Some Topics on Nonlinear Moving-Horizon Observers

Mazen Alamir

French National Research Center
GIPSA-lab. Control Systems Dept.

SYSCO-Team

Email: mazen.alamir@inpg.fr

Summer school, September 2007, Grenoble.

For an updated version

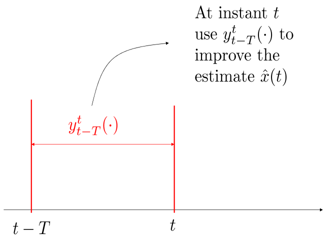
www.lag.ensieg.inpg.fr/alamir/summer_school_no.pdf

Moving-Horizon ?

- $J = \int_{t-T}^{t}[\text{Output pred. err. }(\tau)]d\tau$
- internal state:
  Estimate $z(t)$ of $x(t-T)$
- Update $z(t)$:

$$z^{+} = \varphi(z, J)$$

- The *cost* $J(t)$ is explicitly/implicitly used in the innovation process (correction)

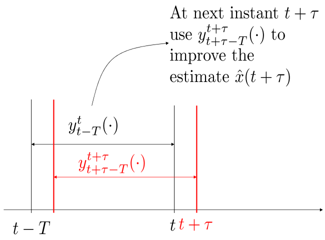


Moving-Horizon ?

- $J = \int_{t-T}^{t} [\text{Output pred. err. } (\tau)] d\tau$
- internal state:
  Estimate $z(t)$ of $x(t-T)$
- Update $z(t)$:

$$z^{+} = \varphi(z, J)$$

- The *cost* $J(t)$ is explicitly/implicitly used in the innovation process (correction)



Overview

State estimation: An optimization problem

Overview

State estimation: An optimization problem

(Dynamic & non convex)

Overview

State estimation: An optimization problem

(Dynamic & non convex)

Local minima

Computation time

Overview

State estimation: An optimization problem

(Dynamic & non convex)

Local minima

Computation time

Singularities avoidance heuristics

Overview

State estimation: An optimization problem

(Dynamic & non convex)

Local minima

Computation time

Singularities avoidance heuristics

Real-time implementation

Ideal Continuous Case (Gradient-based)

Discrete general setting

Some definitions and notations (1): The System

Uncertainty & noise free system

$$\begin{aligned} x(t) &= X(t, t_0, x_0) \\ y(t) &= h(t, x(t)) \end{aligned}$$

Uncertain and noisy system

$$\begin{aligned} x(t) &= X(t, t_0, x_0, w_{t_0}^t) \\ y(t) &= h(t, x(t)) + v(t) \end{aligned}$$

Constraints

- $x(t) \in \mathbb{X}(t) \subset \mathbb{R}^n$
- $w(t) \in \mathbb{W}(t) \subset \mathbb{R}^{n_w}$ Uncertainties/Disturbances.
- $v(t) \in \mathbb{V}(t) \subset \mathbb{R}^{n_y}$ Measurement noise

Definitions and notations (2): Measurements-compatible configurations

Consider

- Time interval $[t-T, t]$
- Measurement profile y_{t-T}^{t}
- $(\xi, \mathbf{w}) \in \mathbb{X}(t-T) \times [\mathbb{R}^{n_w}]^{[t-T,t]}$

Definitions and notations (2): Measurements-compatible configurations

Consider

- Time interval $[t-T,t]$
- Measurement profile y_{t-T}^{t}
- $(\xi,\mathbf{w}) \in \mathbb{X}(t-T) \times [\mathbb{R}^{n_w}]^{[t-T,t]}$

$(\xi,\mathbf{w})$ is (y_{t-T}^{t})-compatible

if for all $\sigma \in [t-T,t]$:

1. $w(\sigma) \in \mathbb{W}(\sigma)$,
2. $X(\sigma,t-T,\xi,\mathbf{w}) \in \mathbb{X}(\sigma)$,
3. $y_{t-T}^{t}(\sigma) - Y(\sigma,t-T,\xi,\mathbf{w}) \in \mathbb{V}(\sigma)$.

gipsa-lab

Definitions and notations (2): Measurements-compatible configurations

Consider

- Time interval $[t-T, t]$
- Measurement profile y_{t-T}^{t}
- $(\xi, \mathbf{w}) \in \mathbb{X}(t-T) \times [\mathbb{R}^{n_w}]^{[t-T,t]}$

$(\xi, \mathbf{w})$ is (y_{t-T}^{t})-compatible

if for all $\sigma \in [t-T, t]$:

1. $w(\sigma) \in \mathbb{W}(\sigma)$,
2. $X(\sigma, t-T, \xi, \mathbf{w}) \in \mathbb{X}(\sigma)$,
3. $y_{t-T}^{t}(\sigma) - Y(\sigma, t-T, \xi, \mathbf{w}) \in \mathbb{V}(\sigma)$.

Notation

$$(\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^{t})$$

gipsa-lab

Definitions and notations (2): Measurements-compatible configurations

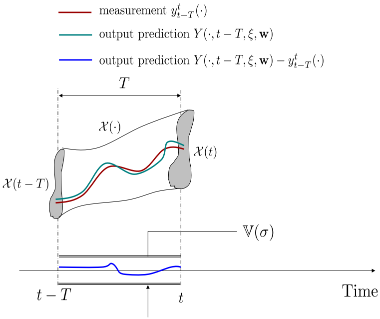
Consider

- Time interval $[t-T,t]$
- Measurement profile y_{t-T}^t
- $(\xi,\mathbf{w}) \in \mathbb{X}(t-T)\times[\mathbb{R}^{n_w}]^{[t-T,t]}$

$(\xi,\mathbf{w})$ is (y_{t-T}^t)-compatible

if for all $\sigma\in[t-T,t]$:

1. $w(\sigma)\in\mathbb{W}(\sigma)$,
2. $X(\sigma,t-T,\xi,\mathbf{w})\in\mathbb{X}(\sigma)$,
3. $y_{t-T}^t(\sigma)-Y(\sigma,t-T,\xi,\mathbf{w})\in\mathbb{V}(\sigma)$.



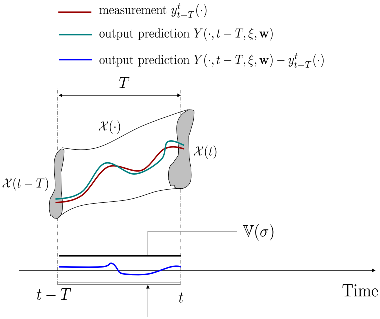
Notation

$$(\xi,\mathbf{w})\in\mathbb{C}(t,y_{t-T}^t)$$

Definitions and notations (2): Measurements-compatible configurations

Consider

- Time interval $[t-T, t]$
- Measurement profile y_{t-T}^{t}
- $(\xi, \mathbf{w}) \in \mathbb{X}(t-T) \times [\mathbb{R}^{n_w}]^{[t-T,t]}$



$(\xi, \mathbf{w})$ is (y_{t-T}^{t})-compatible

if for all $\sigma \in [t-T, t]$:

1. $w(\sigma) \in \mathbb{W}(\sigma)$,
2. $X(\sigma, t-T, \xi, \mathbf{w}) \in \mathbb{X}(\sigma)$,
3. $y_{t-T}^{t}(\sigma) - Y(\sigma, t-T, \xi, \mathbf{w}) \in \mathbb{V}(\sigma)$.

$(\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^{t})$ if the corresponding trajectory

1. meets the constraints
2. explains the measurements

Notation

$$(\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^{t})$$

The finite horizon observation problem

The finite horizon observation problem

The finite Choose $T > 0$ and use at each t, the available information:

1. System equations
2. Past measurements y_{t-T}^{t},
3. Constraints and
4. *Some additional exogenous knowledge*.

in order to produce an estimation $\hat{x}(t)$ of the current state $x(t)$. ♡

The finite horizon observation problem

The finite horizon observation problem

The finite Choose $T > 0$ and use at each t, the available information:

1. System equations
2. Past measurements y_{t-T}^{t},
3. Constraints and
4. *Some additional exogenous knowledge*.

in order to produce an estimation $\hat{x}(t)$ of the current state $x(t)$. ♡

$$\text{Find } (\xi, \mathbf{w}) \quad \longrightarrow \quad \hat{x}(t) = X(t, t-T, \xi, \mathbf{w})$$

The finite horizon observation problem

The finite horizon observation problem

The finite Choose $T > 0$ and use at each t, the available information:

1. System equations
2. Past measurements y_{t-T}^{t},
3. Constraints and
4. *Some additional exogenous knowledge*.

in order to produce an estimation $\hat{x}(t)$ of the current state $x(t)$. ♡

$$\text{Find } (\xi, \mathbf{w}) \quad \longrightarrow \quad \hat{x}(t) = X(t, t-T, \xi, \mathbf{w})$$

The set of candidate estimates $\hat{x}(t)$:

$$\Omega_t = \Big\{ X(t, t-T, \xi, \mathbf{w}) \quad | \quad (\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^{t}) \Big\}.$$

The need for additional knowledge

$$\Omega_t = \Big\{X(t, t-T, \xi, \mathbf{w}) \quad | \quad (\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^t)\Big\}.$$

- Either $\Omega_t = \{x(t)\}$, for instance because
  - $\mathbb{W} = \{0\}$, $\mathbb{V} = \{0\}$ and
  - The system has no indistinguishable states

$$\int_{t-T}^{t} \|Y(\sigma, t-T, x^{(1)}) - Y(\sigma, t-T, x^{(2)})\|^2 d\sigma \geq \alpha(\|x^{(1)} - x^{(2)}\|)$$

for all $t \geq 0$ and all $(x^{(1)}, x^{(2)}) \in \mathbb{X}(t-T) \times \mathbb{X}(t-T)$.

The need for additional knowledge

$$\Omega_t = \Big\{ X(t, t-T, \xi, \mathbf{w}) \quad | \quad (\xi, \mathbf{w}) \in \mathbb{C}(t, y_{t-T}^{t}) \Big\}.$$

- Or $\Omega_t \neq \{x(t)\}$, and a *selection* must be made by solving

$$P(t) \quad : \quad \min_{(\xi,\mathbf{w})\in\Omega_t} J(t, \xi, \mathbf{w}) \quad \rightarrow (\hat{\xi}(t), \hat{\mathbf{w}}(t))$$

estimation: $\hat{x}(t) = X(t, t-T, \hat{\xi}(t), \hat{\mathbf{w}}(t))$

Temporal Parametrization (1)

Solve $P(t) \quad : \quad \min_{(\xi,\mathbf{w})\in\mathbb{C}(t)} J(t,\xi,\mathbf{w})$

In many textbooks, the following parametrization is suggested for **w**:

$$p_w := \{\mathbf{w}(k\tau)\}_{k=k_0}^{k_0+N-1} \in \mathbb{W}(k_0)\times\cdots\times\mathbb{W}(k_0+N-1) \subset \mathbb{R}^{n_w\cdot N}$$

- Decision variable (ξ, p_w) of dimension $n + N\cdot n_w$
- Too rich spectral content *increasing* uselessly Ω_t
- High sensitivity to the knowledge of $\mathbb{W}(\cdot)$.

Unrealistically too many possible interpretations of the measurements

Temporal parametrization (2)

Solve $P(t) \quad : \quad \min_{(\xi,\mathbf{w})\in\mathbb{C}(t)} J(t,\xi,\mathbf{w})$

Use a reduced dimensional parametrization

$$w(t) = \mathcal{W}(t,p_w) \quad ; \quad p_w \in \mathbb{P}.$$

Solve $P(t) \quad : \quad \min_{(\xi,p_w)\in\mathbb{C}(t)} J(t,\xi,\mathcal{W}(\cdot,p_w)) =: J(t,\xi,p_w) \quad \rightarrow (\hat{\xi}(t),\hat{p}_w(t))$

$$\hat{x}(t) = X(t,t-T,\hat{\xi}(t),\mathcal{W}(\cdot,\hat{p}_w(t)))$$

- $\bar{x} := (x^T,p_w^T)^T \in \mathbb{R}^n \times \mathbb{R}^{n_p}$
- $\dot{p}_w = 0$

New uncertainty-free
extended state estimation problem.

gipsa-lab

Analytic vs optimization based observer

Analytic observers

$$\begin{array}{ll}\text{(System)} & \dot{x} = f(x)\ ;\ y = h(x) \\ \text{(Observ)} & \dot{\hat{x}} = f(\hat{x}) + K(\hat{x}, y)\end{array}$$

Try to show asymptotic convergence of $e = x - \hat{x}$ governed by

$$\begin{aligned}\dot{x} &= f(x) \\ \dot{e} &= f(x) - f(x-e) - K(x-e, h(x))\end{aligned}$$

Very Hard Task

- Need for structural properties
- Coordinate transformation
- Constructive assumptions
- Observability $\neq$ Existence of observer

Analytic vs optimization based observer

Analytic observers

$$\begin{array}{ll} \text{(System)} & \dot{x} = f(x) \; ; \; y = h(x) \\ \text{(Observ)} & \dot{\hat{x}} = f(\hat{x}) + K(\hat{x}, y) \end{array}$$

Try to show asymptotic convergence of $e = x - \hat{x}$ governed by

$$\begin{aligned} \dot{x} &= f(x) \\ \dot{e} &= f(x) - f(x-e) - K(x-e, h(x)) \end{aligned}$$

Very Hard Task

- Need for structural properties
- Coordinate transformation
- Constructive assumptions
- Observability $\neq$ Existence of observer

optimization based observers

Rely on the implication

$$\left\{ J(t,\xi) \to 0 \right\} \Rightarrow \left\{ \underbrace{X(t, t-T, \xi)}_{\hat{x}(t)} \to x(t) \right\}$$

+ No need to study the dynamic of e
+ No need for structural assumptions
+ Observability $\Leftrightarrow$ Observer
+ Handling constraints on the state

Potential problems

- Global convergence ?
- Computation time ?

Forthcoming issues

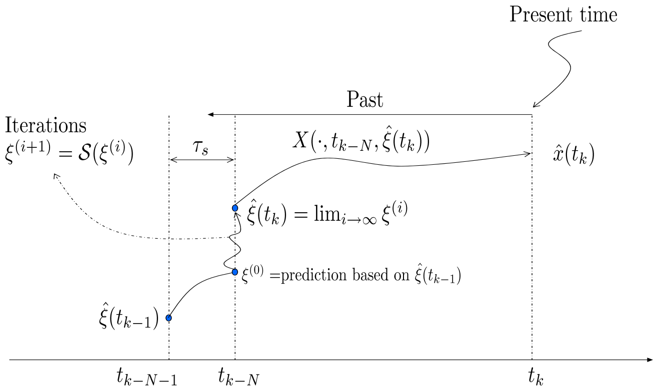
Global convergence ?

- No generic and definitive solution . . . !
- Heuristics for singularities avoidance

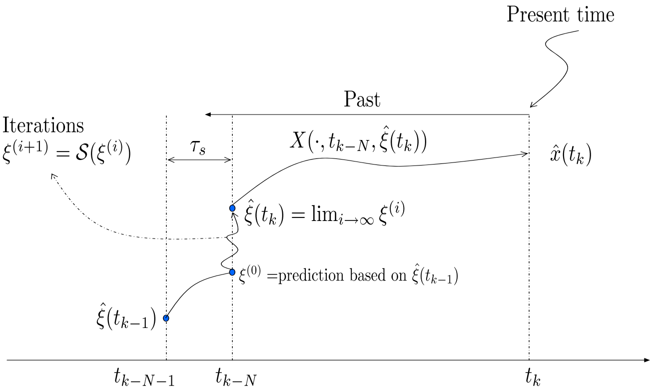
Computation time ?

- Differential form of optimization based observer
- Real-Time iterations / Optimal choice of updating period

Ideal discrete-time estimation scheme

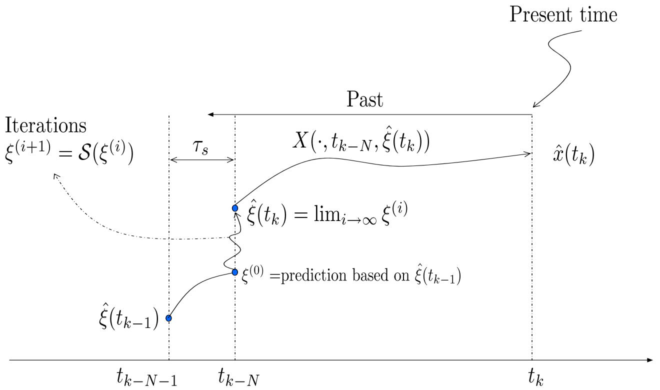


Ideal discrete-time estimation scheme



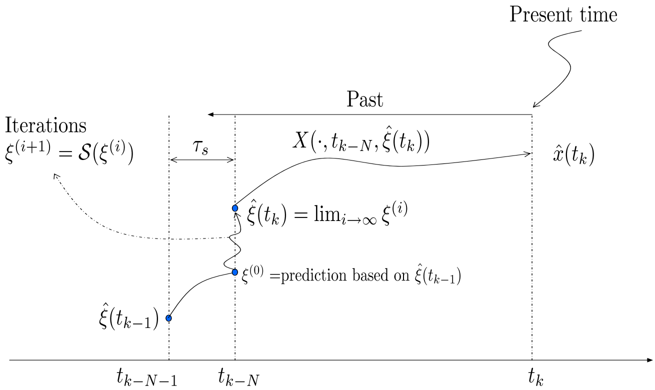
$$\hat{x}(t_k) = X(t_k, t_{k-N}, \hat{\xi}(t_k))$$

Ideal discrete-time estimation scheme



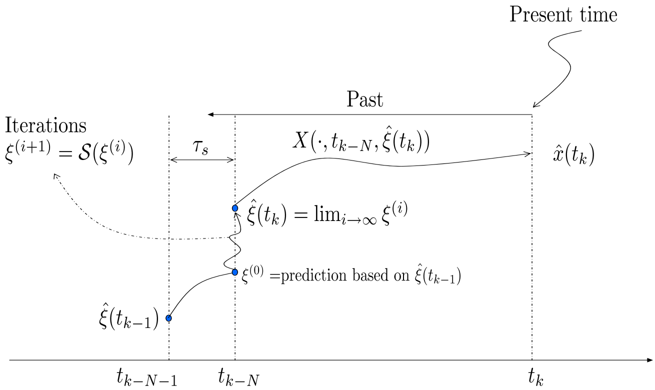
$$\hat{\xi}(t_k) = \underbrace{arg \min_{\xi\in\mathbb{X}(t_{k-N})} \Big[J(t_k,\xi)\Big]}_{\text{Initial Guess } \xi^{(0)}} := \sum_{i=k-N}^{k} \|y(t_i) - Y(t_i, t_{k-N}, \xi)\|_{Q_i(k)}^{2}$$

Ideal discrete-time estimation scheme



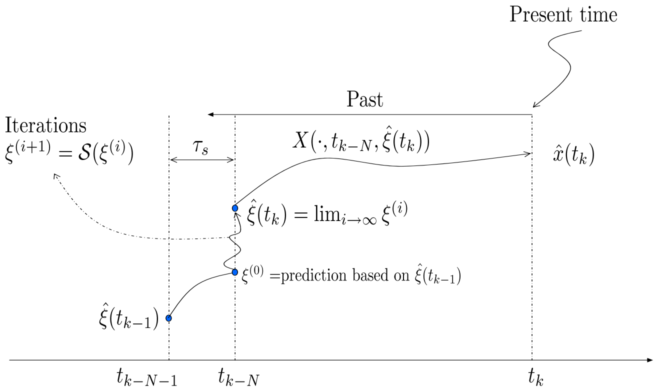
$$J^*(t_k,\xi) := \|\xi - \xi^{(0)}\|_{Q_0} + \sum_{i=k-N}^{k} \|y(t_i) - Y(t_i, t_{k-N}, \xi)\|_{Q_i(k)}^{2}$$

Ideal discrete-time estimation scheme



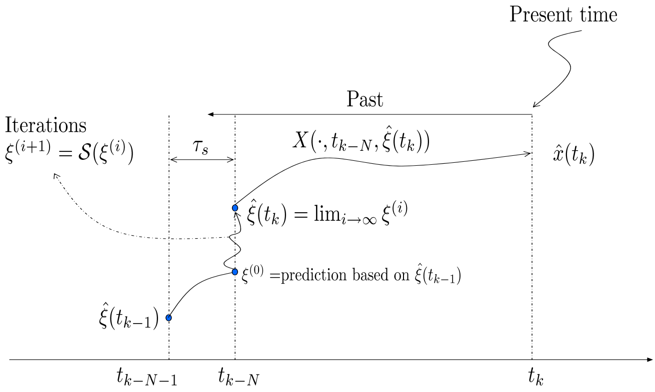
$$\xi^{(0)} = X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$$

Ideal discrete-time estimation scheme



In practice: $\hat{\xi}(t_k) = \xi^{(N_{max})} = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$

Ideal discrete-time estimation scheme



In practice: $\hat{\xi}(t_k) = \xi^{(N_{max})} = \bar{\mathcal{S}}^{N_{max}}(\hat{\xi}(t_{k-1}), t_k, y_{t_{k-N}}^{t_k})$

State estimation is a very particular optimization problem

Particular feature of the state estimation related optimization problem

$x(t_{k-N})$ is the unique global minimum of ALL the optimization problems:

$$\hat{\xi}(t_k) = arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi) \Big] := \sum_{i=k-N}^{k} \| y(t_i) - Y(t_i, t_{k-N}, \xi) \|_{Q_i(k)}{}^2$$

that may be obtained by changing the positive definite weighting matrices $Q_i(k)$.

State estimation is a very particular optimization problem

Particular feature of the state estimation related optimization problem

$x(t_{k-N})$ is the unique global minimum of ALL the optimization problems:

$$\hat{\xi}(t_k) = arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi)\Big] := \sum_{i=k-N}^{k} \|y(t_i) - Y(t_i, t_{k-N}, \xi)\|_{Q_i(k)}{}^2$$

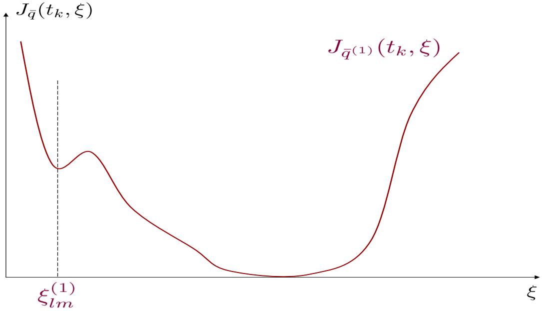
that may be obtained by changing the positive definite weighting matrices $Q_i(k)$.

Let us take the following family subset of weighting choices:

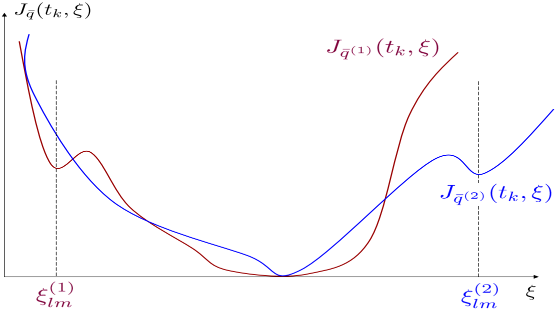
$$Q_i(k) = \gamma^{k-i} \cdot q_i \cdot \mathbb{I}_{n_y} \quad \text{s.t} \quad q_i > 0 \quad \text{and} \quad \sum_i q_i = 1 \tag{1}$$

- $\gamma \in]0, 1]$ Forgetting factor
- $\mathbb{I}_{n_y}$ identity matrix in $\mathbb{R}^{n_y \times n_y}$
- Notations: $\bar{q} = \left(q_1, q_2, \dots, q_N\right)^T$, $J_{\bar{q}}(t_k, \xi)$, $\mathcal{S}_{\bar{q}}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$

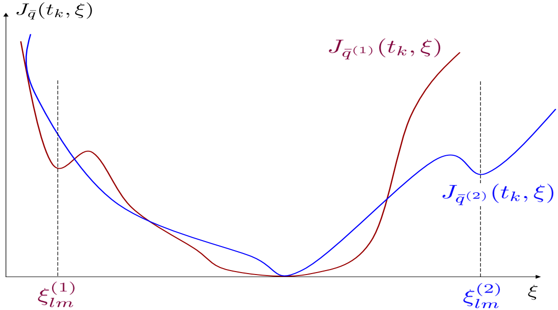
Crossing singularity by swapping the weighting vectors



Crossing singularity by swapping the weighting vectors

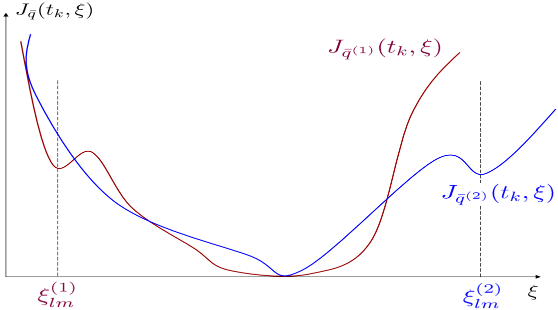


Crossing singularity by swapping the weighting vectors



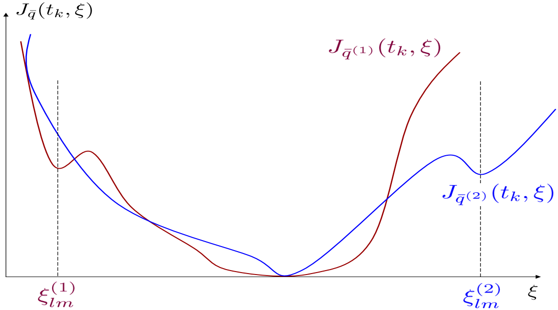
- Generally, $J_{\bar{q}^{(1)}}(t_k, \cdot)$ and $J_{\bar{q}^{(2)}}(t_k, \cdot)$ have *no reasons to share the same* LOCAL *minima*

Crossing singularity by swapping the weighting vectors



- Generally, $J_{\bar{q}^{(1)}}(t_k, \cdot)$ and $J_{\bar{q}^{(2)}}(t_k, \cdot)$ have *no reasons to share the same* LOCAL *minima*
- They DO share the same global minimum $x(t_{k-N})$

Crossing singularity by swapping the weighting vectors



- Generally, $J_{\bar{q}^{(1)}}(t_k, \cdot)$ and $J_{\bar{q}^{(2)}}(t_k, \cdot)$ have *no reasons to share the same* LOCAL *minima*
- They DO share the same global minimum $x(t_{k-N})$
- Think about an infinite number of $J_{\bar{q}}(t_k, \cdot)$ (randomly generated)

This suggests

The crossing singularities heuristic

$$\bar{q} \leftarrow \frac{1}{n_y}\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}$$

$$\hat{\xi}(t_k) \leftarrow X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$$

for $(i = 1 : N_{\text{trials}})$

$$\hat{\xi}(t_k) \leftarrow \mathcal{S}_{\bar{q}}^{N_{max}}(\hat{\xi}(t_k), t_k, y_{t_{k-N}}^{t_k})$$

Generate randomly new admissible $\bar{q}$

end

$$\hat{x}(t_k) \leftarrow X(t_k, t_{k-N}, \hat{\xi}(t_k))$$

This suggests

The crossing singularities heuristic

$$\bar{q} \leftarrow \frac{1}{n_y}\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}$$

$$\hat{\xi}(t_k) \leftarrow X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$$

for $(i = 1 : N_{\text{trials}})$

$$\hat{\xi}(t_k) \leftarrow \mathcal{S}_{\bar{q}}^{N_{max}}(\hat{\xi}(t_k), t_k, y_{t_{k-N}}^{t_k})$$

Generate randomly new admissible $\bar{q}$

end

$$\hat{x}(t_k) \leftarrow X(t_k, t_{k-N}, \hat{\xi}(t_k))$$

- This is not a multiple initial guess trials
- Implementation constraint

$$N_{trial} \times N_{max} \times \tau_{iter} \leq \tau_s$$

- The Trade-off is problem dependent

Example: State estimation of terpolymerization reactors

- Produce polymer from multi-monomer
- Controlling the final properties need the state to be estimated
- State: Polymer composition $\leftrightarrow$ Monomers concetrations
- Complex equations
- Unknown dynamics
- High gain observers need tremendous simplifications to give rather poor performance

Coll. Nida Sheibat-Othman & Sami Othman (LAGEP,Lyon)

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad i = 1, 2, 3$$

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad\qquad i = 1, 2, 3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad\qquad i = 1, 2, 3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

where

$$P_1^P = \frac{\alpha}{\alpha+\beta+\gamma} \quad ; \quad P_2^P = \frac{\beta}{\alpha+\beta+\gamma} \quad ; \quad P_3^P = 1 - P_1^P - P_2^P$$

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad i = 1, 2, 3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

where

$$P_1^P = \frac{\alpha}{\alpha+\beta+\gamma} \quad ; \quad P_2^P = \frac{\beta}{\alpha+\beta+\gamma} \quad ; \quad P_3^P = 1 - P_1^P - P_2^P$$

in which

$$\begin{aligned}
\alpha &= [M_1^P](k_{p21}k_{p31}[M_1^P] + k_{p21}k_{p32}[M_2^P] + k_{p31}k_{p23}[M_3^P]) \\
\beta &= [M_2^P](k_{p12}k_{p31}[M_1^P] + k_{p12}k_{p32}[M_2^P] + k_{p13}k_{p32}[M_3^P]) \\
\gamma &= [M_3^P](k_{p13}k_{p21}[M_1^P] + k_{p21}k_{p23}[M_2^P] + k_{p13}k_{p23}[M_3^P])
\end{aligned}$$

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad i = 1, 2, 3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

where

$$P_1^P = \frac{\alpha}{\alpha+\beta+\gamma} \quad ; \quad P_2^P = \frac{\beta}{\alpha+\beta+\gamma} \quad ; \quad P_3^P = 1 - P_1^P - P_2^P$$

The $[M_i^P]$ depend in the state according to:

$$[M_i^P] = \begin{cases} \dfrac{(1-\phi_p^p)N_i}{\sum_j \dfrac{N_j MW_j}{\rho_j}}, & \text{(Phase II)} \\ \dfrac{N_i}{\sum_j MW_j(\dfrac{N_j^T - N_j}{\rho_{j,h}} + \dfrac{N_j}{\rho_j})} & \text{(Phase III)} \end{cases}$$

gipsa-lab

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad\qquad i = 1, 2, 3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

- μ plays a crucial role
- The dynamic of μ is unknown

Example: terpolymerization reactors (The mathematical model)

$$\dot{N}_i = Q_i - R_{Pi} \qquad\qquad i = 1,2,3$$

$$R_{P_i} = \mu[M_i^P](k_{p1i}P_1^P + k_{p2i}P_2^P + k_{p3i}P_3^P)$$

- μ plays a crucial role
- The dynamic of μ is unknown

- Measurement

The overall monomer conversion measured by calorimetry:

$$y = \frac{\sum_{i=1}^{3} MW_i(N_i^T - N_i)}{\sum_{j=1}^{3} MW_j N_j^T}$$

Example: State reconstruction of terpolymerization reactors (Validation)

1. Simulation results
2. Experimental results

Example: State reconstruction of terpolymerization reactors (Validation)

Simulation results

$$\begin{aligned} \dot{N} &= \begin{pmatrix} 1+d_1 & 0 & 0 \\ 0 & 1+d_2 & 0 \\ 0 & 0 & 1+d_3 \end{pmatrix} \cdot f(x,u) \\ \dot{\mu} &= 0 \\ y &= (1+\nu)\cdot h(x) \end{aligned}$$

- The state $x := \begin{pmatrix} N_1 & N_2 & N_3 & \mu \end{pmatrix} \in \mathbb{R}^4_+$
- The uncertainties

$$\begin{aligned} d_i(k) &= d_{max} \cdot r_i(k) \\ \nu(k) &= \nu_{max} \cdot r_\nu(k) \end{aligned}$$

- r_i and ν randomly chosen in $[-1,+1]$

Simulation results

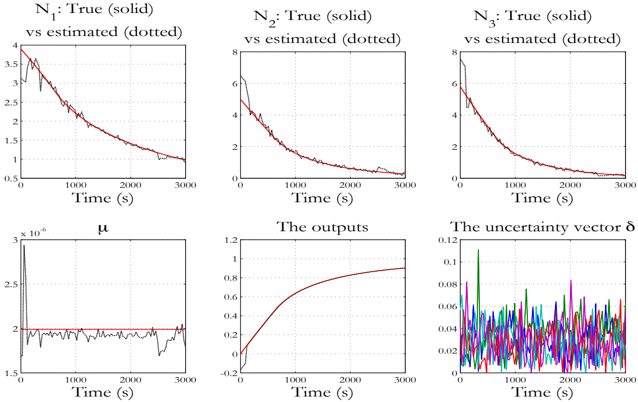


Figure: *Observer behavior under model uncertainty given by (2)-(2) with* $d_{max} = 10\%$ *and no measurement noise* ($\nu_{max} = 0$). *The observation horizon is* $N = 10$ *and the number of trials for the singularity crossing scheme is* $N_{\text{trials}} = 4$. *Initial state of the observer is* $\hat{x}(0) = diag(0.8, 1.3, 1.3) \cdot x(0)$ *and* $\mu_{obs}(0) = 0.8\mu_{model}$.

Simulation results

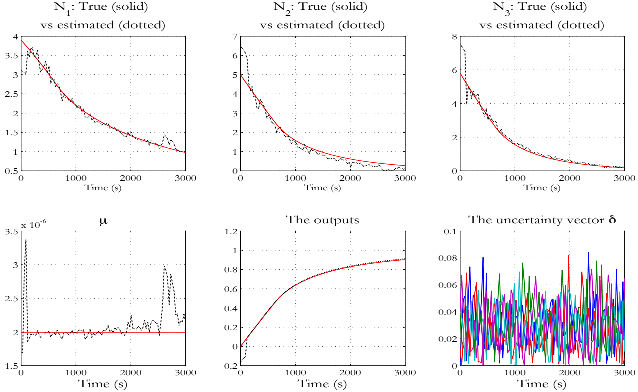


Figure: *Observer behavior under model uncertainty given by (2)-(2) with* $d_{max} = 10\%$ *and in the presence of measurement noise (*$\nu_{max} = 0.01$*). The observation horizon is* $N = 15$ *and the number of trials for the singularity crossing scheme is* $N_{\text{trials}} = 4$*.* $\mu_{obs}(0) = 0.8\mu_{model}$*. Note that concerning the output, only the true output and the estimated one are shown, measurement noise is not presented. This scenario uses a tolerance* $\varepsilon = 10^{-8}$ *for the optimization subroutine.*

Simulation results

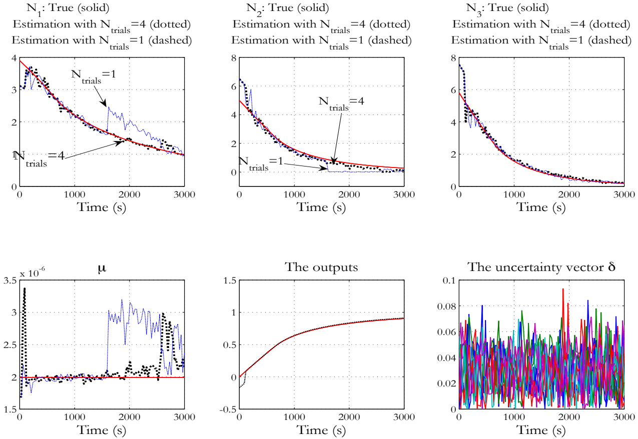


Figure: *Comparison between the observer behavior when $N_{trials} = 1$ and $N_{trials} = 4$ under the scenario depicted on figure 2. Note how the singularity cross mechanism enables to avoid drops in the estimation quality when the observer encounters a singular situation. This scenario uses a tolerance $\varepsilon = 10^{-8}$ for the optimization subroutine.*

Simulation results

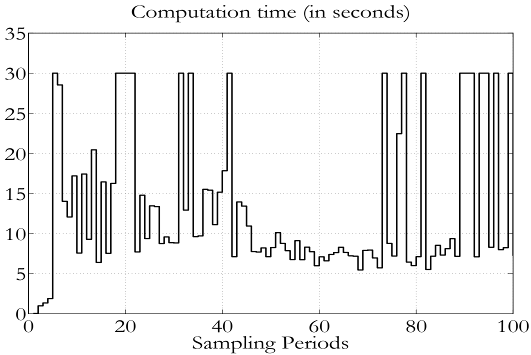


Figure: *Computation times needed to achieve the state estimation depicted on figure 2. Note that an explicit upper bound has been imposed in the internal loop of the optimizer in order to deliver the best estimation that can be obtained within the available computation time defined by the sampling period (30 seconds). This scenario uses a tolerance $\varepsilon = 10^{-8}$ for the optimization subroutine.*

Experimental results

Parameter	Value	Unit
ϕ_p^p	0.4	
MW_1	128.2	(g/mol)
MW_2	100.12	(g/mol)
MW_3	86.09	(g/mol)
ρ_1	0.89	(g/cm^3)
ρ_2	0.94	(g/cm^3)
ρ_3	0.93	(g/cm^3)
$\rho_{1,h}$	1.08	(g/cm^3)
$\rho_{2,h}$	1.15	(g/cm^3)
$\rho_{3,h}$	1.17	(g/cm^3)
k_{p11}	4.5×10^5	$(cm^3/mol/s)$
k_{p22}	1.28×10^6	$(cm^3/mol/s)$
k_{p33}	4.26×10^6	$(cm^3/mol/s)$
r_{12}	0.355	
r_{21}	1.98	
r_{13}	6.635	
r_{31}	0.037	
r_{23}	22.21	
r_{32}	0.07	

Table: Parameter values of the terpolymerization of BuA/MMA/VAc (used in the experimental validation)

Component	Charge (g)
Butyl acrylate	300
Methyl methacrylate	300
Vinyl acetate	60
Sodium dioctyl sulfosuccinate	3
Potassium persulfate	2
Water	2380

Table: Recipe of the terpolymerization of BuA/MMA/VAc

gipsa-lab

Experimental results: $N_{trials} = 10$

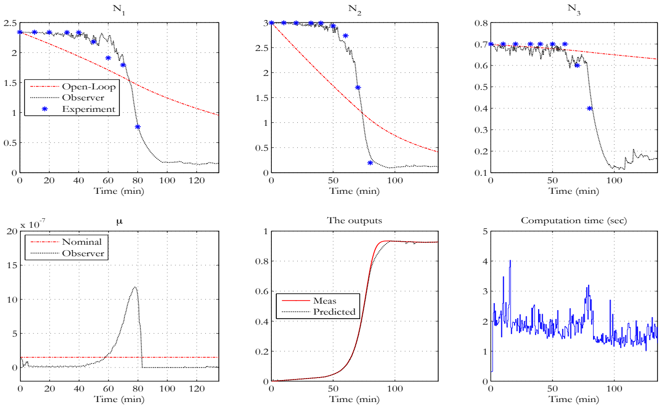


Figure: *Experimental validation with $N_{trials} = 10$ and tolerance threshold $\varepsilon = 10^{-3}$. The same scenario is depicted on figure 6 where $N_{trials} = 1$ is used. The computation time is given in seconds.*

Experimental results: $N_{trials} = 1$

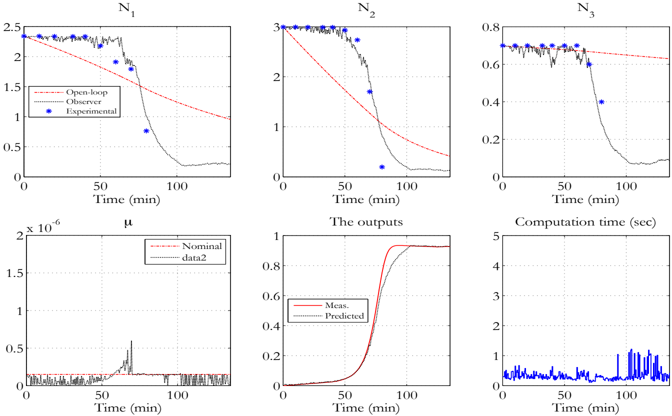


Figure: *Experimental validation with $N_{trials} = 1$ and tolerance threshold $\varepsilon = 10^{-3}$. The same scenario is depicted on figure 5 where $N_{trials} = 10$ is used. The computation time is given in seconds.*

More general formulation for singularities avoidance

Consider a general simulator

$$x(t) = X(t, t_0, x_0) \quad ; \quad y(t) = h(x(t))$$

More general formulation for singularities avoidance

Consider a general simulator

$$x(t) = X(t, t_0, x_0) \quad ; \quad y(t) = h(x(t))$$

Classical observer related optimization problem

$$\min_{z \in \mathbb{X}(t-T)} \Big[J(t,z) \Big] := \int_{t-T}^{t} \| Y(\tau, t-T, z) - y(\tau) \|^2 d\tau$$

More general formulation for singularities avoidance

Consider a general simulator

$$x(t) = X(t, t_0, x_0) \quad ; \quad y(t) = h(x(t))$$

Classical observer related optimization problem

$$\min_{z \in \mathbb{X}(t-T)} \Big[J(t,z) \Big] := \int_{t-T}^{t} \| \underbrace{Y(\tau, t-T, z) - y(\tau)}_{\epsilon_y(\tau,z)} \|^2 d\tau$$

More general formulation for singularities avoidance

Consider a general simulator

$$x(t) = X(t, t_0, x_0) \quad ; \quad y(t) = h(x(t))$$

Classical observer related optimization problem

$$\min_{z \in \mathbb{X}(t-T)} \Big[J(t,z) \Big] := \int_{t-T}^{t} \| \underbrace{Y(\tau, t-T, z) - y(\tau)}_{\epsilon_y(\tau,z)} \|^2 d\tau$$

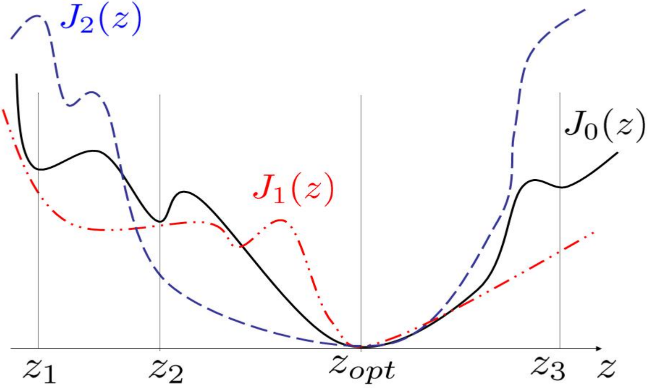
Ideally: $\hat{z}(t) = x(t-T)$ is the unique global minimum for ALL cost functions

$$J_i(t,z) = \int_{t-T}^{t} \Big[\Phi_i(\tau) \cdot \Psi_i(\epsilon_y(\tau, z)) \Big] d\tau$$

for all

- $\Phi_i : [t-T, t] \rightarrow \mathbb{R}_+$
- $\Psi_i(\cdot)$ positive definite function defined on $\mathbb{R}^{n_y}$

Typical scheme behind the intuition



Let us try the special choice

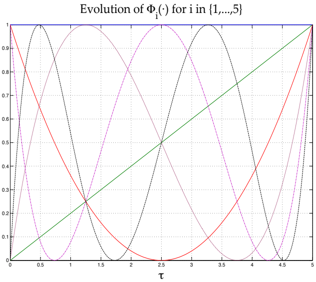
$$J_i(t,z) = \int_{t-T}^{t} \Big[\Phi_i(\tau)\cdot\Psi_i(\epsilon_y(\tau,z))\Big]d\tau$$

with

$$\Psi_i(y) = y^T y$$
$$\Phi_i(\tau) = \frac{1}{2}\Big[T_i(\frac{2\tau}{T}-1)+1\Big]$$

where for $i \in \{1,\dots,N\}$, T_i stands for the ith Tchebychev polynomials of the first kind, namely:

$$\begin{aligned} T_0(x) &= 1 \quad ; \quad T_1(x) = x \\ T_{i+1}(x) &= 2xT_i(x) - T_{i-1}(x) \end{aligned}$$



Example: Recombinant Escherichia Coli

$$\begin{aligned}\dot{X} &= \mu(S)X - k_d \exp(-\frac{k_p}{P})X \\ \dot{S} &= -y_s\mu X - k_m X \\ \dot{P} &= y_p\mu(S)\frac{I}{I+k_I}X - k_d \exp(-\frac{k_p}{P})P\end{aligned}$$

- X: *E. Coli* strain
- S: substrate glycerol
- P: intracellular product β-galactosidase protein
- μ is the growth rate

$$\mu(S) = \frac{\mu_m S}{k_s + S}$$

Figure: Escherichia coli under 15000 magnification factor

Output measurement:
Light produced by the bioluminescence:

$$L = y_l \cdot \mu(S)\frac{I}{I+k_I}XP$$

-lab

Example: Recombinant Escherichia Coli

$$\begin{aligned}\dot{X} &= \mu(S)X - k_d \exp(-\frac{k_p}{P})X \\ \dot{S} &= -y_s\mu X - k_m X \\ \dot{P} &= y_p\mu(S)\frac{I}{I+k_I}X - k_d\exp(-\frac{k_p}{P})P\end{aligned}$$

param.	Val.	Units
μ_m	0.49	h^{-1}
k_s	0.06	g/l
k_p	0.047	g/l
k_d	0.005	g/l
k_m	0.21	h^{-1}
k_I	0.03	g/h
y_s	0.75	g cell/ g glycerol
y_p	0.32	g protein/ β-galactosidase
y_I	17.6	U/β-galactosidase

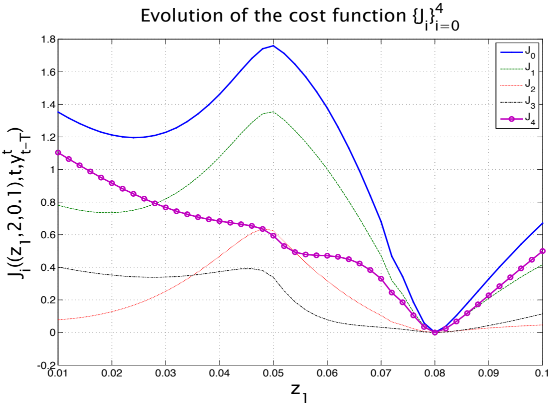
- X: *E. Coli* strain
- S: substrate glycerol
- P: intracellular product β-galactosidase protein
- μ is the growth rate

$$\mu(S) = \frac{\mu_m S}{k_s + S}$$

Output measurement:
Light produced by the bioluminescence:

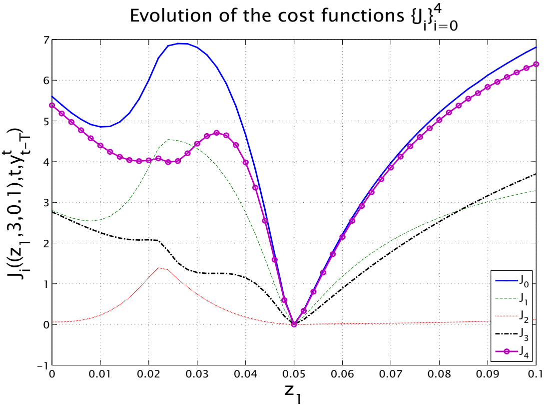
$$L = y_I \cdot \mu(S)\frac{I}{I+k_I}XP$$

-lab



- $T = 10$
- $x(t - T) = (0.08, 2, 0.1)$
- $z_2 = x_2(t - T)$, $z_3 = x_3(t - T)$

gipsa-lab



- $T = 15$
- $x(t-T) = (0.05, 3, 0.1)$
- $z_2 = x_2(t-T)$, $z_3 = x_3(t-T)$

Some general formalism

- Assume some solver iteration:

$$z^{(i+1)} = \mathcal{S}(z^{(i)}, J(\cdot))$$

- Denote multiple-iteration map

$$z^{(i+r)} = \mathcal{S}^{(r)}(z^{(i)}, J)$$

- Define the $(z^{(0)}, J)$-solver path by

$$\left\{\mathcal{S}^{(r)}(z^{(0)}, J)\right\}_{r\in\mathbb{N}}$$

Definition: N-Safely redundant optimization problem

The optimization problem is called *N-safely redundant* iff

1. There exists a finite sequence of N cost functions J_i defined on $\mathbb{R}^n$ that admits $x(t-T)$ as a global minimum.
2. There exists a solver $\mathcal{S}$ and a finite integer $r^* \in \mathbb{N}$ such that:

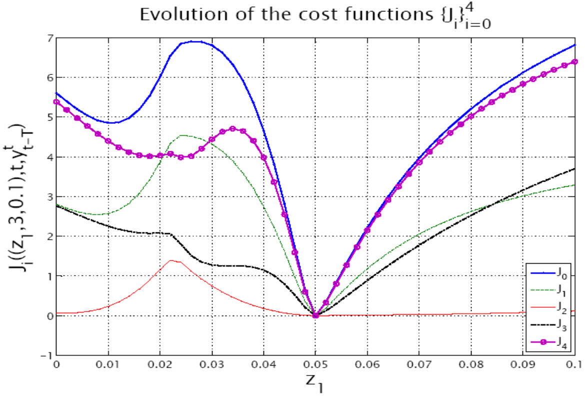
$$\Delta(z) := \min_{i\in\{1,\dots,N\}} \Big[J_0(\mathcal{S}^{(r^*)}(z, J_i)) - \gamma J_0(z) \Big] \leq 0 \tag{2}$$

holds for some $\gamma \in [0, 1[$ and all $z \in \mathcal{Z}$. Moreover:

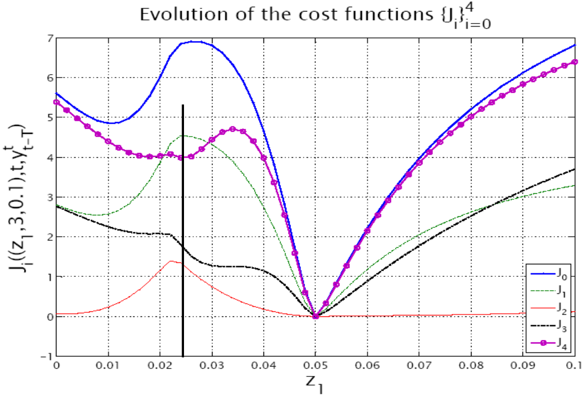
$$\mathcal{S}^{(r^*)}(z, J_{\imath^*}) \in \mathcal{Z}$$

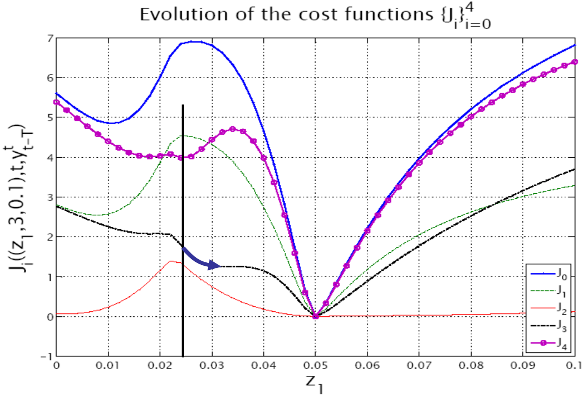
where $\imath^*$ is the optimal argument of the minimization invoked in (2) ♡

gipsa-lab

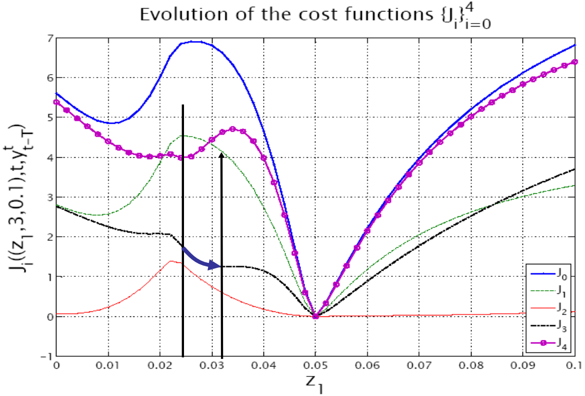


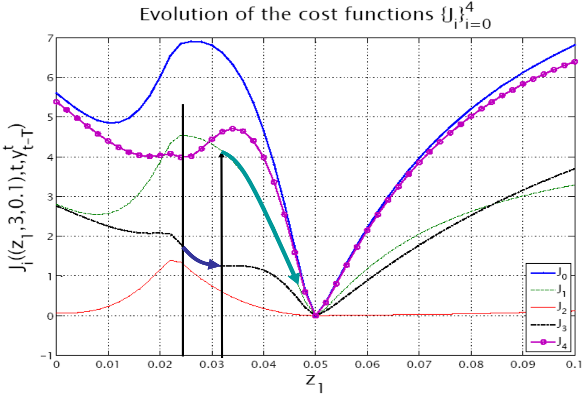
Evolution of the cost functions $\{J_i\}_{i=0}^4$
$J_i((z_1,3,0.1),t,y_{t-T}^t)$
z_1
7
6
5
4
3
2
1
0
−1
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
J_0
J_1
J_2
J_3
J_4


Evolution of the cost functions $\{J_i\}_{i=0}^4$
$J_i((z_1,3,0.1),t,y_{t-T}^t)$
z_1
J_0
J_1
J_2
J_3
J_4



Evolution of the cost functions $\{J_i\}_{i=0}^4$
$J_i((z_1,3,0.1),t,y_{t-T}^t)$
z_1
J_0
J_1
J_2
J_3
J_4


Evolution of the cost functions $\{J_i\}_{i=0}^4$
$J_i((z_1,3,0.1),t,y_{t-T}^t)$
z_1
J_0
J_1
J_2
J_3
J_4



Evolution of the cost functions $\{J_i\}_{i=0}^4$
$J_i((z_1,3,0.1),t,y_{t-T}^t)$
z_1
7
6
5
4
3
2
1
0
−1
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
J_0
J_1
J_2
J_3
J_4

Algorithm A_1

0. Initialization $z^{(0)}$ initial guess, $\sigma \leftarrow 0$
1. while $(J_0(z^{(\sigma)}) > \varepsilon)$ do

End while

Algorithm A_1

0. Initialization $z^{(0)}$ initial guess, $\sigma \leftarrow 0$
1. while $(J_0(z^{(\sigma)}) > \varepsilon)$ do
   - 1.1 $i \leftarrow 1$; again$\leftarrow$ *true*
   - 1.2 while $(\textit{again} \quad \& \quad i \le N)$ do
     - 1.2.1 $\xi^{(\sigma,i)} \leftarrow \mathcal{S}^{r^*}(z^{(\sigma)}, J_i)$
     - 1.2.2 $\textit{again} \leftarrow \Big(J_0(\xi^{(\sigma,i)}) > \gamma J_0(z^{(\sigma)})\Big)$
     - 1.2.3 If *again* then $i \leftarrow i+1$
     - 1.2.4 Else $\sigma \leftarrow \sigma + 1$, $z^{(\sigma)} \leftarrow \xi^{(\sigma,i)}$

   End while

End while

Discussion

- The scheme holds regardless the optimizer $\mathcal{S}$
  - Gradient-based iteration
  - SQP
  - multiple shooting
  - non smooth (simplex, powell's, etc.)

Discussion

- The scheme holds regardless the optimizer $\mathcal{S}$
  - Gradient-based iteration
  - SQP
  - multiple shooting
  - non smooth (simplex, powell's, etc.)

- Easily usable in a parallel architecture
  [parallel $(z^{(0)}, J_i)$ path-solvers exploration]

Discussion

- The scheme holds regardless the optimizer $\mathcal{S}$
  - Gradient-based iteration
  - SQP
  - multiple shooting
  - non smooth (simplex, powell's, etc.)

- Easily usable in a parallel architecture
  [parallel $(z^{(0)}, J_i)$ path-solvers exploration]

- $\neq$ a multiple initial guess scheme
  [It's the cost function that changes not the present solution]

Discussion

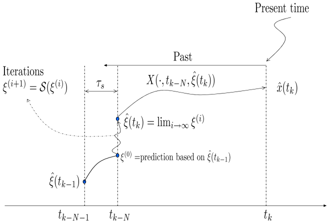
- The scheme holds regardless the optimizer $\mathcal{S}$
  - Gradient-based iteration
  - SQP
  - multiple shooting
  - non smooth (simplex, powell's, etc.)

- Easily usable in a parallel architecture
  [parallel $(z^{(0)}, J_i)$ path-solvers exploration]

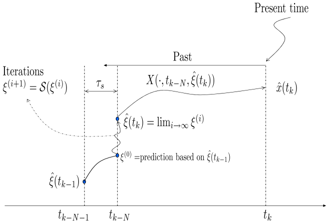
- $\neq$ a multiple initial guess scheme
  [It's the cost function that changes not the present solution]

- Price: Loose optimality that is *loosely defined*

Back to the non-ideal real-time iterations

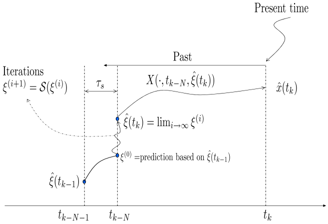


Back to the non-ideal real-time iterations



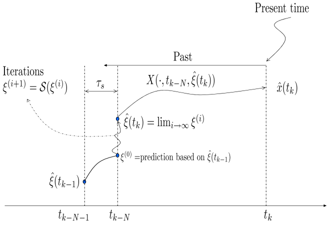
- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi) \Big]}^{\text{Initial Guess } \xi^{(0)}}$

Back to the non-ideal real-time iterations



- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi\in\mathbb{X}(t_{k-N})}\Big[J(t_k,\xi)\Big]}^{\text{Initial Guess } \xi^{(0)}}$
- $\hat{\xi}(t_k) = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$

Back to the non-ideal real-time iterations

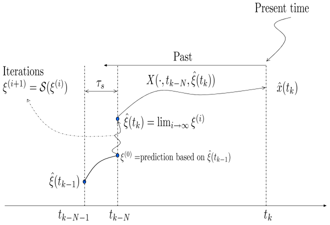


- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi)\Big]}^{\text{Initial Guess } \xi^{(0)}}$

- $\hat{\xi}(t_k) = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$

- $\xi^{(0)} = X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$

Back to the non-ideal real-time iterations

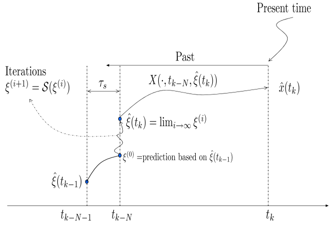


$\Rightarrow$ Implicit updating rule

$$\hat{\xi}(t_k) = F(t_k, \hat{\xi}(t_{k-1}), y_{t_{k-N}}^{t_k})$$

- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi) \Big]}^{\text{Initial Guess } \xi^{(0)}}$
- $\hat{\xi}(t_k) = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$
- $\xi^{(0)} = X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$

Back to the non-ideal real-time iterations



- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi) \Big]}^{\text{Initial Guess } \xi^{(0)}}$
- $\hat{\xi}(t_k) = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$
- $\xi^{(0)} = X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$

$\Rightarrow$ Implicit updating rule

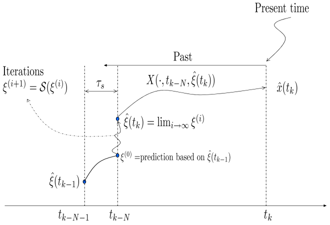
$$\hat{\xi}(t_k) = F(t_k, \hat{\xi}(t_{k-1}), y_{t_{k-N}}^{t_k})$$

$\downarrow$

Is there a differential form of this updating rule ?

$$\frac{d\xi}{dt}(t) = f(t, \xi, y_{t-T}^{t})$$

Back to the non-ideal real-time iterations



- $\hat{\xi}(t_k) = \overbrace{arg \min_{\xi \in \mathbb{X}(t_{k-N})} \Big[J(t_k, \xi) \Big]}^{\text{Initial Guess } \xi^{(0)}}$
- $\hat{\xi}(t_k) = \mathcal{S}^{N_{max}}(\xi^{(0)}, t_k, y_{t_{k-N}}^{t_k})$
- $\xi^{(0)} = X(t_{k-N}, t_{k-N-1}, \hat{\xi}(t_{k-1}))$

$\Rightarrow$ Implicit updating rule

$$\hat{\xi}(t_k) = F(t_k, \hat{\xi}(t_{k-1}), y_{t_{k-N}}^{t_k})$$

$\downarrow$

Is there a differential form of this updating rule ?

$$\frac{d\xi}{dt}(t) = f(t, \xi, y_{t-T}^{t})$$

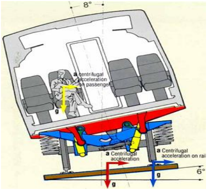
$\rightarrow$ Differential form for moving horizon observers

gipsa-lab

Differential Form of Moving-Horizon Observers: Outline

For system of the form $\dot{x} = f(t,x)$:

$$\begin{aligned}\dot{\xi}(t) &= f(t-T,\xi(t)) + c\big(t,\xi(t)\big)\\ \hat{x}(t) &= X(t,t-T,\xi(t))\end{aligned}$$



- The correction term

$$c(t,\xi) := \gamma\left[\frac{J_\xi^T(t,\xi)}{\|J_\xi\|^2+\varepsilon}\right]\left[-\left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi))\right| - \big[1+\phi(t,\xi)\big]\sqrt{J}\right]$$

- Post-Stabilization technique → improve (Sampling period)/Precision ratio
- Alstom-Transport Patent

Preliminary definitions & assumptions

System model

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)) \\ y(t) &= h(t, x(t)) \end{aligned}$$

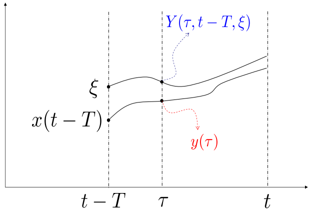
Notations

- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

Preliminary definitions & assumptions

System model

$$\begin{aligned}\dot{x}(t) &= f(t, x(t)) \\ y(t) &= h(t, x(t))\end{aligned}$$



Notations

- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

The cost function

$$J(t, \xi) = \int_{t-T}^{t} \|Y(\tau, t-T, \xi) - y(\tau)\|^2 d\tau$$

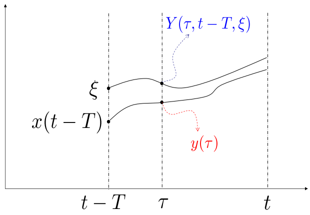
NOTA:

Dependence w.r.t y_{t-T}^{t} are implicitly assumed through dependence on t

Preliminary definitions & assumptions

System model

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)) \\ y(t) &= h(t, x(t)) \end{aligned}$$



Notations

- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

The cost function

$$J(t, \xi) = \int_{t-T}^{t} \|Y(\tau, t-T, \xi) - y(\tau)\|^2 d\tau$$

- if $\xi = x(t-T)$ then $J(t, \xi) = 0$

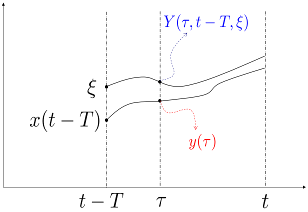
NOTA:

Dependence w.r.t y_{t-T}^{t} are implicitly assumed through dependence on t

Preliminary definitions & assumptions

System model

$$\dot{x}(t) = f(t, x(t))$$
$$y(t) = h(t, x(t))$$



Notations

- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

The cost function

$$J(t, \xi) = \int_{t-T}^{t} \|Y(\tau, t-T, \xi) - y(\tau)\|^2 d\tau$$

- if $\xi = x(t-T)$ then $J(t, \xi) = 0$
- Let ξ be a dynamic variable $\xi(t)$

NOTA:

Dependence w.r.t y_{t-T}^{t} are implicitly assumed through dependence on t

Preliminary definitions & assumptions

System model

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)) \\ y(t) &= h(t, x(t)) \end{aligned}$$

Notations

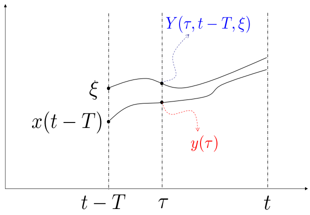
- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

The cost function

$$J(t, \xi) = \int_{t-T}^{t} \|Y(\tau, t-T, \xi) - y(\tau)\|^2 d\tau$$

NOTA:

Dependence w.r.t y_{t-T}^{t} are implicitly assumed through dependence on t



- if $\xi = x(t-T)$ then $J(t, \xi) = 0$
- Let ξ be a dynamic variable $\xi(t)$
- Look for a dynamic on ξ:

$$\dot{\xi}(t) = ?$$

such that $\lim_{t\to\infty} \xi(t) = x(t-T)$

Preliminary definitions & assumptions

System model

$$\dot{x}(t) = f(t, x(t))$$
$$y(t) = h(t, x(t))$$

Notations

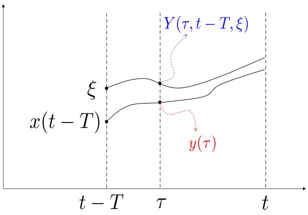
- $X(t, t_0, x_0)$: State evolution
- $Y(t, t_0, x_0)$: Output evolution

The cost function

$$J(t,\xi) = \int_{t-T}^{t} \|Y(\tau, t-T, \xi) - y(\tau)\|^2 d\tau$$

Observer

$$\dot{\xi}(t) = ?$$
$$\hat{x}(t) = X(t, t-T, \xi(t))$$



- if $\xi = x(t-T)$ then $J(t,\xi) = 0$
- Let ξ be a dynamic variable $\xi(t)$
- Look for a dynamic on ξ:

$$\dot{\xi}(t) = ?$$

such that $\lim_{t\to\infty} \xi(t) = x(t-T)$

Reformulating the condition on $\dot{\xi}(t)$

By the very definition of observability, the two following formulations are equivalent

Formulation 1

Look for a dynamic on ξ:

$$\dot{\xi}(t) = ?$$

such that $\lim_{t\to\infty} \xi(t) = x(t-T)$

Formulation 2

Look for a dynamic on ξ:

$$\dot{\xi}(t) = ?$$

such that $\lim_{t\to\infty} J(t, \xi(t)) = 0$

Formulation 2

Look for a dynamic on ξ:

$$\dot{\xi}(t) = ?$$

such that $\lim_{t\to\infty} J(t,\xi(t)) = 0$

Observer

$$\dot{\xi}(t) = ?$$
$$\hat{x}(t) = X(t, t-T, \xi(t))$$

Conditions on the dynamic of $\dot{\xi}$

1. Consistency

$$\dot{\xi}(t) = f(t-T,\xi(t)) + \underbrace{c(t,\xi(t))}_{\text{correction term } O(J(t,\xi(t)))}$$

2. Convergence

$$\lim_{t\to\infty} J(t,\xi(t)) = 0$$

gipsa-lab

Consistency

$$\dot{\xi}(t) = f(t - T, \xi(t)) + c(t, \xi(t))$$

Convergence

$$\lim_{t \to \infty} J(t, \xi(t)) = 0$$

gipsa-lab

Consistency

$$\dot{\xi}(t) = f(t - T, \xi(t)) + c(t, \xi(t))$$

Convergence

$$\lim_{t \to \infty} J(t, \xi(t)) = 0$$

$$\frac{dJ}{dt}(t, \xi(t)) = J_t(t, \xi(t)) + \left[J_\xi(t, \xi(t))\right]\dot{\xi}$$

Consistency

$$\dot{\xi}(t) = f(t-T, \xi(t)) + c(t, \xi(t))$$

Convergence

$$\lim_{t \to \infty} J(t, \xi(t)) = 0$$

$$\frac{dJ}{dt}(t, \xi(t)) = J_t(t, \xi(t)) + \left[J_\xi(t, \xi(t))\right]\dot{\xi}$$

$$\frac{dJ}{dt} = J_t(t, \xi(t)) + \left[J_\xi(t, \xi(t))\right] \cdot \left[f(t-T, \xi(t)) + c\big(t, \xi(t)\big)\right]$$

Consistency

$$\dot{\xi}(t) = f(t-T,\xi(t)) + c(t,\xi(t))$$

Convergence

$$\lim_{t\to\infty} J(t,\xi(t)) = 0$$

$$\frac{dJ}{dt}(t,\xi(t)) = J_t(t,\xi(t)) + \Big[J_\xi(t,\xi(t))\Big]\dot{\xi}$$

$$\frac{dJ}{dt} = J_t(t,\xi(t)) + \Big[J_\xi(t,\xi(t))\Big]\cdot\Big[f(t-T,\xi(t)) + c(t,\xi(t))\Big]$$

This writes

$$\frac{dJ}{dt} = \frac{dJ}{dt}|_{c(\cdot,\cdot)\equiv 0} + \Big[J_\xi(t,\xi(t))\Big]\cdot c(t,\xi(t))$$

gipsa-lab

$$\frac{dJ}{dt} = \frac{dJ}{dt}|_{c(\cdot,\cdot)\equiv 0} + \Big[J_\xi(t,\xi(t))\Big]\cdot c(t,\xi(t))$$

Lemma

The correction-free time derivative of J satisfies:

$$\frac{dJ}{dt}|_{c(\cdot,\cdot)\equiv 0} \le |\epsilon_y(t,\xi(t)) - \epsilon_y(t-T,\xi(t))| + \Big[\phi(t,\xi(t))\Big]\cdot\sqrt{J(t,\xi(t))}$$

where

$$J(t,\xi) = \int_{t-T}^{t} \|Y(\tau,t-T,\xi) - y(\tau)\|^2 d\tau$$

$$\epsilon_y(\tau,\xi(t)) = Y(\tau,t-T,\xi(t)) - y(\tau) \quad \forall \tau\in[t-T,t]$$

$$\Big[\phi(t,\xi(t))\Big] = \sup_{\tau\in[t-T,t]}\left|\frac{dY}{dt}|_{c\equiv 0}(\tau,t-T,\xi(t)) - \dot{y}(\tau)\right|$$

$$\frac{dJ}{dt} = \frac{dJ}{dt}|_{c(\cdot,\cdot)\equiv 0} + \Big[J_\xi(t,\xi(t))\Big]\cdot c(t,\xi(t))$$

Therefore

The correction-free time derivative of J satisfies:

$$\frac{dJ}{dt}|_{c(\cdot,\cdot)\equiv 0} \le \left|\Delta_{t-T}^{t}(\epsilon(\cdot,\xi(t)))\right| + \Big[\phi(t,\xi(t))\Big]\cdot\sqrt{J(t,\xi(t))}$$

where

- $\Delta_{t-T}^{t}(\cdot,\xi(t)) = \epsilon_y(t,\xi(t)) - \epsilon_y(t-T,\xi(t))$ and
- $\Big[\phi(t,\xi(t))\Big]$

are computable quantities

psa-lab

$$\frac{dJ}{dt} \le \left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right]\cdot\sqrt{J} + \left[J_\xi(t,\xi(t))\right]\cdot c(t,\xi(t))$$

Therefore

The correction-free time derivative of J satisfies:

$$\frac{dJ}{dt}\Big|_{c(\cdot,\cdot)\equiv 0} \le \left|\Delta_{t-T}^{t}(\epsilon(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right]\cdot\sqrt{J(t,\xi(t))}$$

where

- $\Delta_{t-T}^{t}(\cdot,\xi(t)) = \epsilon_y(t,\xi(t)) - \epsilon_y(t-T,\xi(t))$ and
- $\left[\phi(t,\xi(t))\right]$

are computable quantities

psa-lab

$$\frac{dJ}{dt} \leq \left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right] \cdot \sqrt{J} + \left[J_\xi(t,\xi(t))\right] \cdot c(t,\xi(t))$$

The ability to decrease J depend on the *rank of J_ξ*

Uniform Global Regularity Assumption

There is a K-function $\Upsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that the following inequality holds:

$$\|J_\xi(t,\xi)\|^2 \geq \Upsilon(J(t,\xi))$$

for all (t,ξ) ♡

- Non constructive assumption (only for proof)
- Locally linked to the observability of the linearized system

$$\frac{dJ}{dt} \leq \left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right]\cdot\sqrt{J} + \left[J_\xi(t,\xi(t))\right]\cdot c(t,\xi(t))$$

The ability to decrease J depend on the *rank of J_ξ*

Uniform Global Regularity Assumption

There is a K-function $\Upsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that the following inequality holds:

$$\|J_\xi(t,\xi)\|^2 \geq \Upsilon(J(t,\xi))$$

for all (t,ξ) ♡

This suggests the following correction term

$$c(t,\xi(t)) := \gamma\left[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+\varepsilon}\right]\left[-\left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| - \left[1+\phi(t,\xi(t))\right]\sqrt{J}\right]$$

$$\frac{dJ}{dt} \le \left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right]\cdot\sqrt{J} + \left[J_\xi(t,\xi(t))\right]\cdot c(t,\xi(t))$$

The ability to decrease J depend on the *rank of J_ξ*

Uniform Global Regularity Assumption

There is a K-function $\Upsilon : \mathbb{R}_+ \to \mathbb{R}_+$ such that the following inequality holds:

$$\|J_\xi(t,\xi)\|^2 \ge \Upsilon(J(t,\xi))$$

for all (t,ξ) ♡

This suggests the following correction term

$$c(t,\xi(t)) := 1\left[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+0}\right]\left[-\left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| - \left[1+\phi(t,\xi(t))\right]\sqrt{J}\right]$$

$$\frac{dJ}{dt} \leq -\sqrt{J(t,\xi(t))}$$

The ability to decrease J depend on the *rank of J_ξ*

Uniform Global Regularity Assumption

There is a K-function $\Upsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that the following inequality holds:

$$\|J_\xi(t,\xi)\|^2 \geq \Upsilon(J(t,\xi))$$

for all (t,ξ) ♡

This suggests the following correction term

$$c(t,\xi(t)) := 1\left[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+0}\right]\left[-\left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| - \left[1+\phi(t,\xi(t))\right]\sqrt{J}\right]$$

gipsa-lab

$$\frac{dJ}{dt} \leq \left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| + \left[\phi(t,\xi(t))\right]\cdot\sqrt{J} + \left[J_\xi(t,\xi(t))\right]\cdot c(t,\xi(t))$$

The ability to decrease J depend on the *rank of J_ξ*

Uniform Global Regularity Assumption

There is a K-function $\Upsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that the following inequality holds:

$$\|J_\xi(t,\xi)\|^2 \geq \Upsilon(J(t,\xi))$$

for all (t,ξ) ♡

This suggests the following correction term

$$c(t,\xi(t)) := \gamma\left[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+\varepsilon}\right]\left[-\left|\Delta_{t-T}^{t}(\epsilon_y(\cdot,\xi(t)))\right| - \left[1+\phi(t,\xi(t))\right]\sqrt{J}\right]$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\Big(\frac{\varepsilon}{\gamma - 1}\Big)\right\}$$

is invariant and globally attractive set.

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\Big(\frac{\varepsilon}{\gamma - 1}\Big)\right\}$$

is invariant and globally attractive set.

But according to the definition of observability

$$\int_{t-T}^{t} \|Y(\sigma, t-T, x^{(1)}) - Y(\sigma, t-T, x^{(2)})\|^2 d\sigma \geq \alpha(\|x^{(1)} - x^{(2)}\|)$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\left(\frac{\varepsilon}{\gamma - 1}\right)\right\}$$

is invariant and globally attractive set.

Taking $x^{(1)} = x(t-T)$ and $x^{(2)} = \xi(t)$

$$\int_{t-T}^{t} \|Y(\sigma, t-T, x^{(1)}) - Y(\sigma, t-T, x^{(2)})\|^2 d\sigma \geq \alpha(\|x^{(1)} - x^{(2)}\|)$$

gipsa-lab

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\left(\frac{\varepsilon}{\gamma - 1}\right)\right\}$$

is invariant and globally attractive set.

one has

$$J(t,\xi(t)) \geq \alpha(\|x(t - T) - \xi(t)\|)$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma-1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\left(\frac{\varepsilon}{\gamma-1}\right)\right\}$$

is invariant and globally attractive set.

one has

$$\alpha^{-1}\Big(J(t,\xi(t))\Big) \geq \|x(t-T)-\xi(t)\|$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\left(\frac{\varepsilon}{\gamma - 1}\right)\right\}$$

is invariant and globally attractive set.

one has

$$\alpha^{-1} \circ \Upsilon^{-1}\left(\frac{\varepsilon}{\gamma - 1}\right) \underbrace{\geq}_{\text{asymptotically}} \alpha^{-1}\Big(J(t,\xi(t))\Big) \geq \|x(t-T) - \xi(t)\|$$

More generally ($\gamma > 1$ and $\varepsilon \neq 0$), one has the following implication

$$\left\{\Upsilon(J(t,\xi(t)) > \frac{\varepsilon}{\gamma - 1}\right\} \quad \Rightarrow \quad \left\{\dot{J}(t,\xi(t)) < 0\right\}.$$

therefore, the following set

$$\mathcal{A}_J := \left\{(t,\xi) \quad | \quad J(t,\xi) \leq \Upsilon^{-1}\Big(\frac{\varepsilon}{\gamma - 1}\Big)\right\}$$

is invariant and globally attractive set.

one has

$$\alpha^{-1} \circ \Upsilon^{-1}\Big(\frac{\varepsilon}{\gamma - 1}\Big) \underbrace{\geq}_{\text{asymptotically}} \alpha^{-1}\Big(J(t,\xi(t))\Big) \geq \|x(t-T) - \xi(t)\|$$

$$\lim_{(\varepsilon/\gamma)\to 0}\left[\lim_{t\to\infty}\|\hat{x}(t) - x(t)\|\right] = 0.$$

Convergence result

If the following conditions hold:

1. The maps are continuously differentiable
2. The system is uniformly observable
3. The uniform regularity assumption is satisfied

then for any a priori fixed desired precision $\eta > 0$ on the state estimation error, there is a sufficiently high ratio γ/ε such that the dynamic system given by:

$$\begin{aligned} \dot{\xi}(t) &= f(t-T,\xi(t)) + c\big(t,\xi(t)\big) \\ \hat{x}(t) &= X(t,t-T,\xi(t)) \end{aligned}$$

where the correction term $c(t,\xi)$ is given by:

$$c(t,J) := \gamma\Big[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+\varepsilon}\Big]\Big[-\big|\Delta_{t-T}^t(\epsilon_y(\cdot,\xi(t)))\big| - \big[1+\phi(t,\xi(t))\big]\sqrt{J}\Big]$$

leads to an estimation error that is asymptotically lower than η. ♡

psa-lab

The Post Stabilization Technique

$$\dot{\xi}(t) = f(t-T,\xi(t)) + c\big(t,\xi(t)\big)$$

$$c(t,J) := \gamma\Big[\frac{J_\xi^T(t,\xi(t))}{\|J_\xi\|^2+\varepsilon}\Big]\Big[-\big|\Delta_{t-T}^t(\epsilon_y(\cdot,\xi(t)))\big| - \big[1+\phi(t,\xi(t))\big]\sqrt{J}\Big]$$

The Post Stabilization Technique

$$
\begin{aligned}
\dot{\xi}(t) &= f_c(t, \xi(t), J_\xi(t)) \\
\hat{x}(t) &= X(t, t-T, \xi(t)) \\
J(t, \xi(t)) &= 0 \quad \text{(Ideally)}
\end{aligned}
$$

The Post Stabilization Technique

$$\begin{aligned} \dot{\xi}(t) &= f_c(t, \xi(t), J_\xi(t)) \\ \hat{x}(t) &= X(t, t-T, \xi(t)) \\ J(t, \xi(t)) &= 0 \quad \text{(Ideally)} \end{aligned}$$

Bad News: The computation of the r.h.s of the observer equation is expensive.

[Integration of a differential system of order $n(n+1)$]

$\rightarrow$ For a given sampling period, need for lower order integration methods

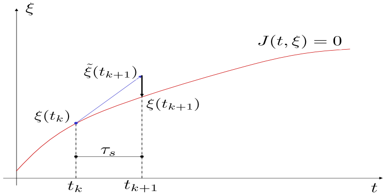
gipsa-lab

The Post Stabilization Technique

$$\begin{aligned}\dot{\xi}(t) &= f_c(t, \xi(t), J_\xi(t)) \\ \hat{x}(t) &= X(t, t-T, \xi(t)) \\ J(t, \xi(t)) &= 0 \quad \text{(Ideally)}\end{aligned}$$

Good News: Ideally, DAE's with invariant submanifold

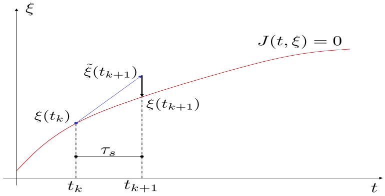
→ There are techniques (Ascher, Num. Alg. 1997) to accurately integrate with lower order methods



1. Integrate over $[t_k, t_{k+1}]$ starting from the initial condition $(t_k, \xi(t_k))$

$$\dot{\xi}(t) = f_c(t, \xi(t), J_\xi(t_k)) \quad ; \quad t \in [t_k, t_{k+1}]$$

to obtain $\tilde{\xi}(t_{k+1})$



1. Integrate over $[t_k, t_{k+1}]$ starting from the initial condition $(t_k, \xi(t_k))$

$$\dot{\xi}(t) = f_c(t, \xi(t), J_\xi(t_k)) \quad ; \quad t \in [t_k, t_{k+1}]$$

to obtain $\tilde{\xi}(t_{k+1})$

2. Correct the *rough* approximation $\tilde{\xi}(t_{k+1})$ by projection

$$\xi(t_{k+1}) = \tilde{\xi}(t_{k+1}) - \frac{J_\xi(t_{k+1}, \tilde{\xi}(t_{k+1}))}{\|J_\xi(t_{k+1}, \tilde{\xi}(t_{k+1}))\|^2 + \nu} \cdot J(t_{k+1}, \tilde{\xi}(t_{k+1}))$$

to obtain the update $\xi(t_{k+1})$

Convergence analysis

Regardless the order of the integration scheme, one has

$$\lim_{k\to\infty} J(\xi(t_k)) = O(\tau_s^4)$$

[Alamir: Int. J. Contr. 1999]

1. Integrate over $[t_k, t_{k+1}]$ starting from the initial condition $(t_k, \xi(t_k))$

$$\dot{\xi}(t) = f_c(t, \xi(t), J_\xi(t_k)) \quad ; \quad t \in [t_k, t_{k+1}]$$

   to obtain $\tilde{\xi}(t_{k+1})$

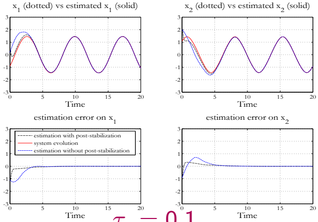
2. Correct the *rough* approximation $\tilde{\xi}(t_{k+1})$ by projection

$$\xi(t_{k+1}) = \tilde{\xi}(t_{k+1}) - \frac{J_\xi(t_{k+1}, \tilde{\xi}(t_{k+1}))}{\|J_\xi(t_{k+1}, \tilde{\xi}(t_{k+1}))\|^2 + \nu} \cdot J(t_{k+1}, \tilde{\xi}(t_{k+1}))$$

   to obtain the update $\xi(t_{k+1})$

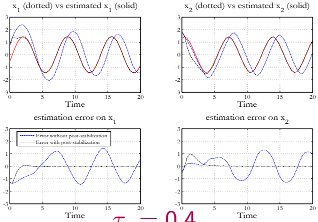
The benefit from the post-stabilization technique

$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -\sin(x_1) - 0.2x_1\cos(x_1x_2) \\
y &= x_1 + x_2
\end{aligned}
$$



$\tau_s = 0.1$

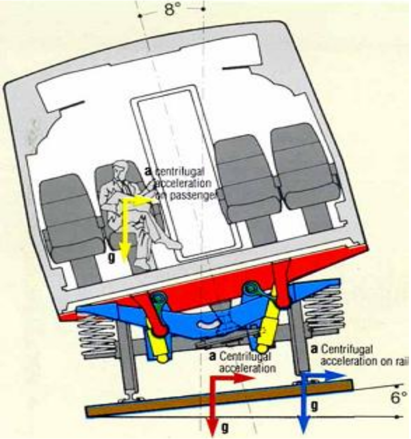
Almost no need for Post-stabilization step



$\tau_s = 0.4$

Post-stabilization is mandatory to keep precision under high sampling period.

The Moving-Horizon Observer in Tilting Trains



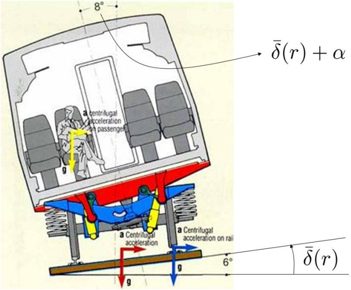
Context

- Coll. Alstom-Transport (Villuerbanne)
- 1997-1999
- Patented solution

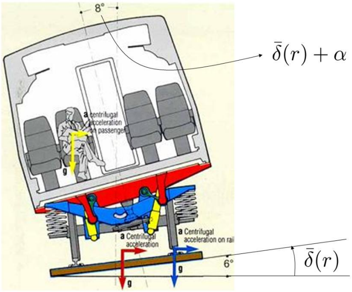
Outline

- Tilting train control problem
- Estimation problem
- Appl. of the Diff. Form. of MHO

Why tilting trains ?

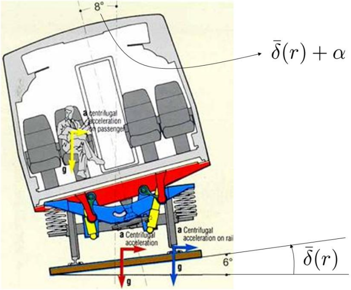


Why tilting trains ?



- Centrifugal acceleration $V^2\rho(r)$

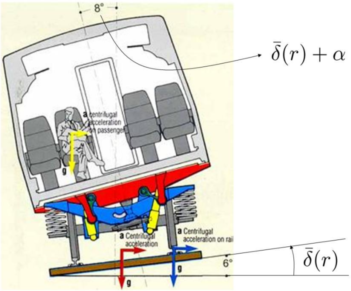
Why tilting trains ?



- Centrifugal acceleration $V^2\rho(r)$
- By construction (comfort)

$$\bar{\delta}(r) = \tan^{-1}\Big(\frac{V_{nom}^2 \cdot \rho(r)}{g}\Big)$$

Why tilting trains ?



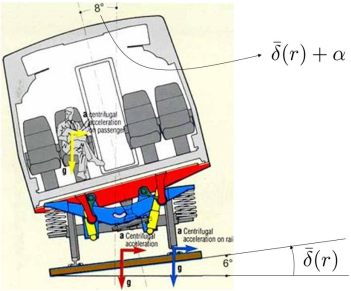
- Centrifugal acceleration $V^2\rho(r)$
- By construction (comfort)

$$\bar{\delta}(r) = \tan^{-1}\left(\frac{V_{nom}^2 \cdot \rho(r)}{g}\right)$$

- If $V > V_{nom}$, we need an additional inclinaison α_d

$$\alpha_d(V(t), r(t)) := \tan^{-1}\left(\frac{1}{g} \cdot V^2(t) \cdot \rho(r(t))\right) - \bar{\delta}(r(t))$$

Why tilting trains ?



- Centrifugal acceleration $V^2\rho(r)$
- By construction (comfort)

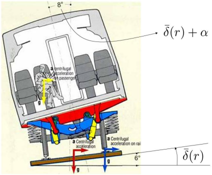
$$\bar{\delta}(r) = \tan^{-1}\Big(\frac{V_{nom}^2 \cdot \rho(r)}{g}\Big)$$

- If $V > V_{nom}$, we need an additional inclinaison α_d

$$\alpha_d(V(t), r(t)) := \tan^{-1}\Big(\frac{1}{g} \cdot V^2(t) \cdot \rho(r(t))\Big) - \bar{\delta}(r(t))$$

Tilting trains enables higher velocities on old rails while maintaining almost the same comfort level

Tilting train: The Control Problem



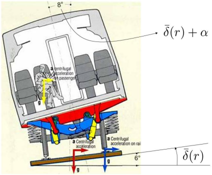
Track the reference trajectory

$$\alpha_d(V(t), r(t)) := \tan^{-1}\left(\frac{1}{g} \cdot V^2(t) \cdot \rho(r(t))\right) - \bar{\delta}(r(t))$$

that depends at each instant t on

- The train's velocity $V(t)$
- The curvilinear abscissa of the train on the rail $r(t)$
- The geometric characteristics of the rail $\rho(r(t))$ and $\bar{\delta}(r(t))$

Tilting train: The Control Problem

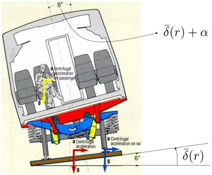


The time derivative of the reference

$$\dot{\alpha}_d = \frac{\partial \alpha_d}{\partial V}\dot{V} + \frac{\partial \alpha_d}{\partial r}V \approx \frac{\partial \alpha_d}{\partial r}(V,r)V$$

is high precisely at high velocities when tilting is needed.

Tilting train: The Control Problem



The time derivative of the reference

$$\dot{\alpha}_d = \frac{\partial \alpha_d}{\partial V}\dot{V} + \frac{\partial \alpha_d}{\partial r}V \approx \frac{\partial \alpha_d}{\partial r}(V, r)V$$

is high precisely at high velocities when tilting is needed.

- Need for an anticipative action (Predictive control)
- Any delay may give the exact inverse effect
- Anticipation needs $r(t)$ to be known
- Estimating $V(t)$ is a classical trouble in railways applications

The Estimation Problem

- State

$$x := \begin{pmatrix} \text{crvilinear abscissa } r \\ \text{relative error on velocity} \end{pmatrix}$$

- Measurement

$$\begin{aligned} y &= \text{Yaw angular velocity} \\ V_m &= \text{velocity sensor output} \end{aligned}$$

System model for observation

$$\begin{aligned} \dot{x}_1 &= (1 + x_2) \cdot V_m \\ \dot{x}_2 &= 0 \\ y &= (1 + x_2) V_m \cdot \rho(x_1) \end{aligned}$$

The map $\rho(\cdot)$ is available after *off-line* careful crossing of the line under consideration

[Paris-toulouse profile]

Computation of $J(t,\xi)$ and $J_\xi(t,\xi)$

Sensitivity System

$$
\begin{aligned}
\dot{z}_1 &= (1+z_2)V \\
\dot{z}_2 &= 0 \\
\dot{z}_3 &= \Big[(1+z_2)V\rho(z_1) - y(t)\Big]^2 \\
\dot{A}(\tau) &= \begin{pmatrix} 0 & V(\tau) & 0 \\ 0 & 0 & 0 \\ \mathcal{X}_1(\tau) & \mathcal{X}_2(\tau) & 0 \end{pmatrix} A
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{X}_1 &= \frac{\partial \dot{z}_3}{\partial z_1} \\
\mathcal{X}_2 &= \frac{\partial \dot{z}_3}{\partial z_2}
\end{aligned}
$$

Computation of $J(t,\xi)$ and $J_\xi(t,\xi)$

Sensitivity System

$$
\begin{aligned}
\dot z_1 &= (1+z_2)V\\
\dot z_2 &= 0\\
\dot z_3 &= \Big[(1+z_2)V\rho(z_1) - y(t)\Big]^2\\
\dot A(\tau) &= \begin{pmatrix} 0 & V(\tau) & 0\\ 0 & 0 & 0\\ \mathcal{X}_1(\tau) & \mathcal{X}_2(\tau) & 0\end{pmatrix} A
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{X}_1 &= \frac{\partial \dot z_3}{\partial z_1} = 2\Big[(1+z_2)V\cdot\rho(z_1) - y\Big](1+z_2)V\frac{\partial\rho}{\partial z_1}(z_1)\\
\mathcal{X}_2 &= \frac{\partial \dot z_3}{\partial z_2} = 2\Big[(1+z_2)V\cdot\rho(z_1) - y\Big]\cdot V\cdot\rho(z_1)
\end{aligned}
$$

Computation of $J(t,\xi)$ and $J_\xi(t,\xi)$

Sensitivity System

$$\begin{aligned}
\dot{z}_1 &= (1+z_2)V \\
\dot{z}_2 &= 0 \\
\dot{z}_3 &= \Big[(1+z_2)V\rho(z_1) - y(t)\Big]^2 \\
\dot{A}(\tau) &= \begin{pmatrix} 0 & V(\tau) & 0 \\ 0 & 0 & 0 \\ \mathcal{X}_1(\tau) & \mathcal{X}_2(\tau) & 0 \end{pmatrix} A
\end{aligned}$$

Initial Conditions

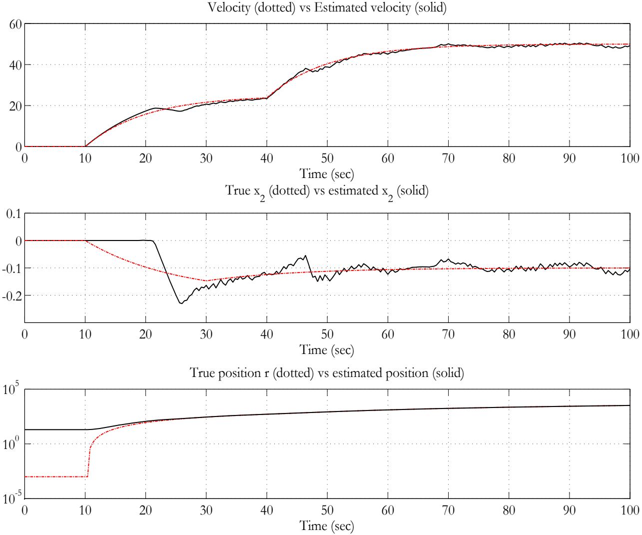
$$\begin{aligned}
z(t-T) &= (\xi^T, 0)^T \\
A(t-T) &= \mathbb{I}_{3\times 3}
\end{aligned}$$

$$J(t,\xi) = z_3(t)$$
$$J_\xi(t,\xi) = (A_{31}(t), A_{32}(t))$$

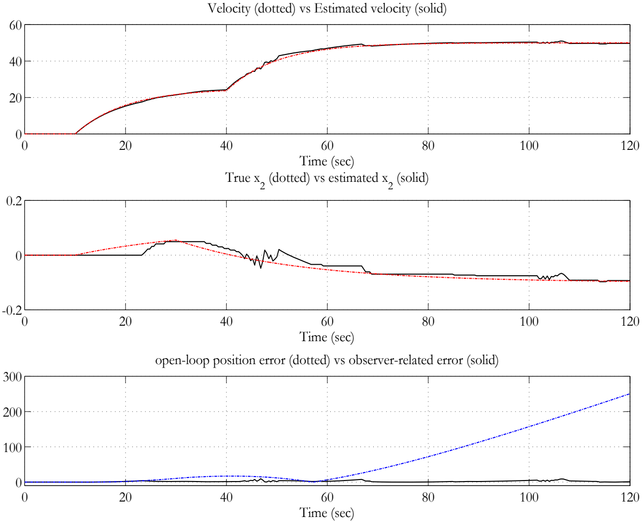
where

$$\begin{aligned}
\mathcal{X}_1 &= \frac{\partial \dot{z}_3}{\partial z_1} = 2\Big[(1+z_2)V\cdot\rho(z_1) - y\Big](1+z_2)V\frac{\partial\rho}{\partial z_1}(z_1) \\
\mathcal{X}_2 &= \frac{\partial \dot{z}_3}{\partial z_2} = 2\Big[(1+z_2)V\cdot\rho(z_1) - y\Big]\cdot V\cdot\rho(z_1)
\end{aligned}$$

Validation on the Paris-Toulouse line ($T = 5$ *sec* and $\tau_s = 0.4$ *sec*)



Validation on the Paris-Toulouse line ($T = 5$ *sec* and $\tau_s = 0.4$ *sec*)



Moving-Horizon Observers with Distributed Optimization

- The system

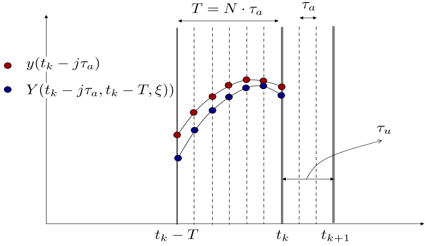
$$\begin{aligned} x(t) &= X(t, t_0, x_0), \\ y(t) &= h(t, x(t)), \end{aligned}$$

- Measurement acquisition period τ_a
- Updating period $\tau_u = N_u \cdot \tau_a$
- Updating instants $t_k = k \cdot \tau_u$
- Observation horizon $T = N \cdot \tau_a$
- Cost function at instant t_k: $J(t_k, \xi)$

Moving-Horizon Observers with Distributed Optimization

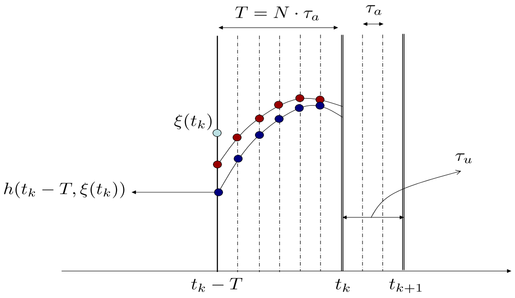
- The system

$$
\begin{aligned}
x(t) &= X(t, t_0, x_0), \\
y(t) &= h(t, x(t)),
\end{aligned}
$$

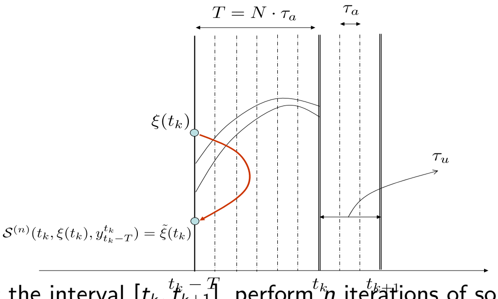


- Measurement acquisition period τ_a
- Updating period $\tau_u = N_u \cdot \tau_a$
- Updating instants $t_k = k \cdot \tau_u$
- Observation horizon $T = N \cdot \tau_a$
- Cost function at instant t_k: $J(t_k, \xi)$

Moving-Horizon Observers with Distributed Optimization



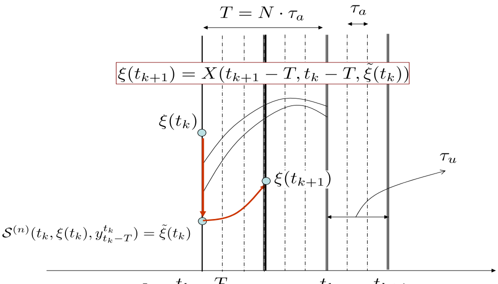
Moving-Horizon Observers with Distributed Optimization



During the interval $[t_k, t_{k+1}]$, perform n iterations of some iterative process $\mathcal{S}$:

$$\tilde{\xi}(t_k) = \mathcal{S}^{(n)}\Big(t_k, \xi(t_k), y_{t_k - T}^{t_k}\Big)$$

Moving-Horizon Observers with Distributed Optimization



During the interval $[t_k, t_{k+1}]$, perform n iterations of some iterative process $\mathcal{S}$:

$$\tilde{\xi}(t_k) = \mathcal{S}^{(n)}\left(t_k, \xi(t_k), y_{t_k-T}^{t_k}\right)$$

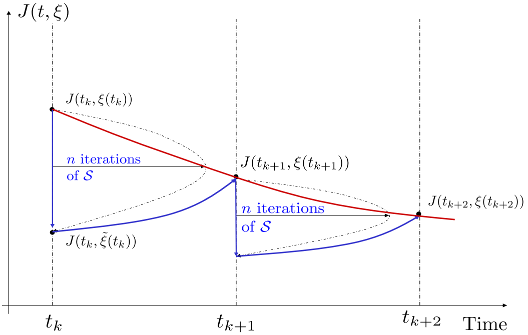
Using $\tilde{\xi}(t_k)$, update the value of $\xi(t_{k+1})$ according to

$$\xi(t_{k+1}) = X(t_{k+1} - T, t_k - T, \tilde{\xi}(t_k))$$

Two opposite processes

The updating mechanism involves two opposite effects on $J(t_k, \xi(t_k))$:

1. A decreasing effect from the n-iterations of the optimization process
2. An increasing effect from the open loop prediction over $\tau_u = N\tau_a$.



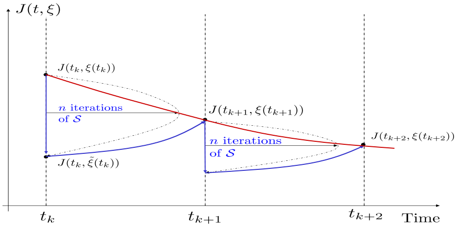
Moving-Horizon Observers with Distributed Optimization

Assumption: Efficiency of the optimizer

The iterative process $\mathcal{S}$ is efficient in the sense that there exists some efficiency map $\alpha_{eff} : \mathbb{N} \rightarrow [0,1[$ such that for all t and ξ, one has:

$$J\Big(t, \mathcal{S}^{(n)}(t,\xi,y_{t-T}^{t})\Big) \leq \alpha_{eff}(n)\cdot J(t,\xi)$$

where $\alpha(\cdot)$ is a decreasing function such that $\alpha(0)=1$. ♡



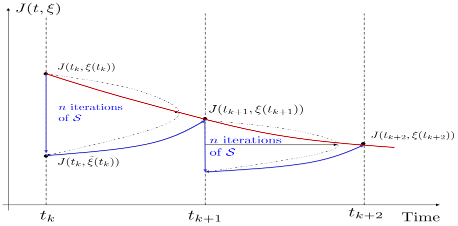
Moving-Horizon Observers with Distributed Optimization

Assumption: Open-loop behavior of the cost function

When using open-loop prediction, the only inequality one can guarantee is given by:

$$J(t+\tau, X(t+\tau-T, t-T, \xi)) \leq \Big[J(t,\xi)\Big] \cdot \vartheta(\tau) \qquad (3)$$

♡



Moving-Horizon Observers with Distributed Optimization

A rather qualitative result

Under the assumptions above, the convergence of the distributed in time optimization based observer is guaranteed provided that the following inequality holds:

$$\varpi(N_u) := \alpha_{eff}\Big(E(\frac{N_u\tau_a}{\tau_{iter}})\Big)\cdot\vartheta(N_u\tau_a) < 1 \tag{4}$$

Moreover, the *convergence time* is given by:

$$t_r(N_u) \approx \Big[\frac{3N_u}{|\log\big(\varpi(N_u)\big)|}\Big]\cdot\tau_a \tag{5}$$

where

- ✓ τ_a is the measurement acquisition period
- ✓ $N_u\tau_a$ is the updating period
- ✓ τ_{iter} is the time necessary to perform one iteration of the process $\mathcal{S}$ ♠

psa-lab

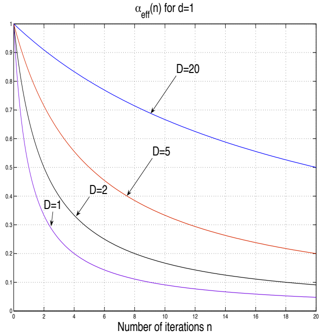
Moving-Horizon Observers with Distributed Optimization

Take the following example for illustration

$$\alpha_{eff}(n) = \frac{D}{n^d + D} \quad ; \quad \vartheta(\tau) = \exp(\beta \cdot \tau)$$

Note that:

- $d \nearrow$ increases the efficiency
- $D \nearrow$ decreases the efficiency
- $\alpha_{eff}(0) = 1$
- $\beta \nearrow$ assumes high model discrepancy



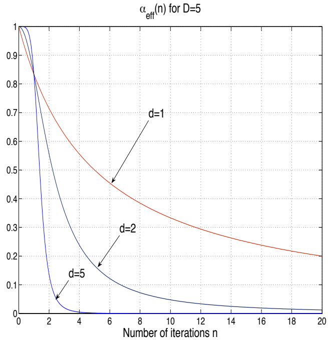
Moving-Horizon Observers with Distributed Optimization

Take the following example for illustration

$$\alpha_{eff}(n) = \frac{D}{n^d + D} \quad ; \quad \vartheta(\tau) = \exp(\beta \cdot \tau)$$

Note that:

- $d \nearrow$ increases the efficiency
- $D \nearrow$ decreases the efficiency
- $\alpha_{eff}(0) = 1$
- $\beta \nearrow$ assumes high model discrepancy



ipsa-lab

Moving-Horizon Observers with Distributed Optimization

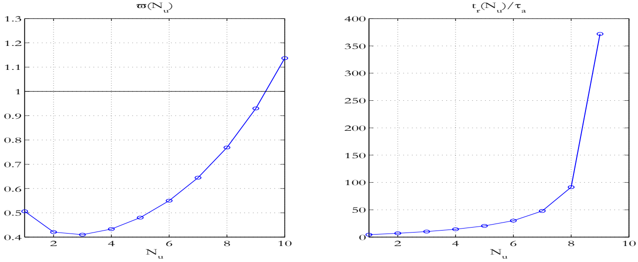


Figure: Evolutions of the stability indicator $\varpi(N_u)$ and the settling time $t_r(N_u)$ vs the number of iterations N_u used to update the state estimation. $D = 3$, $d = 1$, $\beta \cdot \tau_a = 0.3$ and $\tau_a/\tau_{iter} = 5$. Under these conditions, stability cannot be guaranteed when more that 9 iterations are used. The optimal choice (in term of settling time) is the one where only one iteration is used to perform the updating.

Moving-Horizon Observers with Distributed Optimization

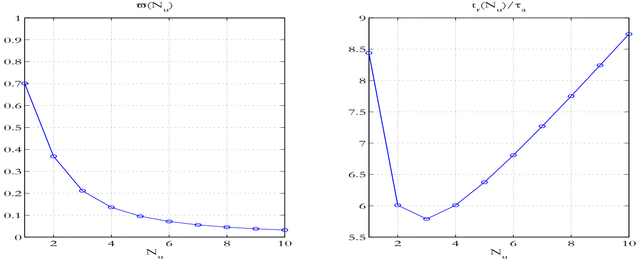


Figure: Evolutions of the stability indicator $\varpi(N_u)$ and the settling time $t_r(N_u)$ vs the number of iterations N_u used to update the state estimation. $D = 50$, $d = 2$, $\beta \cdot \tau_a = 0.05$ and $\tau_a/\tau_{iter} = 5$. Under these conditions, while stability seems guaranteed regardless the number of iterations used to perform the updating, the use of 3 iterations gives the best result in term of settling time.

Moving-Horizon Observers with Distributed Optimization

The success and the quality of the Moving-Horizon-Observer depend on

- The quality of the optimizer (d, D)
- The quality of the model (β)
- The iteration complexity (τ_{iter})
- The problem itself (The very existence of such parameters)

On-line identification of the problem parameters ?

$\Rightarrow$ On-line adaptation of the updating rate ?

Further readings

- Michalska, H & Mayne D. Q. *Moving Horizon Observers and Observer Based Control* IEEE TAC 1995.
- Zimmer G.*State Observation by On-line Optimization* Int. J. of Control 1994.
- Rao C. V., Rawlings J. B. and Mayne D. Q. *Constrained State Estimation* ... IEEE TAC 2003.
- Rao C. V. and Rawlings J. B. *Constrained Process Monitoring: Moving-Horizon* ... AIChE Journal 2004.
- Ohtsuka T. and Fujii H. A. *Nonlinear Receding Horizon* ... J. Guidance, Control, Dynamics 1996.
- Alvarado I. A., Findeisen R., Kuhl P., ... *State Estimation for Repetitive* ... IFAC WC barcelona, 2002.
- Wei Kang *Moving horizon numerical observers of nonlinear control systems* IEEE TAC 2006.
- Alamir M. *Optimization Based Nonlinear Observer Revisited* ... Int. J. of Control 1999.
- Alamir M. and Calvillo-Corona L. *Further Results on Nonlinear Receding Horizon Observers* ... IEEE TAC 2002.
- Ascher U M. *Stabilization of Invariants of Discretized Diff. Syst.* Numerical Algorithms, 1997.
- Othman N., McKenna T. F. and Févotte G. *On-Line Monitoring* ... Polymer Reaction Engineering, 2001.
- Ferrari-Trecate G. Mignone D. and Morari M. *MHE for Hybrid Systems* ... ACC 2000.
- Zavala V. M., Laird C. D. and Biegler L. T. *A Fast Computation Framework for Large Scale MHE* Int. Symp. On Dyn. and Contr. of Process Syst 2007.

General Conclusion

- The progress in MHO $\longleftarrow$ progress of optimization tools
- MHO-related problem is NOT ONLY an optimization problem
- Promising direction:

Combine Analytic and Optimization Based Observers

(Let the MHO concentrate on the structure-free part of the problem)

Example of a selection index

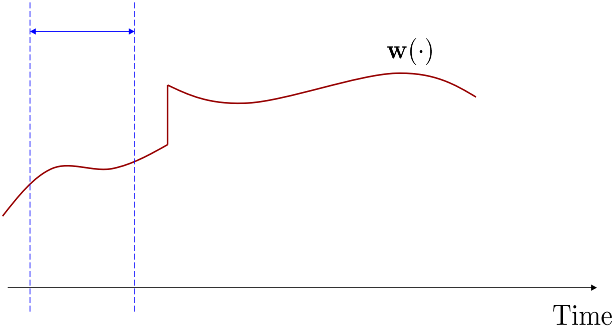
$$J(t,\xi,\mathbf{w}) := \Gamma(t,\xi-\xi^*(t)) + \int_{t-T}^{t} L\Big(\mathbf{w}(\sigma),\varepsilon_y(\sigma)\Big)$$

- $\varepsilon_y(\sigma) = y_{t-T}^t(\sigma) - Y(\sigma, t-T, \xi, \mathbf{w})$ output prediction error
- $\xi^*(t)$ condenses the past knowledge.
- For Kalman filter

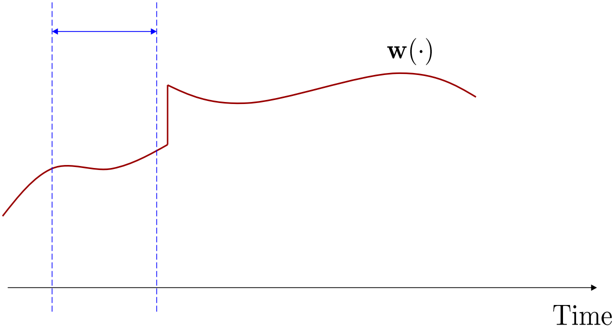
$$L(w,\varepsilon_y) = w^T Q^{-1} w + \varepsilon_y^T R^{-1}\varepsilon_y$$

$\xi^*(t)$ Induced by the past estimate (discrete KF)

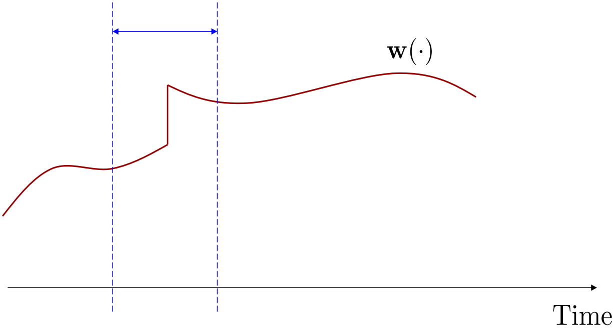
Handling abrupt behavior of uncertainties



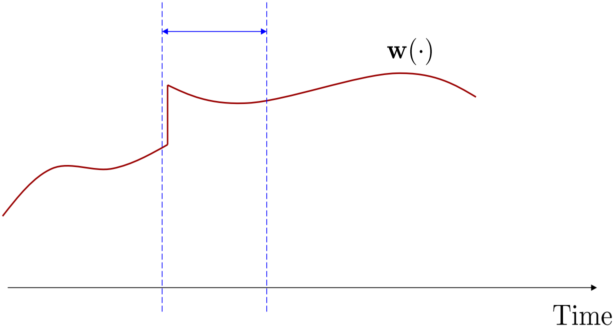
Handling abrupt behavior of uncertainties



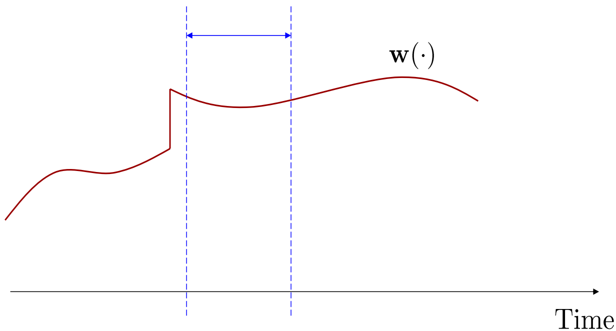
Handling abrupt behavior of uncertainties



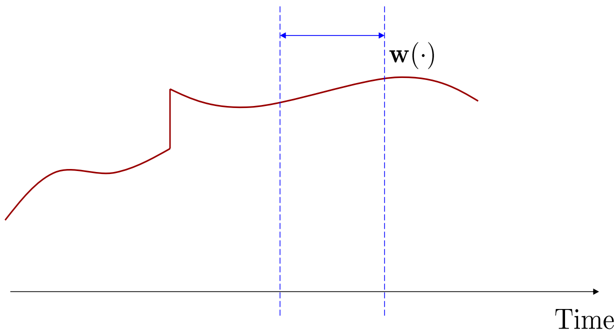
Handling abrupt behavior of uncertainties



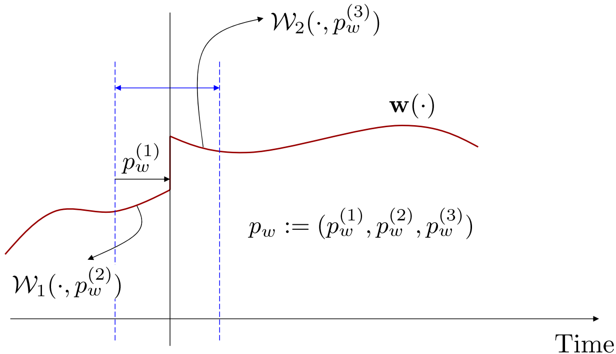
Handling abrupt behavior of uncertainties



Handling abrupt behavior of uncertainties



Non smooth behaviors can be parametrized



$$\mathcal{W}(\cdot, p_w) := \begin{cases} \mathcal{W}_1(\tau, p_w^{(2)}) & \text{if } \tau \le p_w^{(1)} \\ \mathcal{W}_2(\tau, p_w^{(3)}) & \text{otherwise} \end{cases}$$

Definition of the phase II: Existence of monomer droplets

$$N_1\delta_1 + N_2\delta_2 + N_3\delta_3 - \frac{(1-\phi_p^p)}{\phi_p^p}\sigma > 0 \tag{6}$$

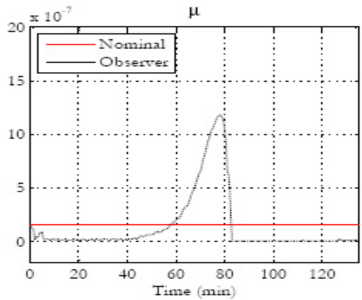
where

$$\delta_i = MWi(\frac{1}{\rho_i} + \frac{(1-\phi_p^p)}{\rho_{i,h}\phi_p^p})\ ,\quad i = 1, 2, 3 \tag{7}$$

and

$$\sigma = \sum_{j=1}^{3} \frac{MW_j N_j^T}{\rho_j, h} \tag{8}$$

Example of dynamic evolution of μ



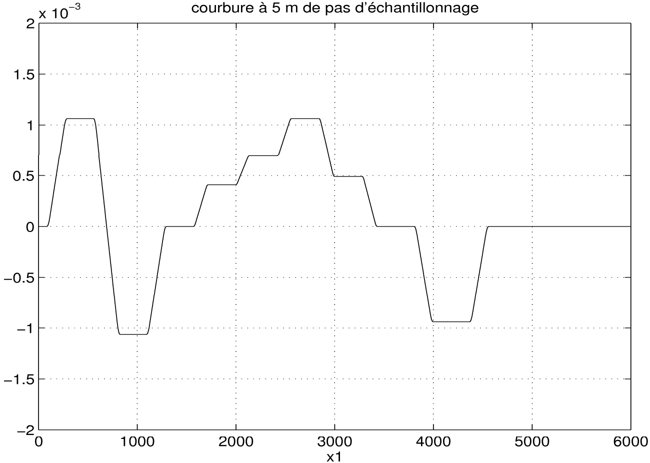
$$\frac{dY}{dt}|_{c\equiv 0}(\tau, t-T, \xi(t)) = Y_{t_2}(\tau, t-T, \xi(t)) + Y_{\xi}(\tau, t-T, \xi(t))f(t-T, \xi(t))$$

Computing $\frac{dY}{dt}|_{c\equiv 0}(\tau, t-T, \xi(t))$ amounts to compute the sensitivity of solutions of ODE's to initial conditions.

Sensitivity Computation

If $X(t, x_0)$ is solution of $\dot{x} = f(x)$ then $\dfrac{\partial X}{\partial x_0}(\tau, x_0)$ is given by $A(\tau) \in \mathbb{R}^{n\times n}$ where

$$\dot{A}(\sigma) = \Big[\frac{\partial f}{\partial x}(X(\sigma, x_0))\Big]A(\sigma) \quad ; \quad A(0) = \mathbb{I}_n$$

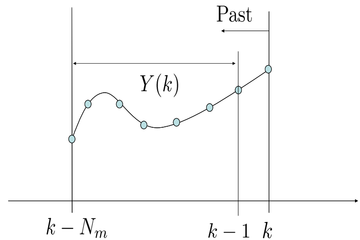


Evolution of th
Toulouse line.

gipsa-lab

Consider a dynamic simulator

$$x(t) = X(t, x_0, p)$$
$$y(t) = h(x(t), p) \in \mathbb{R}$$

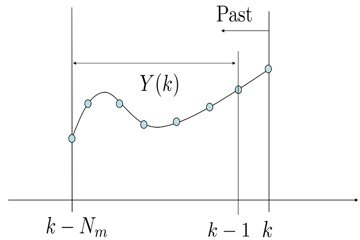


- x state vector
- p a vector of parameter/faults
- y vector of measurements

$$Y(k) := \big(y(k - N_m) \quad \dots \quad y(k-1)\big) \in \mathbb{R}^{N_m}$$

Consider a dynamic simulator

$$x(t) = X(t, x_0, p)$$
$$y(t) = h(x(t), p) \in \mathbb{R}$$



- x state vector
- p a vector of parameter/faults
- y vector of measurements

$$Y(k) := \big(y(k - N_m) \quad \dots \quad y(k-1)\big) \in \mathbb{R}^{N_m}$$

For N_m sufficiently high

$$y(k) = F(Y(k), p)$$

For N_m sufficiently high

$$y(k) = F(Y(k), p)$$

- $\forall p \in \mathbb{P}$, $F(\cdot, p) = F_p(\cdot) : \mathbb{R}^{N_m} \rightarrow \mathbb{R}$
- $F_p(\cdot)$ is unknown
- But $F_p(\cdot)$ is sensitive to variations on p

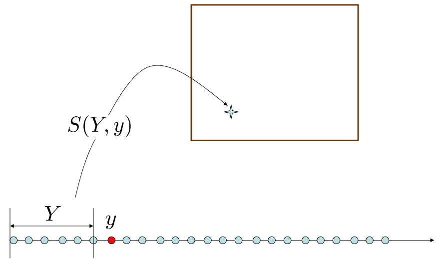
Any graphical signature of $F_p(\cdot)$ is generically sensitive to variations on p.

Graphical signatures generation

Any map

$$S : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^2$$

may be used as a *pensil* to draw a signature of $F_p(\cdot)$ when applied to $(Y, F_p(Y))$



A family of signatures can be defined

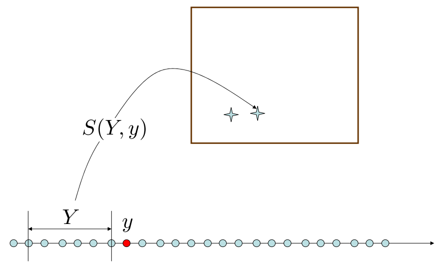
- Either by fixing N and changing S
- Or by *fixing* S and changing $N \in \{n, \ldots, \infty\}$

Graphical signatures generation

Any map

$$S : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^2$$

may be used as a *pensil* to draw a signature of $F_p(\cdot)$ when applied to $(Y, F_p(Y))$



A family of signatures can be defined

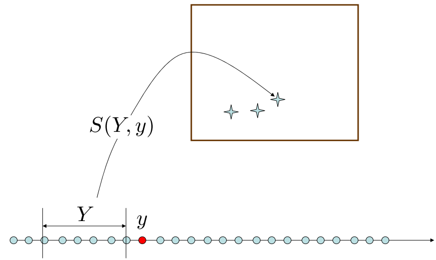
- Either by fixing N and changing S
- Or by *fixing* S and changing $N \in \{n, \dots, \infty\}$

Graphical signatures generation

Any map

$$S : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^2$$

may be used as a *pensil* to draw a signature of $F_p(\cdot)$ when applied to $(Y, F_p(Y))$



A family of signatures can be defined

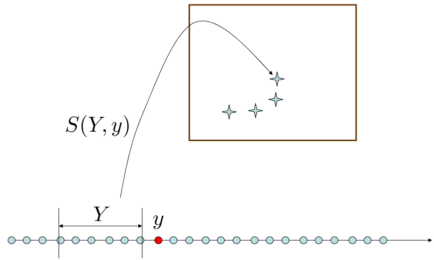
- Either by fixing N and changing S
- Or by *fixing* S and changing $N \in \{n, \dots, \infty\}$

Graphical signatures generation

Any map

$$S : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^2$$

may be used as a *pensil* to draw a signature of $F_p(\cdot)$ when applied to $(Y, F_p(Y))$



A family of signatures can be defined

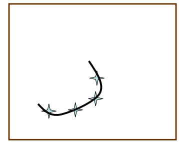
- Either by fixing N and changing S
- Or by *fixing* S and changing $N \in \{n, \dots, \infty\}$

Graphical signatures generation

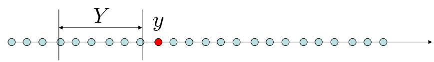
Any map

$$S : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^2$$

may be used as a *pensil* to draw a signature of $F_p(\cdot)$ when applied to $(Y, F_p(Y))$



Signature of $F_p(\cdot)$

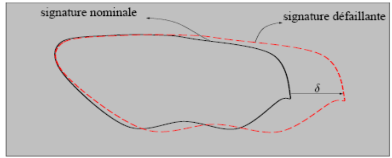


A family of signatures can be defined

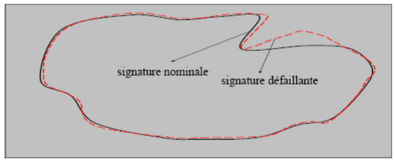
- Either by fixing N and changing S
- Or by *fixing S* and changing $N \in \{n, \ldots, \infty\}$

The key intuition (1)

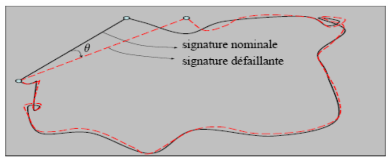
Signature S_{N_1} when p_1 varies



Signature S_{N_2} when p_2 varies



Signature S_{N_3} when p_3 varies



The key intuition (2)

Road map

1. Find as many N_i's as necessary such that all the p_j's are discriminated.
   - The information is somewhere there !
   - There are degrees of freedom
   - Human brain's classification skills
2. Translate geometrical deformations into mathematical expressions
3. Encode it on-line to perform identification and/or diagnosis

Example 1: Parameterized Van-der-Pol oscillator

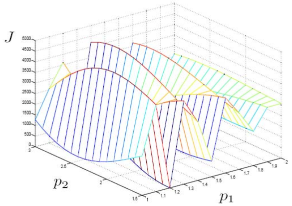
Modified Vand-der-Pol oscillator

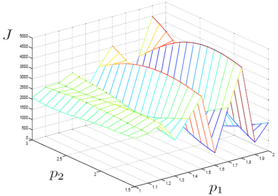
$$\begin{aligned} \dot{x}_1 &= p_1 x_2 \\ \dot{x}_2 &= -9x_1 + p_2(1-(x_1+p_3)^2)x_2 \\ y &= x_1 \end{aligned}$$

with $p \in [1,2] \times [1.5,3] \times [0,0.1]$

$\rightarrow$ See illustrations.

Signatures vs Least Squares identification





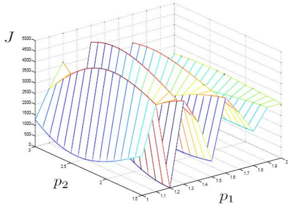
Output prediction error related cost for $p = \begin{pmatrix}1.2 & 1.5 & 0\end{pmatrix}^T$

Output prediction error related cost for $p = \begin{pmatrix}2 & 2.5 & 0.1\end{pmatrix}^T$

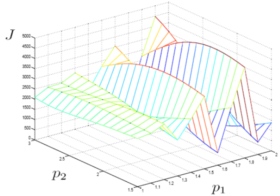
Outp

gipsa-lab

Signatures vs Least Squares identification



Output prediction error related cost for $p = \begin{pmatrix} 1.2 & 1.5 & 0 \end{pmatrix}^T$

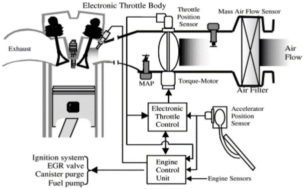


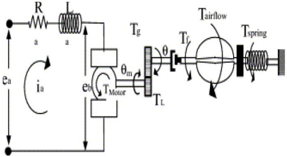
Output prediction error related cost for $p = \begin{pmatrix} 2 & 2.5 & 0.1 \end{pmatrix}^T$

Signatures enable

- Decoupling
- Convexification

Example 2: Electronic Throttle Control System (ETC)





$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \frac{1}{J}\Big[-K_{sp}(x_1+\theta_0) - K_f x_2 + (NK_t)x_3\Big] \\
\dot{x}_3 &= -\frac{1}{L_a}\Big[NK_b x_2 + R_a x_3 + u\Big]
\end{aligned}
$$

Problem: Estimate the coefficients: K_t, R_a, K_b and K_f
Measures: $y = (\theta, i_a)^T$.

$\rightarrow$ See illustrations.

Properties of a signature

- Let S_N be a signature
- Let

$$\mathcal{C}_N = \Big\{\xi_1(i), \xi_2(i)\Big\}_i$$

be the corresponding $2D$ curves.

Properties of a signature

- Let S_N be a signature
- Let

$$\mathcal{C}_N = \Big\{\xi_1(i), \xi_2(i)\Big\}_i$$

be the corresponding $2D$ curves.

- A property of S_N is a scalar function $P(\mathcal{C}_N)$.

Examples

- $P(\mathcal{C}_N) = \max_i\{\xi_j(i)\}$, $std(\xi_j(\cdot)$, $mean(\xi_j(\cdot)$, $\max_i\{\xi_j(i)\} - \min_i\{\xi_j(i)\}$
- But also,

$$x_1(\imath^*) \quad ; \quad \text{where} \quad \imath^* = arg\max_i |\hat{\hat{y}}(i)|$$

- etc.

A pair

$$(S_N, P)$$

of a signature and a property is called a a coordinate

A pair

$$(S_N, P)$$

of a signature and a property is called a a coordinate

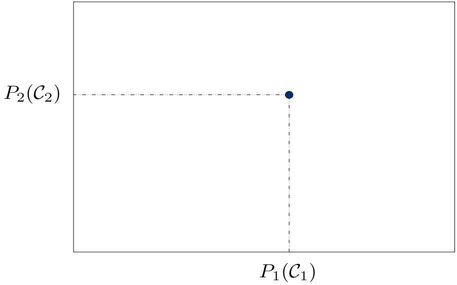
By choosing two coordinates (S_{N_1}, P_1), (S_{N_2}, P_2), an *experiment* can be represented by one point in the $2D$ plane

A pair

$$(S_N, P)$$

of a signature and a property is called a a coordinate

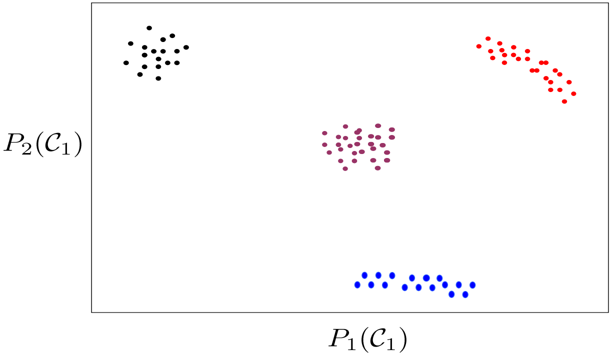
By choosing two coordinates (S_{N_1}, P_1), (S_{N_2}, P_2), an *experiment* can be represented by one point in the $2D$ plane



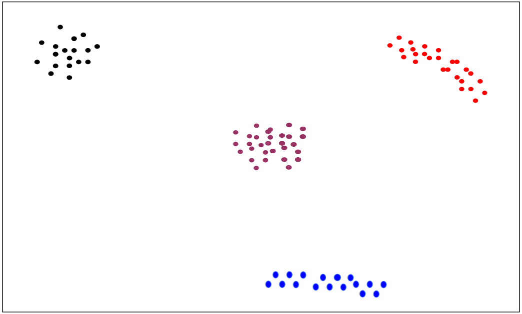
Using signature for faults classification

$$
\begin{array}{lcl}
y^{(1,1)}(\cdot),\dots,y^{(1,n_1)}(\cdot) & ; & \text{experiments / config 1} \\
y^{(2,1)}(\cdot),\dots,y^{(2,n_2)}(\cdot) & ; & \text{experiments / config 2} \\
y^{(3,1)}(\cdot),\dots,y^{(3,n_3)}(\cdot) & ; & \text{experiments / config 3} \\
y^{(4,1)}(\cdot),\dots,y^{(4,n_4)}(\cdot) & ; & \text{experiments / config 4}
\end{array}
$$

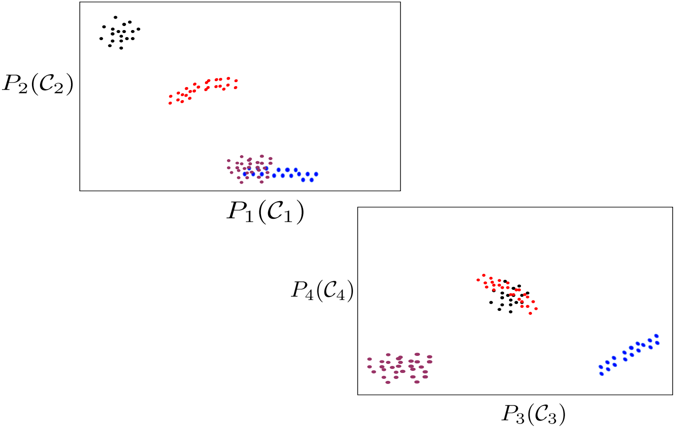
- These can be obtained
    - Either using real data collected after fault occurrence
    - Or using faithful models representing faulty configurations


$P_2(\mathcal{C}_1)$
$P_1(\mathcal{C}_1)$

$P_2(\mathcal{C}_2)$



$P_1(\mathcal{C}_1)$


$P_2(\mathcal{C}_2)$
$P_1(\mathcal{C}_1)$
$P_4(\mathcal{C}_4)$
$P_3(\mathcal{C}_3)$

DiagSign

- Generic software
- Fast Prototyping of classification algorithms
- Automated/Manual Modes
- Cheap Computations

See more on www.diagsign.com