)) + \underline{\epsilon}(t_k) \end{aligned}$$

where $\underline{\eta}(t) \sim \mathcal{N}(0, R_\eta(t))$, $\underline{\epsilon}(t_k) \sim \mathcal{N}(0, R_\epsilon(t_k))$ and $\underline{x}(t_0) \sim \mathcal{N}(\underline{x}_0, P_0)$

A simple idea: use a linearization along the state estimate trajectory

$$F(\hat{\underline{x}}(t), \underline{u}(t)) = \left. \frac{\partial \underline{f}(\underline{x}(t), \underline{u}(t))}{\partial \underline{x}} \right|_{\underline{x}(t)=\widehat{\underline{x}}(t)}$$

$$C(\hat{\underline{x}}(t_k)) = \left. \frac{\partial \underline{h}(\underline{x}(t))}{\partial \underline{x}} \right|_{\underline{x}(t_k)=\hat{\underline{x}}(t_k)}$$

The EKF is an approximate (suboptimal) method. Nonlinear evolution equations do not preserve Gaussian distributions.

Extension to nonlinear systems: The extended Kalman filter

Continuous prediction equations :

$$
\begin{aligned}
\dot{\underline{\hat{x}}}(t) &= \underline{f}(\underline{\hat{x}}(t), \underline{u}(t)) \\
\dot{P}(t) &= F(\underline{\hat{x}}(t), \underline{u}(t))P(t) + P(t)F^T(\underline{\hat{x}}(t), \underline{u}(t)) + R_\eta(t)
\end{aligned}
$$

Discrete correction equations

$$
\begin{aligned}
K(t_k) &= P_{t_k|t_{k-1}} C^T(t_k)(C(t_k)P_{t_{k|k-1}} C^T(t_k) + R_\epsilon(t_k))^{-1} \\
\underline{m}_{t_k|t_k} &= \underline{m}_{t_k|t_{k-1}} + K(t_k)(\underline{y}(t_k) - \underline{h}(\underline{m}_{t_k|t_{k-1}})) \\
P_{t_k|t_k} &= P_{t_k|t_{k-1}} - K(t_k)C(t_k)P_{t_k|t_{k-1}}
\end{aligned}
$$

The extended Kalman filter: an example

Prediction

$$\dot{\hat{X}}(t) = -D(t)\hat{X}(t) + \mu_m \frac{\hat{S}(t)}{\hat{S}(t) + K_S}\hat{X}(t)$$

$$\dot{\hat{S}}(t) = D(t)\left(S_{in}(t) - \hat{S}(t)\right) - \nu\mu_m \frac{\hat{S}(t)}{\hat{S}(t) + K_S}\hat{X}(t)$$

$$\dot{P}(t) = F(\underline{\hat{x}}(t))P(t) + P(t)F^T(\underline{\hat{x}}(t)) + R_\eta(t)$$

where

$$F\left(\hat{X}, \hat{S}\right) = \begin{bmatrix} \mu_m \frac{\hat{S}(t)}{\hat{S}(t)+K_S} - D(t) & \mu_m \frac{K_S}{\left(\hat{S}(t)+K_S\right)^2}\hat{X}(t) \\ -\nu\mu_m \frac{\hat{S}(t)}{\hat{S}(t)+K_S} & -D(t) - \nu\mu_m \frac{K_S}{\left(\hat{S}(t)+K_S\right)^2}\hat{X}(t) \end{bmatrix}$$

The extended Kalman filter: an example

Correction

$$K(t_k) = \frac{1}{\sigma^2_{\hat{X}^-}(t_k) + \sigma^2_X} \begin{bmatrix} \sigma^2_{\hat{X}^-}(t_k) \\ \sigma_{\hat{X}^-\hat{S}^-}(t_k) \end{bmatrix}$$

$$\begin{bmatrix} \hat{X}^+(t_k) \\ \hat{S}^+(t_k) \end{bmatrix} = \begin{bmatrix} \hat{X}^-(t_k) \\ \hat{S}^-(t_k) \end{bmatrix} + K(t_k)\left(X(t_k) - \hat{X}^-(t_k)\right)$$

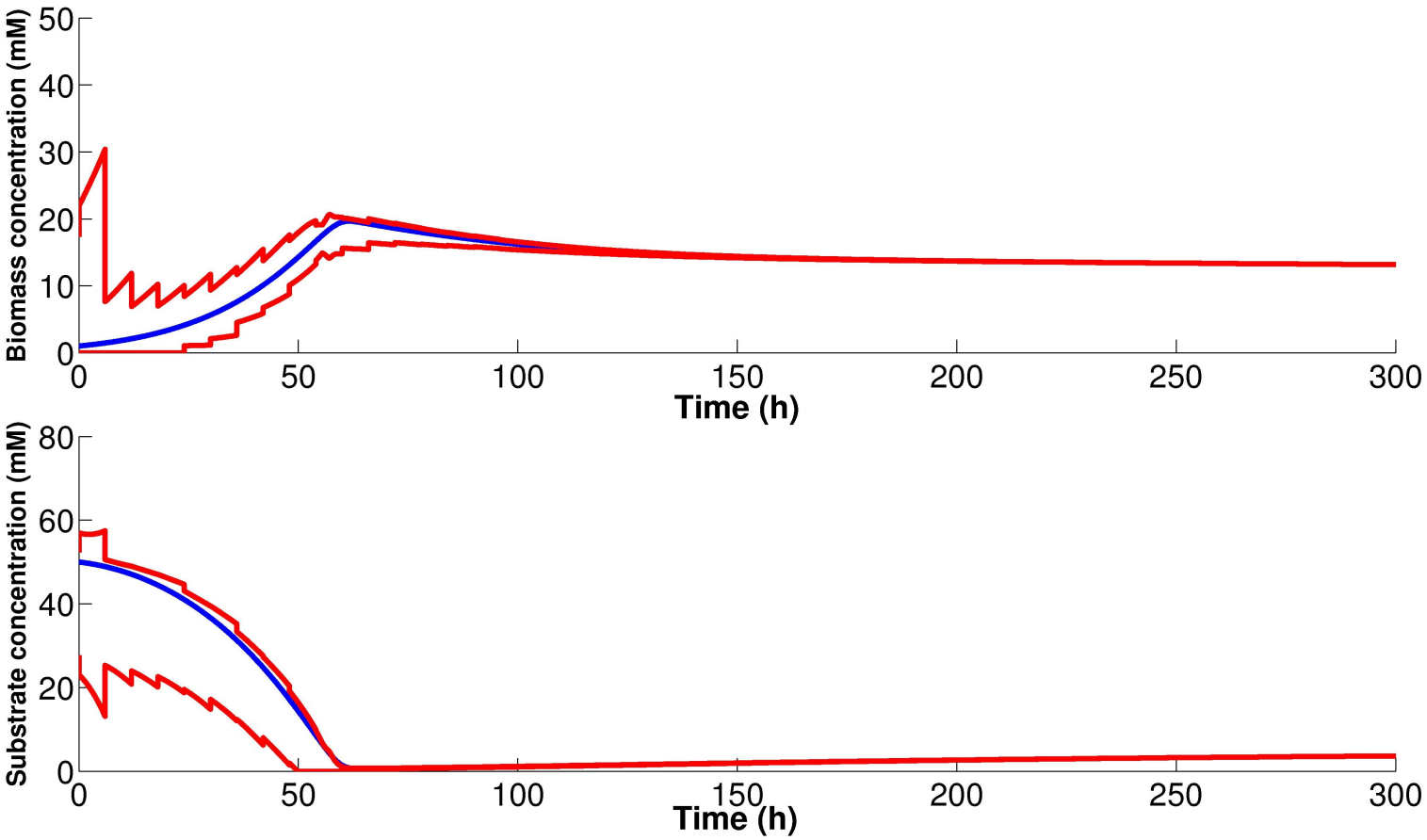
$$P_{t_k|t_k} = \begin{bmatrix} \sigma^2_{\hat{X}^+}(t_k) & \sigma_{\hat{X}^+\hat{S}^+}(t_k) \\ \sigma_{\hat{X}^+\hat{S}^+}(t_k) & \sigma^2_{\hat{S}^+}(t_k) \end{bmatrix}$$

$$\sigma^2_{\hat{X}^+}(t_k) = \frac{\sigma^2_X}{\sigma^2_{\hat{X}^-}(t_k) + \sigma^2_X}\sigma^2_{\hat{X}^-}(t_k) \qquad \sigma_{\hat{X}^+\hat{S}^+}(t_k) = \frac{\sigma^2_X}{\sigma^2_{\hat{X}^-}(t_k) + \sigma^2_X}\sigma_{\hat{X}^-\hat{S}^-}(t_k)$$

$$\sigma^2_{\hat{S}^+}(t_k) = \sigma^2_{\hat{S}^-}(t_k) - \frac{\sigma^2_{\hat{X}^-\hat{S}^-}(t_k)}{\sigma^2_{\hat{X}^-}(t_k) + \sigma^2_X}$$

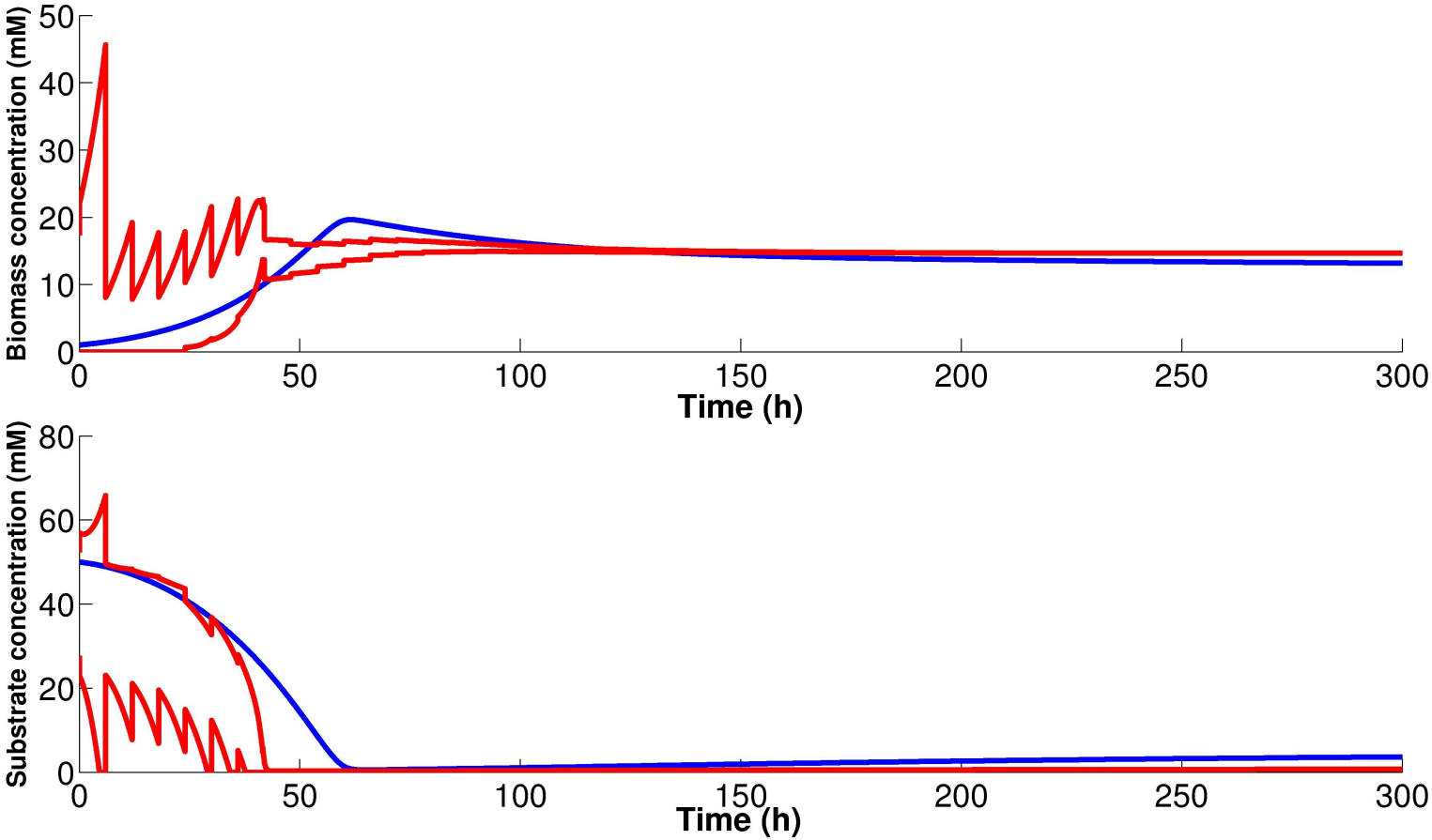
The extended Kalman filter: an example

Exact model ($R_\eta = 0$) - measurement of biomass every 6 hours with a standard deviation of 1 mM



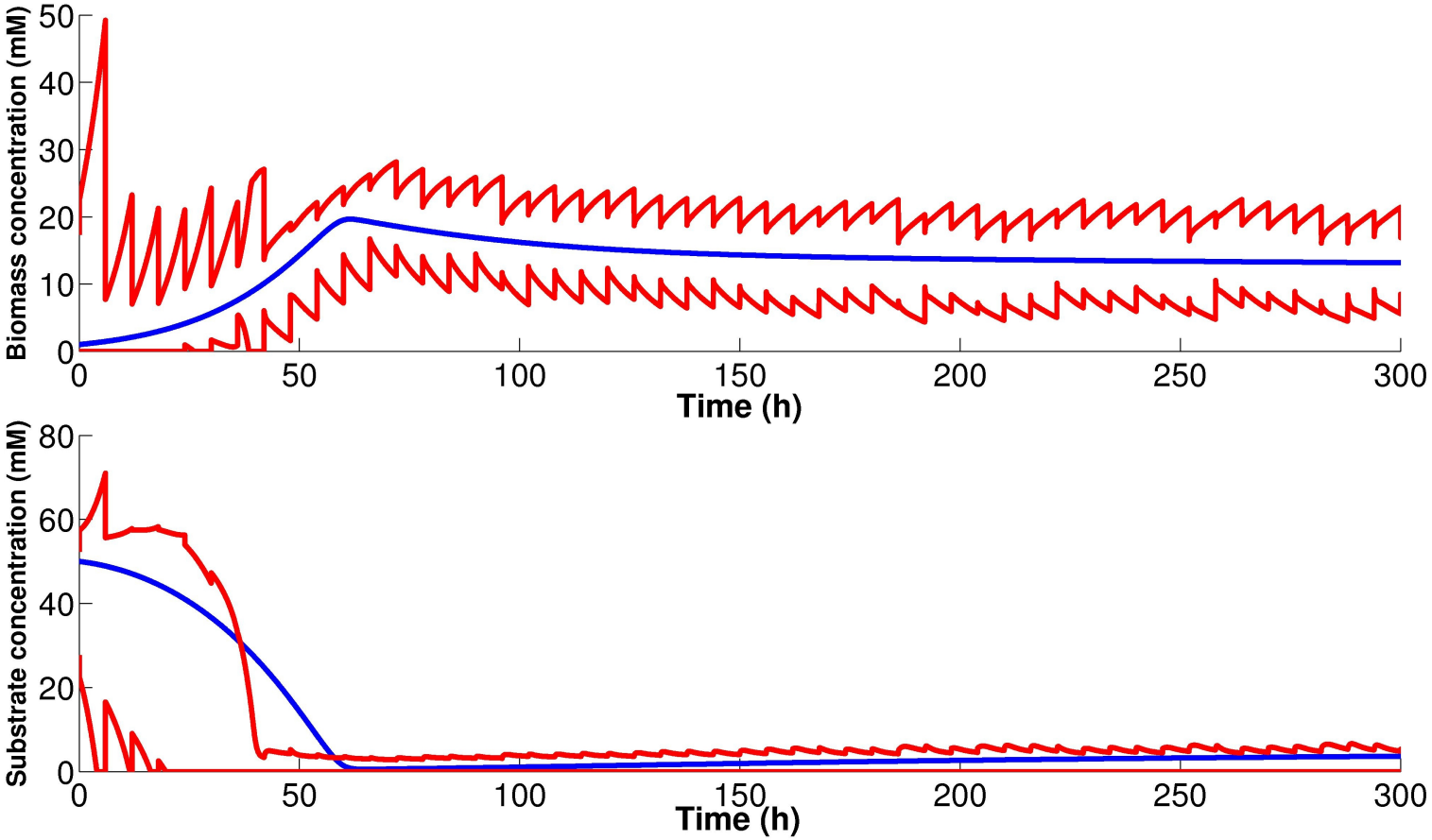
The extended Kalman filter: an example

Kinetic parameter error (but still $R_\eta = 0$) - measurement of biomass every 6 hours with a standard deviation of 1 mM



The extended Kalman filter: an example

Kinetic parameter error (and R_η is adjusted) - measurement of biomass every 6 hours with a standard deviation of 1 mM



Asymptotic observer

A reaction scheme:

$$\sum_{i \in R_m} (-k_{i,m})\xi_i \xrightarrow{\varphi_m} \sum_{j \in P_m} k_{j,m}\xi_j \qquad m \in [1, ..., M]$$

A general macroscopic model:

$$\frac{d\underline{\xi}(t)}{dt} = K\underline{\phi}(\xi, t) - D(t)\underline{\xi}(t) + \underline{F}(t) - \underline{Q}(t)$$

There is a partition $\underline{\xi}^T = [\underline{\xi}_a^T \underline{\xi}_b^T]$ inducing a partition $K^T = [K_a^T K_b^T]$ where $K_a \in \Re^{p \times M}$ is full row rank: $rank(K_a) = rank(K) = p \leq M$

$$\begin{aligned} \dot{\underline{\xi}}_a &= K_a\underline{\phi}(\underline{\xi}_a, \underline{\xi}_b) - D\underline{\xi}_a + \underline{F}_a - \underline{Q}_a \\ \dot{\underline{\xi}}_b &= K_b\underline{\phi}(\underline{\xi}_a, \underline{\xi}_b) - D\underline{\xi}_b + \underline{F}_b - \underline{Q}_b \end{aligned}$$

Asymptotic observer

$$\dot{\underline{\xi}}_a = K_a \underline{\phi}(\underline{\xi}_a, \underline{\xi}_b) - D\underline{\xi}_a + \underline{F}_a - \underline{Q}_a$$
$$\dot{\underline{\xi}}_b = K_b \underline{\phi}(\underline{\xi}_a, \underline{\xi}_b) - D\underline{\xi}_b + \underline{F}_b - \underline{Q}_b$$

If we introduce a state transformation

$$\underline{\xi}_a = \underline{\xi}_a$$
$$\underline{z} = C\underline{\xi}_a + \underline{\xi}_b$$

with the matrix $C \in \Re^{(N-p)\times p}$ solution of $CK_a + K_b = 0$, then

$$\dot{\underline{z}} = -D\underline{z} + C(\underline{F}_a - \underline{Q}_a) + \underline{F}_b - \underline{Q}_b$$
$$\underline{\xi}_b = \underline{z} - C\underline{\xi}_a$$

The state transformation allows the elimination of the - uncertain - kinetics

Asymptotic observer

If $\underline{\xi}_1 \in \Re^L$ denotes the measured concentrations, and $\underline{\xi}_2 \in \Re^{N-L}$, then

$$\underline{z} = A_1 \underline{\xi}_1 + A_2 \underline{\xi}_2$$

$$\begin{aligned} \dot{\hat{\underline{z}}} &= -D\hat{\underline{z}} + C(\underline{F}_a - \underline{Q}_a) + \underline{F}_b - \underline{Q}_b \\ \hat{\underline{\xi}}_2 &= A_2^+(\hat{\underline{z}} - A_1\underline{\xi}_1) \end{aligned}$$

where $A_2^+ = (A_2^T A_2)^{-1} A_2^T \in \Re^{(N-L)\times(N-p)}$ is the left pseudo inverse of A_2.

*For this matrix to exist (**necessary condition of asymptotic observability**) $L \geq p$*

In the particular case where $L = p$ and $\underline{\xi}_1 = \underline{\xi}_a$ then the observer reduces to

$$\begin{aligned} \dot{\hat{\underline{z}}} &= -D\hat{\underline{z}} + C(\underline{F}_1 - \underline{Q}_1) + \underline{F}_2 - \underline{Q}_2 \\ \hat{\underline{\xi}}_2 &= \hat{\underline{z}} - C\underline{\xi}_1 \end{aligned}$$

Asymptotic observer

If the measurements occur at discrete times (always the case), extrapolation is required

Zero-order extrapolation

$$y(t) = y(t_k) \quad t_k \leq t \leq t_{k+1}$$

First-order extrapolation

$$y(t) = y(t_k) + (t - t_k)\frac{y(t_k) - y(t_{k-1})}{t_k - t_{k-1}} \quad t_k \leq t \leq t_{k+1}$$

The estimation error converges asymptoticaly to zero

$$\frac{d\underline{\tilde{\xi}}_2}{dt} = \frac{d(\underline{\hat{\xi}}_2 - \underline{\xi}_2)}{dt} = -D(\underline{\hat{\xi}}_2 - \underline{\xi}_2)$$

provided that D is persistently exiting

$$\exists\, \delta > 0, \beta > 0 \quad : \quad \int_t^{t+\delta t} D(\tau)d\tau \geq \beta$$

Hybrid Kalman-asymptotic observer

Get the best from both solutions: an exponential convergence if the model is accurate, and an asymptotic convergence if the kinetic model is uncertain

The starting point is the EKF

$$\dot{\hat{\underline{\xi}}}(t) = K\underline{\phi}(\hat{\underline{\xi}}(t)) - D(t)\hat{\underline{\xi}}(t) + \underline{u}(t)$$
$$\dot{P}(t) = F(\hat{\underline{\xi}}(t))P(t) + P(t)F^T(\hat{\underline{\xi}}(t)) + R_\eta(t)$$

with $F(\hat{\underline{\xi}}(t)) = K \left.\frac{\partial \underline{\phi}(\underline{\xi})}{\partial \underline{\xi}}\right|_{\underline{\xi}=\hat{\underline{\xi}}(t)} - D(t)I_N$

$$K(t_k) = P^-(t_k)H^T\left[CP^-(t_k)H^T + R_\varepsilon(t_k)\right]^{-1}$$
$$\hat{\underline{\xi}}^+(t_k) = \hat{\underline{\xi}}^-(t_k) + K(t_k)\left(\underline{y}(t_k) - H\hat{\underline{\xi}}^-(t_k)\right)$$
$$P^+(t_k) = P^-(t_k) - K(t_k)HP^-(t_k)$$

Hybrid Kalman-asymptotic observer

We assume that the measurements $\underline{y}(t_k) = H\underline{\xi}(t_k) + \underline{\varepsilon}(t_k) = \underline{\xi}_1(t_k) + \underline{\varepsilon}(t_k)$ induce a partition $K^T = \begin{bmatrix} K_1^T & K_2^T \end{bmatrix}$ where $rank(K_1) = L = M$
A state transformation is introduced using the C matrix wich is solution of $CK_1 + K_2 = 0$

$$\begin{aligned} \underline{z}_1 &= \underline{\xi}_1 \\ \underline{z}_2 &= \underline{\xi}_2 + (1-\delta)C\underline{\xi}_1 \end{aligned}$$

Two extreme situations can be defined:

- $\delta = 1$: just a simple change of notation (plain EKF)
- $\delta = 0$: asymptotic observer
- $0 \leq \delta \leq 1$: intermediate situations to be specified - δ will play the role of a confidence level in the model !

Hybrid Kalman-asymptotic observer

This state transformation can be used to express the Kalman filter in an alternative form

$$
\begin{aligned}
\frac{d\hat{\underline{z}}_1}{dt} &= K_1\underline{\phi}(\hat{\underline{z}}_1, \hat{\underline{\xi}}_2) - D\hat{\underline{z}}_1 + \underline{u}_1 \\
\frac{d\hat{\underline{z}}_1}{dt} &= \delta K_2\phi(\hat{\underline{z}}_1, \hat{\underline{\xi}}_2) - D\hat{\underline{z}}_2 + \underline{u}_2 + (1-\delta)C\underline{u}_1 \\
\hat{\underline{\xi}}_1 &= \hat{\underline{z}}_1 \\
\hat{\underline{\xi}}_2 &= \hat{\underline{z}}_2 - (1-\delta)C\hat{\underline{\xi}}_1 \\
\frac{dP}{dt} &= F(\hat{\underline{z}})P + PF^T(\hat{\underline{z}}) + R_\eta
\end{aligned}
$$

with

$$
\begin{aligned}
\underline{z}^T &= \begin{bmatrix} \underline{z}_1^T & \underline{z}_2^T \end{bmatrix} \\
\underline{f}(\underline{z}) &= \begin{bmatrix} K_1\underline{\phi}(\underline{z}_1, \underline{\xi}_2(\underline{z}_1, \underline{z}_2)) - D\underline{z}_1 \\ \delta K_2\underline{\phi}(\underline{z}_1, \underline{\xi}_2(\underline{z}_1, \underline{z}_2)) - D\underline{z}_2 \end{bmatrix} \\
F(\hat{\underline{z}}) &= \left.\frac{\partial \underline{f}(\underline{z})}{\partial \underline{z}}\right|_{\underline{z}=\hat{\underline{z}}(t)}
\end{aligned}
$$

Hybrid Kalman-asymptotic observer

This state transformation can be used to express the Kalman filter in an alternative form

$$
\begin{aligned}
&K(t_k) = P(t_k^-)H^T\left[HP(t_k^-)H^T + R_\varepsilon(t_k)\right]^{-1}\\
&\underline{\hat{z}}^+(t_k) = \hat{z}^-(t_k) + K(t_k)\left(\underline{y}(t_k) - H\underline{\hat{z}}^-(t_k)\right) = \underline{\hat{z}}^-(t_k) + K(t_k)\left(\underline{y}(t_k) - \underline{\hat{\xi}}_1^-(t_k)\right)\\
&P^+(t_k) = P^-(t_k) - K(t_k)HP^-(t_k)
\end{aligned}
$$

At this stage, more influencial modifications are introduced:

- an extension of the state $\underline{\hat{z}}^T = \begin{bmatrix} \underline{\hat{z}}_1^T & \underline{\hat{z}}_2^T & \delta \end{bmatrix}$ with an additional state equation $\frac{d\delta}{dt} = 0$ and a new measurement matrix $\mathrm{H}_\delta = \begin{bmatrix} H & O_{L,1} \end{bmatrix}$
- a weighted sum of the estimation $\hat{z}_1$ and the measurement y to build the estimation of ξ_1: $\underline{\hat{\xi}}_1 = \delta\underline{\hat{z}}_1 + (1-\delta)\underline{y}$
- an extrapolation of the discrete-time measurements (as in the asymptotic observer)

Hybrid Kalman-asymptotic observer

Summing up all these modifications leads to the prediction equations

$$
\begin{aligned}
\frac{d\hat{\underline{z}}_1}{dt} &= K_1\underline{\phi}(\hat{\underline{z}}_1, \hat{\underline{\xi}}_2) - D\hat{\underline{z}}_1 + \underline{u}_1 \\
\frac{d\hat{\underline{z}}_2}{dt} &= \delta K_2\underline{\phi}(\hat{\underline{z}}_1, \hat{\underline{\xi}}_2) - D\hat{\underline{z}}_2 + \underline{u}_2 + (1-\delta)C\underline{u}_1 \\
\frac{d\delta}{dt} &= 0 \\
\frac{dP}{dt} &= F(\hat{\underline{z}})P + PF^T(\hat{\underline{z}}) + R_\eta
\end{aligned}
$$

$$
\begin{aligned}
\hat{\underline{\xi}}_1 &= \delta\hat{\underline{z}}_1 + (1-\delta)\underline{y} \\
\hat{\underline{\xi}}_2 &= \hat{\underline{z}}_2 - (1-\delta)C\hat{\underline{\xi}}_1
\end{aligned}
$$

with

$$
\underline{z}^T = \begin{bmatrix} \underline{z}_1^T & \underline{z}_2^T & \delta \end{bmatrix}
$$

$$
\underline{f}(\underline{z}) = \begin{bmatrix} K_1\underline{\phi}(\underline{z}_1, \underline{\xi}_2(\underline{z}_1, \underline{z}_2, \delta)) - D\underline{z}_1 \\ \delta K_2\underline{\phi}(\underline{z}_1, \underline{\xi}_2(\underline{z}_1, \underline{z}_2, \delta)) - D\underline{z}_2 + (1-\delta)C\underline{u}_1 \\ 0 \end{bmatrix}
$$

$$
F(\hat{\underline{z}}) = \left.\frac{\partial \underline{f}(\underline{z})}{\partial \underline{z}}\right|_{\underline{z}=\hat{\underline{z}}(t)}
$$

Hybrid Kalman-asymptotic observer

... and the correction equations

$$
\begin{aligned}
K(t_k) &= P^-(t_k)H_\delta^T\left[H_\delta P^-(t_k)H_\delta^T + R_\varepsilon(t_k)\right]^{-1} \\
\hat{z}^+(t_k) &= \hat{z}^-(t_k) + K(t_k)\left(y(t_k) - \hat{\xi}_1^-(t_k)\right) \\
P^+(t_k) &= P^-(t_k) - K(t_k)H_\delta P^-(t_k)
\end{aligned}
$$

with positivity constraints on the concentration estimates $\hat{\xi}_1(t)$ and $\hat{\xi}_2(t)$ such that

$$
\begin{aligned}
&\hat{z}_1^+(t_k) \geq 0 \\
&\hat{z}_2^+(t_k) \geq (1-\delta)C\hat{\xi}_1^+(t_k) \\
&0 \leq \delta \leq 1
\end{aligned}
$$

Hybrid Kalman-asymptotic observer

In practice ...

- $0 \leq \delta \leq 1$ can be seen as a degree of confidence in the kinetic model
- $\delta = 1$ is used as initialization to give a chance to the model ! (which means that we start with the EKF)
- If δ ends up in 0, then the observer behaves as the asymptotic observer
- the covariance matrix of the estimation errors

$$P(t_0) = \begin{bmatrix} P_z I_{N,N} & O_{N,1} \\ O_{1,N} & P_\delta \end{bmatrix}$$
$$(\mathrm{e.g.} P_z = 10^6 \mathrm{and} P_\delta = 0.1)$$

can be tuned to adjust the sensitivity to the kinetic model discrepancies

Full horizon observer

Consider a nonlinear dynamic system with discrete-time measurements

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t)) \qquad \underline{x}(t_0) = \underline{x}_0$$
$$\underline{y}(t_k) = \underline{h}(\underline{x}(t_k)) + \underline{\varepsilon}(t_k)$$

The solution $\underline{g}(t, \underline{u}, \underline{x}_0)$ can be obtained by numerical integration.

The idea of the **full-horizon observer** is to compute the most likely initial conditions on the absis of all the measurement data collected so far

$$\hat{\underline{x}}_{0/k} = \arg\min_{\underline{x}_0} \sum_{i=1}^{k} [\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0)]^T R_\varepsilon^{-1}(t_i) [\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0)]$$

and to predict a state estimate trajectory between two measurement times by simulating the model equations from these estimated initial conditions

$$\hat{\underline{x}}(t) = g(t, \underline{u}(t), \hat{\underline{x}}_{0/k}) \qquad t_k \leq t \leq t_{k+1}$$

Full horizon observer

The algorithm can be summarized as follows:

- **Initialization**: numerical solution of the model equations starting from an initial condition built using the values of the measured states, and an initial guess of the nonmeasured states
- **Correction**: at each new measurement time, identification of the most likely initial conditions, on the basis of all the measurement information collected so far
- **Prediction**: between two measurement times, simulation of the model equations using the latest initial condition estimates

Full horizon observer

Algorithm properties

- the nonlinear model is used without any approximation (in contrast to the extended Kalman filter which requires an on-line linearization)
- the drawback is the use of a nonlinear optimization procedure (with its associated problems)
- The existence of a solution to the minimimization problem requires that the number of available measurements is larger than the number of states $km \geq n$

Full and receding horizon observers

Possible extensions

- constraints on the initial conditions (e.g. positivity constraints)
- use of a **receding horizon** of length L (so as to forget the past)

$$\underline{\hat{x}}_{0/k} = \arg\min_{\underline{x}_0} \sum_{i=k-L+1}^{k} [\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0)]^T R_\varepsilon^{-1}(t_i)[\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0)]$$

with $k \geq L$ and $mL \geq n$

- jointly estimate a few unknown parameters

$$[\underline{\hat{x}}_{0/k}, \underline{\hat{\theta}}_k] = \arg\min_{\underline{x}_0, \underline{\theta}} \sum_{i=1}^{k} [\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0, \underline{\theta})]^T R_\varepsilon^{-1}(t_i)[\underline{y}(t_i) - \underline{h}(g(t_i, \underline{u}(t_i), \underline{x}_0, \underline{\theta})]$$

with $mk \geq n + p$