Data Assimilation: A Deterministic Vision – Theory and Applications
Least-square estimation

Philippe Moireau
Inria Project-Team MΞDISIM

Inria Université Paris-Saclay
LMS, Ecole Polytechnique, CNRS

Outline

Few words on inverse problems

The least square approach

Data assimilation seen as an inverse problem

The variational approach – the so-called 4D-Var

The Kalman estimator formalism

Kalman in an infinite dimension context

The discrete-time Kalman filter as a time discretization of the continuous-time Kalman estimator

Non-linear optimal estimator

Reduced Order Optimal Filtering approaches

Beyond the deterministic approach

Few words on inverse problems

Inverse problems in the sense of Hadamard

- Problem setting: Given z, find y such that

$$z = \Psi(y),$$

  where

  - State space: Separable Hilbert $\mathcal{Y}, (\cdot,\cdot)_{\mathcal{Y}}$;
  - Observation space: Separable Hilbert $\mathcal{Z}, (\cdot,\cdot)_{\mathcal{Z}}$.

- A "well-posed" problem:
  - There exists at least one solution – *i.e.* surjectivity of Ψ;
  - There exists at most one solution – *i.e.* injectivity of Ψ;
  - The solution is continuous with respect to the data,

$$\|z^{\delta} - \check{z}\|_{\mathcal{Z}} \leq \delta \Rightarrow \|y - \check{y}\|_{\mathcal{Y}} \leq \alpha^{\delta}.$$

One example (not so far from biology), i

- Model: From u_0 and with $u(\ell) = 0$

$$\partial_t u(t,x) - b\partial_x u(x,t) = 0, \quad 0 \le x \le \ell, \quad 0 \le t \le T.$$

- Observations:

$$\mu_k(t) = \int_0^\ell x^k u(t,x)dx, \quad 0 \le t \le T.$$

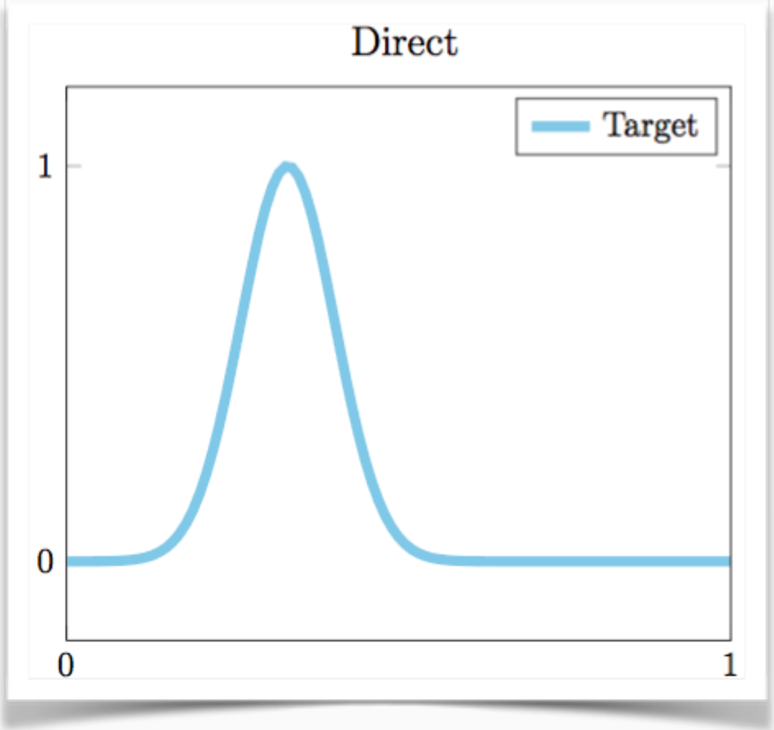
- Operator: $\Psi_{|T}$

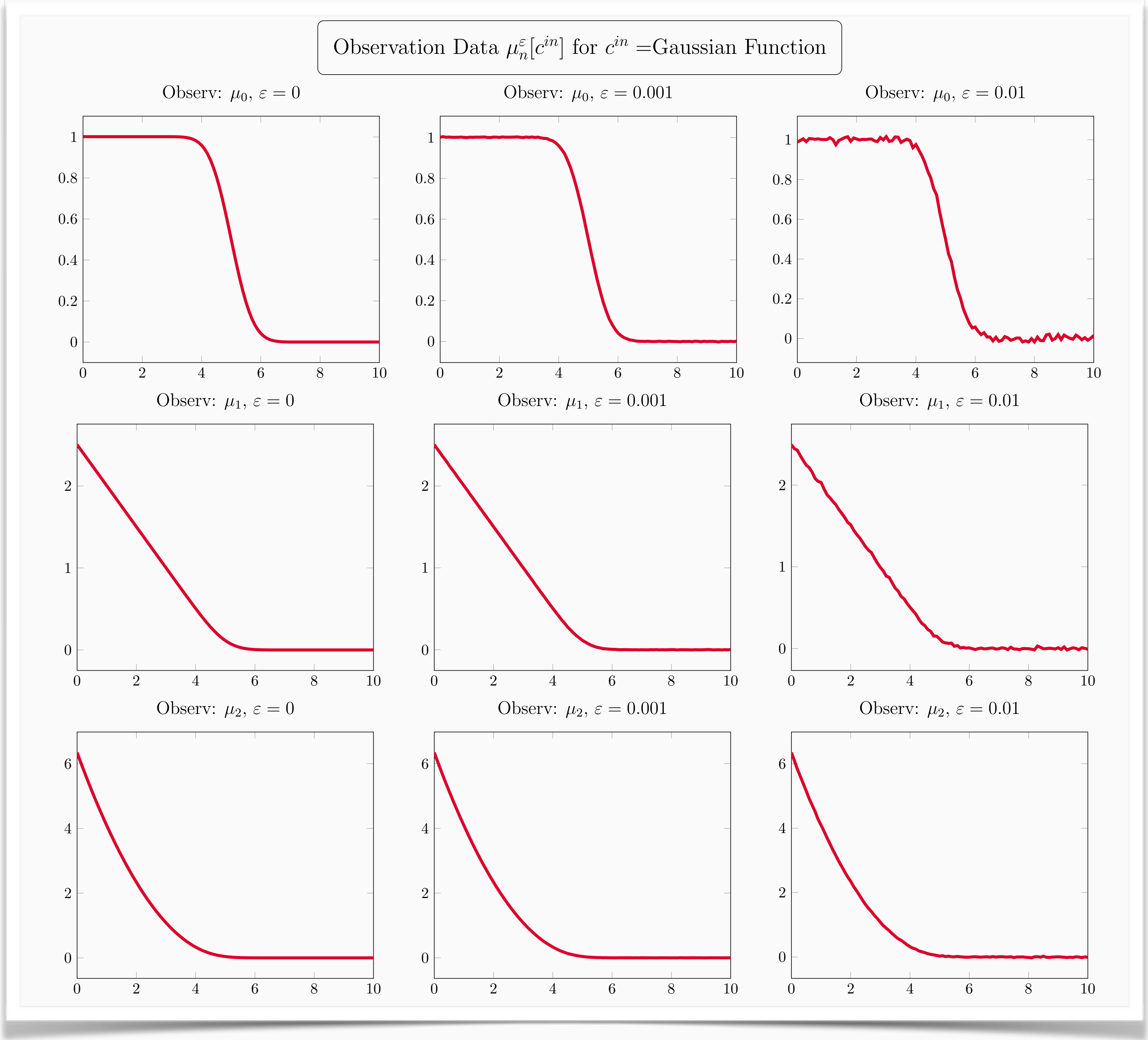
$$\Psi_{|T}(u_0) = \int_0^\ell x^k u(t,x)\,\mathrm{d}x = \int_0^\ell x^k u_0(x+bt)\,\mathrm{d}x = \int_{bt}^\ell (x-bt)^k u_0(x)\,\mathrm{d}x$$

Hence

$$\frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}}(\Psi_{|T} u_0)(t) = (-b)^{k+1} k! u_0(bt).$$

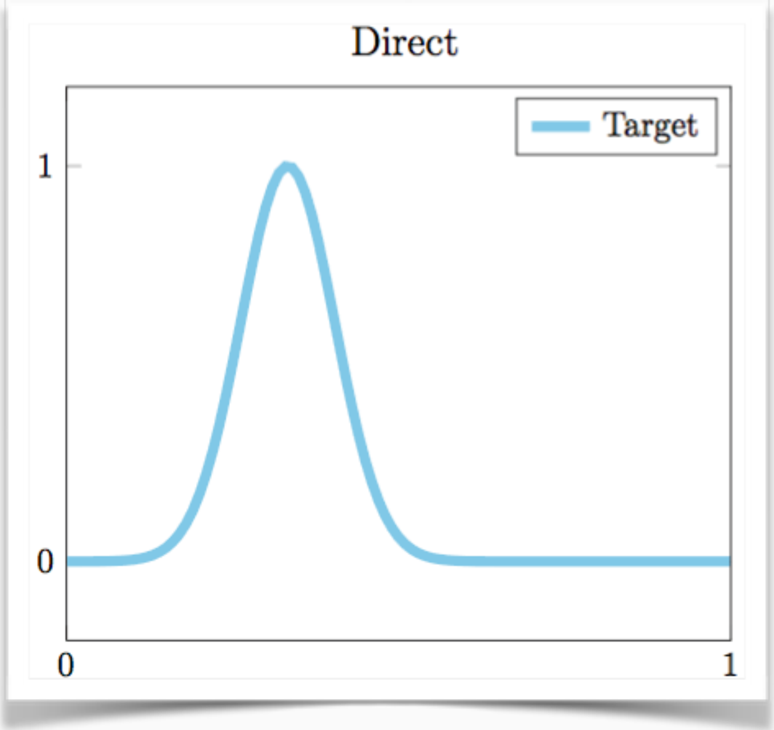
One example (not so far from biology), ii

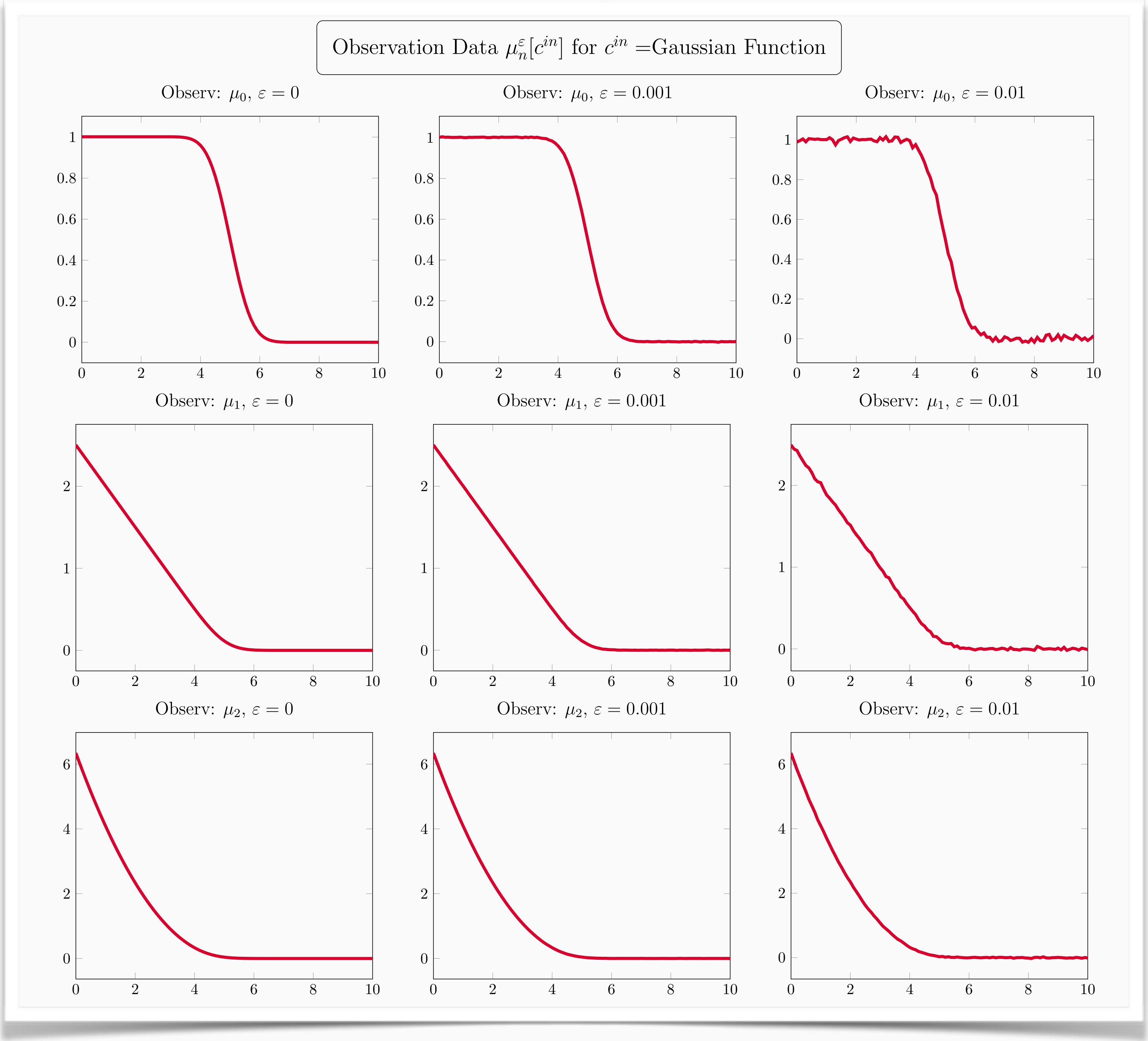




(Armiento et al 2017)

One example (not so far from biology), ii





(Armiento et al 2017)

One example (not so far from biology), iii

- Conclusion:
  - When $bT \geq \ell$

    The observation z is related to $k+1$-th derivative of the state $y(= u_0)$

    $$\mathrm{Im}(\Psi_{|T}) := \left\{z^\delta \in H^{k+1}([0,T]), \qquad z^\delta(T) = \cdots = z^{(k)}(T) = 0\right\},$$

    Therefore Ψ is compact with dense image, injective but not surjective.

  - When $bT \leq \ell$, the operator is not even injective

The least square approach

- Minimizing

$$\mathscr{J}(y) = \frac{1}{2\alpha^2}\|y - y_\diamond\|^2_{\mathcal{Y}} + \frac{1}{2\delta^2}\|z^\delta - \Psi(y)\|^2_{\mathcal{Z}}$$
$$= \frac{1}{2}\|y - y_\diamond\|^2_{\mathcal{Y},\Lambda_\diamond} + \frac{1}{2}\|z^\delta - \Psi(y)\|^2_{\mathcal{Z},R}.$$

- Weight Operators $\Lambda_\diamond = \frac{1}{\alpha^2}\mathbb{1}$, $R = \frac{1}{\delta^2}\mathbb{1}$,

$$(z_1, z_2)_{\mathcal{Z},R} = (z_1, Rz_2)_{\mathcal{Z}}, \quad (y_1, y_2)_{\mathcal{Y},\Lambda_\diamond} = (y_1, \Lambda_\diamond y_2)_{\mathcal{Y}}.$$

- Optimality condition:

$$\forall \delta y \quad \mathrm{d}\mathscr{J}(y)(\delta y) = (y_\diamond - y, \Lambda_\diamond \delta y)_{\mathcal{Y}} + (z^\delta - \Psi(y), R\,\mathrm{d}\Psi(y)\delta y)_{\mathcal{Z}} = 0.$$

- Then the classical Tikhonov approach gives in the linear case

$$\left(\frac{1}{\delta^2}\Psi^*\Psi + \frac{1}{\alpha^2}\mathbb{1}\right)\bar{y}^{\delta,\alpha} = \frac{1}{\delta^2}\Psi^* z^\delta + \frac{1}{\delta^2}\bar{y}_\diamond.$$

Least square estimation ii

- A more general formulation in the general (non-linear) case: replace $\frac{1}{\alpha^2}\mathbb{1}$ by $\Lambda_\diamond = B^* B$.
- A practical algorithm: a gradient-descent method

$$y_{k+1} = y_k - \rho_k^{\delta,\alpha} \nabla \mathscr{J}^{\delta,\alpha}(y_k)$$

- Example: Ψ is linear compact with dense image, injective but not surjective: With a constant step $\rho^{\delta,\alpha} = \frac{\alpha^2}{\delta^2+\alpha^2\sigma_0^2},$, where σ_0 is the first eigenvalue of the Ψ SVD.

Recursive least square estimation i

Let's assume that each day, we have a new data z_n^δ, with linear operator Ψ_n

- Augmented Spaces:

$$
{}^\natural z = \begin{pmatrix} y_\diamond \\ z_0 \\ \vdots \\ z_n \end{pmatrix}, \quad {}^\natural\Psi_n = \begin{pmatrix} \mathbb{1} \\ \Psi_0 \\ \vdots \\ \Psi_n \end{pmatrix}, \quad {}^\natural R = \begin{pmatrix} \Lambda_\diamond & & & 0 \\ & R_0 & & \\ & & \ddots & \\ 0 & & & R_n \end{pmatrix}.
$$

- the LSE is given by $\bar{y}_n = \underbrace{\left(\Lambda_\diamond + \sum_{k=0}^{n} \Psi_k^* R_k \Psi_k\right)^{-1}}_{\Pi_n} \left(\Lambda_\diamond y_\diamond + \sum_{k=0}^{n} \Psi_k^* R_k z_k^\delta\right)$.

Lemma (Lemma of inversion)

Under suitable obvious conditions

$$
(M_1 - M_{12} M_2^{-1} M_{21})^{-1} = M_1^{-1} + M_1^{-1} M_{12} (M_2 - M_{21} M_1^{-1} M_{12})^{-1} M_{21} M_1^{-1}.
$$

Recursive least square estimation ii

- Recursive relationships on the covariance operator:

$$\begin{aligned}\Pi_n &= (\Psi_n^* R_n \Psi_n + \Pi_{n-1}^{-1})^{-1} \\ &= \Pi_{n-1} - \Pi_{n-1}\Psi_n^*(W_n + \Psi_n \Pi_{n-1}\Psi_n^*)^{-1}\Psi_n \Pi_{n-1}\end{aligned}$$

- Recursive estimator ($W_n = R_n^{-1}$):

$$\begin{aligned}&\bar{y}_n = \bar{y}_{n-1} + G_n(z_n^\delta - \Psi_n \bar{y}_{n-1}) \\ &\text{with } G_n = \Pi_{n-1} C_n^*(W_n + C_n \Pi_{n-1} C_n^\intercal)^{-1} = \Pi_n C_n^* R_n\end{aligned}$$

Data assimilation seen as an inverse problem

Formalism (hence formal) i

Model: **A dynamical system**

$$\begin{cases} \dot{y}_{|\zeta,\nu}(t) = A(y,t) + B\nu(t), \ t \in [0,T] \\ y_{|\zeta,\nu}(0) = y_\diamond + \zeta \end{cases}$$

where

$$\|\check{y}(0) - y_\diamond\|_{\mathcal{Y}} = \|\check{\zeta}\|_{\mathcal{Y}} \leq \alpha$$

and

$$\|\dot{y}_{|\zeta,\nu} - A(\check{y},t)\|_{\mathcal{Y}} \propto \|\check{\nu}(t)\|_{\mathcal{U}} \leq \beta$$

Observations

$$z^\delta(t) = C(\check{y}(t),t) + \eta(t), \quad t \in [0,T]$$

where

$$\|z^\delta - \check{z}\|_{\mathcal{Z}} = \|\check{\eta}\|_{\mathcal{Z}} \leq \delta$$

The least square criterion

(Non-linear) operator:

$$\Psi_T : \left| \begin{aligned} \mathcal{Q} = \mathcal{Y} \times \mathcal{U}_T &\to \mathcal{Z}_T \\ \xi = (\zeta, \nu) &\mapsto z \end{aligned} \right.$$

And we minimize $\min_{\xi} \mathscr{J}(\xi) = \frac{1}{2}\|\xi - \xi_\diamond\|^2_{\mathcal{Q},\Lambda} + \frac{1}{2}\|z^\delta - \Psi(\xi)\|^2_{\mathcal{Z},R}$

Least square estimation

We minimize – under the "constraint" of the dynamics – a criterion $\mathscr{J}^{\alpha,\beta,\delta}_{|T}$

$$\min_{\zeta,\nu} \Bigg\{ \mathscr{J}_{|T}(\zeta,\nu) = \frac{1}{2}\|\zeta\|^2_{\mathcal{Y},\Lambda_\diamond} + \frac{1}{2}\int_0^T \|\nu(t)\|^2_{\mathcal{U},S}\,\mathrm{d}t + \frac{1}{2}\int_0^T \|z^\delta(t) - C(y_{|\zeta,\nu}(t))\|^2_{\mathcal{Z},R}\,\mathrm{d}t \Bigg\},$$

The swiss knife

- Every operator can be time-dependent
- The formalism encompasses identification problems:
  – Augmented state ${}^{\natural}y = (y, \theta)^{\intercal}$

$$\begin{cases} \dot{y}_{|{}^{\natural}\zeta,\nu} = A(y,\theta,t) + B\nu, \;\; t \in [0,T] \\ \dot{\theta}_{|{}^{\natural}\zeta,\nu} = 0, \;\; t \in [0,T] \\ y_{|{}^{\natural}\zeta,\nu}(0) = y_{\diamond} + \zeta_y \\ \theta_{|{}^{\natural}\zeta,\nu}(0) = \theta_{\diamond} + \zeta_{\theta} \end{cases}$$

- Observations can have a more complex structure:

$$J(z^{\delta}(t), \check{y}(t)) = \eta, \quad t \in [0,T]$$

- Dynamics, observations and errors can be discrete in time...
  $\rightarrow$ (see slides on discretization aspects)

But the devil is in the detail

- Linear VS Non-linear (even for a linear dependency w.r.t parameters)
- Spaces: for instance for the state:

$$L^2([0,T],\mathcal{Y}),\ C_0([0,T],\mathcal{Y}),\ C_1([0,T],\mathcal{Y}) \text{ or } C_0^a([0,T],\mathcal{Y})^1,\ H^1((0,T);\mathcal{Y})?$$

- Bounded or unbounded operators A, C, B: Essentially
  - ODE (A bounded) VS PDE (A unbounded).
  - Intern. obs. (C bounded?) VS Bound. obs. (C unbounded?).

[1] $f \in C_0^a([0,T],\mathcal{Y})$ if $\exists g \in L^1([0,T]) \mid f(t) - f(0) = \int_0^t g(s)\,\mathrm{d}s$

The variational approach – the so-called 4D-Var

Let $\mathcal{Y}_T = H^1((0,T);\mathcal{Y}), \quad \mathcal{U}_T = L^2((0,T);\mathcal{Y}$

Lagrangian

Let $(\bar{y},\bar{\nu},\bar{q})$ be a satisfying the saddle-point optimality condition of

$$\inf_{\substack{y\in\mathcal{Y}_T,\\ \nu\in\mathcal{U}_T}} \sup_{q\in\mathcal{Y}_T} \left\{ \mathscr{L}(y,\nu,q) = \mathscr{J}_{|T}(y-y(0),\nu) + \int_0^T (\dot{y} - A(y,t) - B\nu, q)_{\mathcal{Y}} \,\mathrm{d}t \right\}$$

Then $(\bar{y}(0) - y_\diamond, \bar{\nu})$ is a minimum of of $\mathscr{J}_{|T}$.

Proof.

Remember: $J(z^\delta, y)$ plays the role of $z^\delta - C(y)$.

Let's $\delta y \in \mathcal{D}([0,T];\mathcal{Y})$, then

$$\begin{aligned}\langle \partial_y \mathscr{L}, \delta y\rangle &= \int_0^T (J(z^\delta,\bar{y}),\, \mathrm{d}_y J(z^\delta,\bar{y})(\delta y))_{\mathcal{Z}} - (\dot{\bar{q}},\delta y)_{\mathcal{Y}} - (\,\mathrm{d}A(\bar{y},t)(\delta y),\bar{q})_{\mathcal{Y}} \,\mathrm{d}t \\ &= 0 \qquad \Rightarrow \dot{\bar{q}} + \mathrm{d}A(\bar{y},t)^*\bar{q} = \mathrm{d}_y J(z^\delta,\bar{y})^* J(z^\delta,\bar{y}).\end{aligned}$$

Proof cont'd i.

Let's take now $\delta y \in H^1_R([0,T];\mathcal{Y})$, then $\langle \partial_y \mathscr{L}, \delta y\rangle = 0 \Rightarrow \bar{q}(T) = 0$.
Let's take now $\delta y \in H^1_L([0,T];\mathcal{Y})$, then with $\mathit{\Pi}_\diamond = \Lambda_\diamond^{-1}$,

$$\langle \partial_y \mathscr{L}, \delta y\rangle = (\bar{y}(0) - y_\diamond, \Lambda_\diamond \delta y(0))_{\mathcal{Y}} + \Big[(\bar{q}, \delta y)\Big]_0^T = 0 \Rightarrow \bar{y}(0) = y_\diamond + \mathit{\Pi}_\diamond \bar{q}(0).$$

With $\delta\nu \in L^2([0,T];\mathcal{U})$, we get with $Q = S^{-1}$, $\langle \partial_\nu \mathscr{L}, \delta\nu\rangle = 0 \Rightarrow \bar{\nu} = QB^*\bar{q}$

Finally taking $\delta q \in \mathcal{D}([0,T];\mathcal{Y})$, then

$$\langle \partial_q \mathscr{L}, \delta q\rangle = \int_0^T (\dot{\bar{y}}, \delta q)_{\mathcal{Y}} - (A(\bar{y},t), \delta\bar{q})_{\mathcal{Y}} \,\mathrm{d}t = 0 \Rightarrow \dot{\bar{y}} = A(\bar{y},t) + BQB^*\bar{q}.$$

Proof cont'd ii.

Using this two-ends dynamics

$$\begin{cases} \dot{\bar{y}}_{|T} = A(\bar{y}_{|T}, t) + BQB^* \bar{q}_T, & t \in (0, T) \\ \dot{\bar{q}}_{|T} + \mathrm{d}A(\bar{y}_{|T}, t)^* \bar{q}_{|T} = -\mathrm{d}C(\bar{y}_{|T}, t)^* R(z^\delta - C(\bar{y}_{|T}, t)), & t \in (0, T) \\ \bar{y}_{|T}(0) = y_0 + \Pi_\diamond \bar{q}_{|T}(0) \\ \bar{q}_{|T}(T) = 0 \end{cases}$$

We get for all $(\delta\zeta, \delta\nu) \in \mathcal{Y} \times L^2((0, T); \mathcal{U})$,

$$\begin{aligned} \langle \mathrm{d}\mathscr{J}_{|T}(\bar{\zeta}, \bar{\nu}), \delta\zeta \otimes \delta\nu \rangle = &(\Lambda_\diamond \bar{\zeta}, \delta\zeta)_\mathcal{Y} + (\bar{q}(0), \delta\zeta)_\mathcal{Y} \\ &+ \int_0^T (S\delta\nu(t), \bar{\nu}(t))_\mathcal{U} + (\bar{q}(t), B\delta\nu(t))_\mathcal{Y} \, \mathrm{d}s = 0. \end{aligned}$$

The linear case

- The criterion reads

$$\mathscr{J}_{|T}(\zeta,\nu) = \frac{1}{2}\|\zeta\|^2_{\mathcal{Y},\Lambda_\diamond} + \frac{1}{2}\int_0^T \left(\|z^\delta(s) - Cy_{|\zeta,\nu}(s)\|^2_{\mathcal{Z},R} + \|\nu(s)\|^2_{\mathcal{U},R}\right)\,\mathrm{d}s,$$

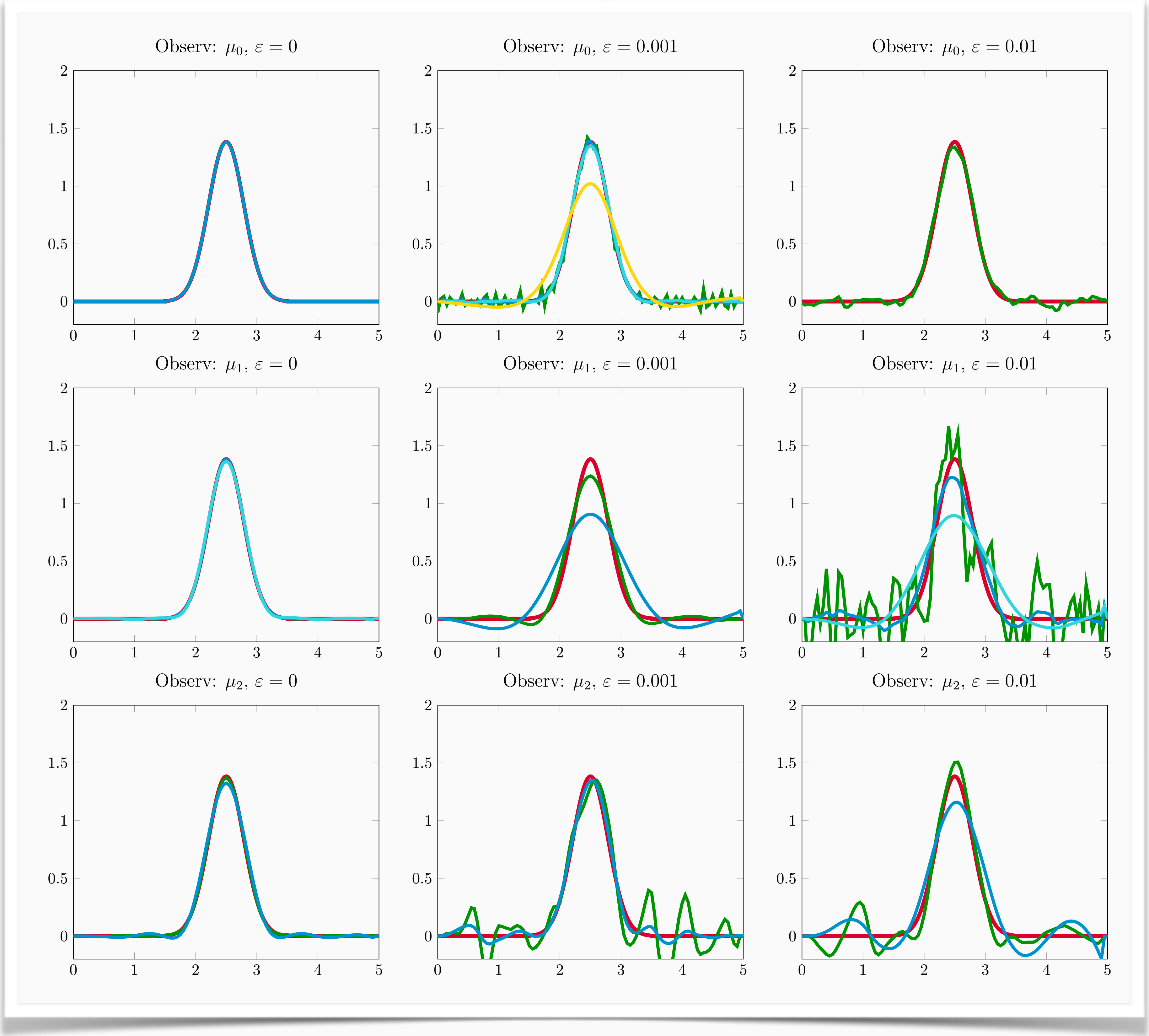
- $\mathscr{J}_{|T}$ is a quadratic functional in the Hilbert space $\mathcal{Y}\times L^2((0,T);\mathcal{U})$ with

$$\exists c_{\text{st}} > 0,\quad \forall(\bar{\zeta},\bar{\nu}) \in \mathcal{Y}\times L^2((0,T);\mathcal{U}),\quad \mathscr{J}(\zeta,\nu) \geq c_{\text{st}}\left(\|\zeta\|^2_{\mathcal{Y}} + \|\nu\|^2_{L^2((0,T);\mathcal{U})}\right).$$

- Therefore, there exists one, and only one, optimal estimation $(\bar{\zeta},\bar{\nu}) = \operatorname{argmin}\mathscr{J}(\zeta,\nu)$.
- The optimality system then reads

$$\begin{cases}\dot{\bar{y}}_{|T} = A\bar{y}_{|T} + BQB^*\bar{q}_{|T}, & \text{in } (0,T)\\ \dot{\bar{q}}_{|T} + A^*\bar{q}_{|T} = -C^*R(z^\delta - C\bar{y}_{|T}(t)), & \text{in } (0,T)\\ \bar{y}_{|T}(0) = y_\diamond + \mathit{\Pi}_\diamond\bar{q}_{|T}(0) & \\ \bar{q}_{|T}(T) = 0 & \end{cases}$$

Example



(Armiento et al 2017)

The Kalman estimator formalism

Optimal sequential estimation in a linear context & in finite dimension

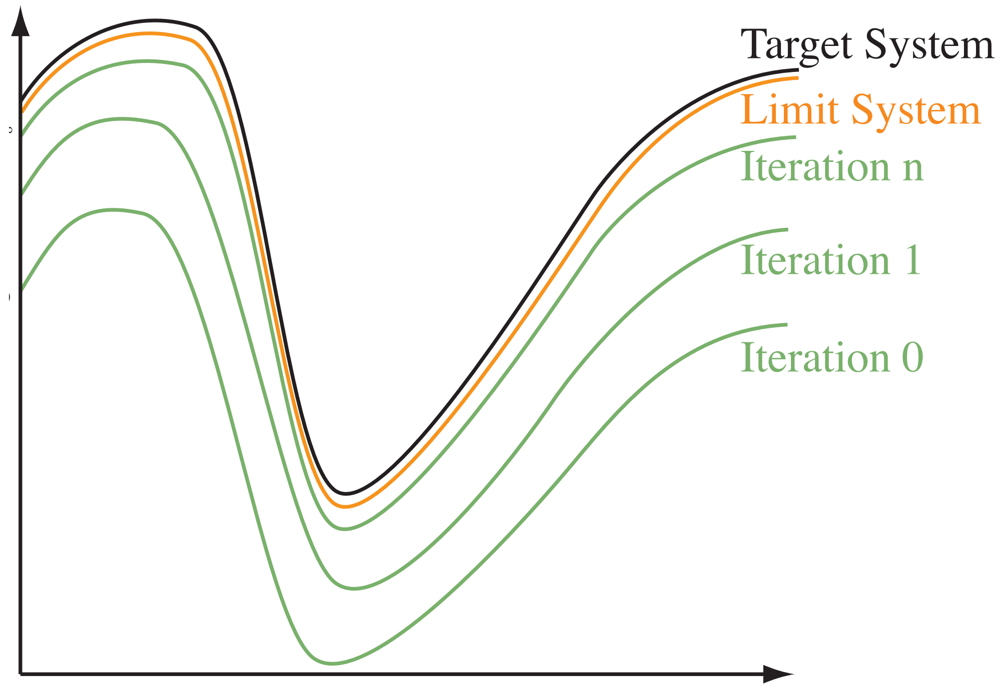
- *Assumption:* $\dim \mathcal{Y} < \infty$; A linear (bounded) and $\sup_{[0,T]} \|A(t)\| \leq c_{\mathrm{st}}$
- Cauchy-Lipchitz gives a solution $y \in C^a([0,T];\mathcal{Y})$
- One unique minimizer for $\mathscr{J}_{|T} \Rightarrow \bar{y} = y_{|\bar{\zeta},\bar{\nu}}$

Definition (Optimal estimator)

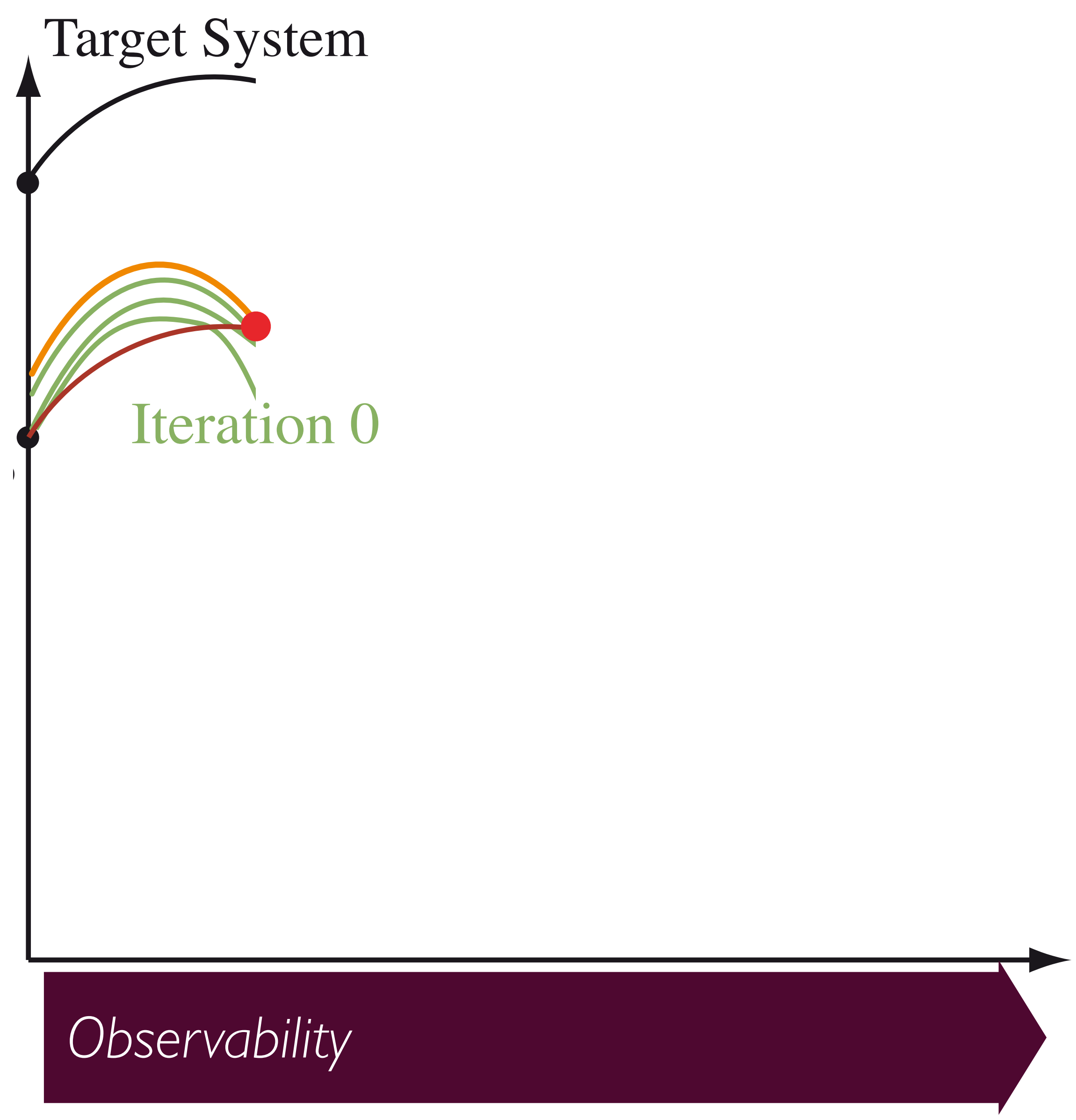
The optimal sequential estimator is defined by

$$\forall t \in [0,T], \quad \hat{y}(t) = \bar{y}_{|t}(t).$$

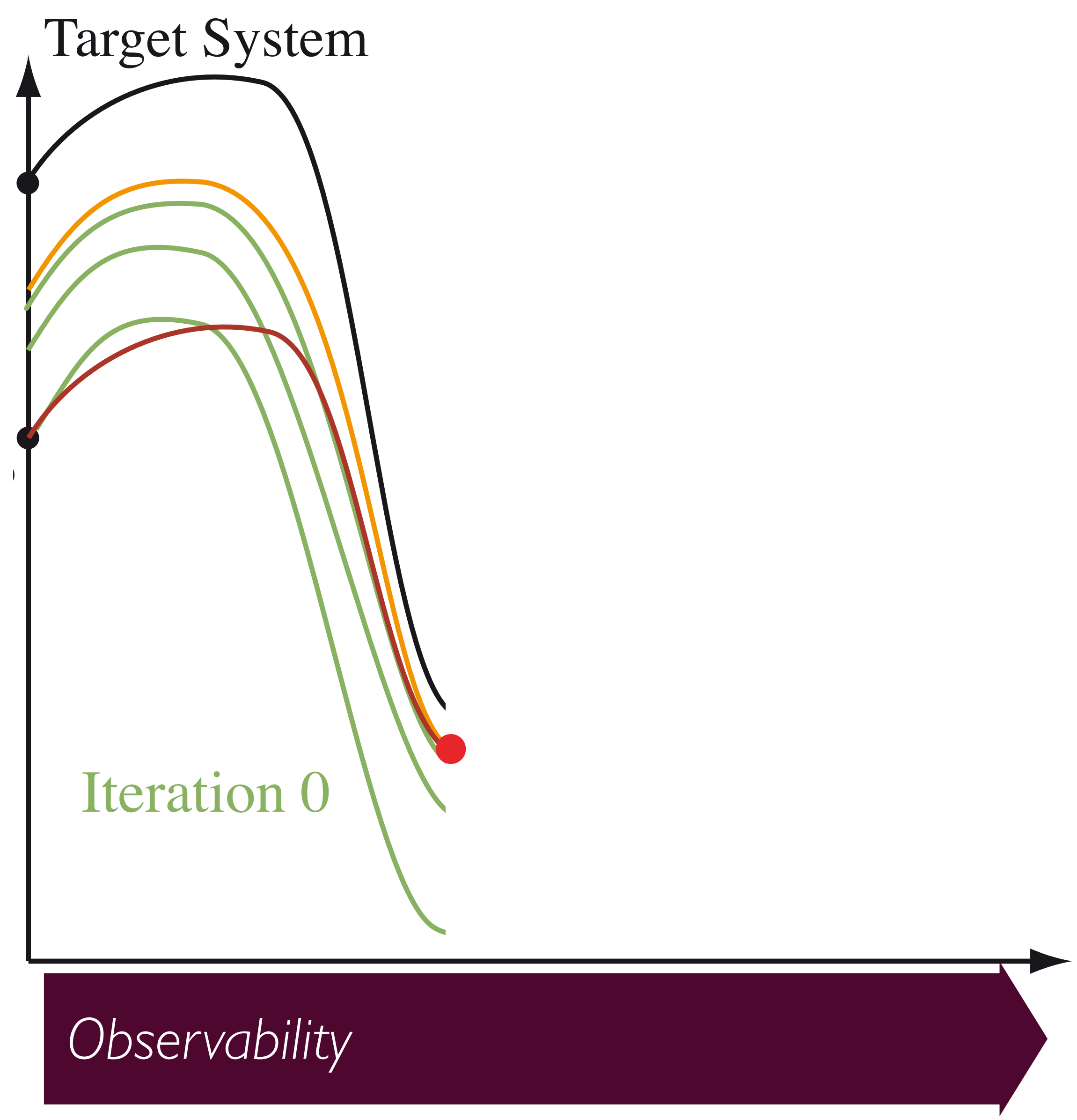
Illustration



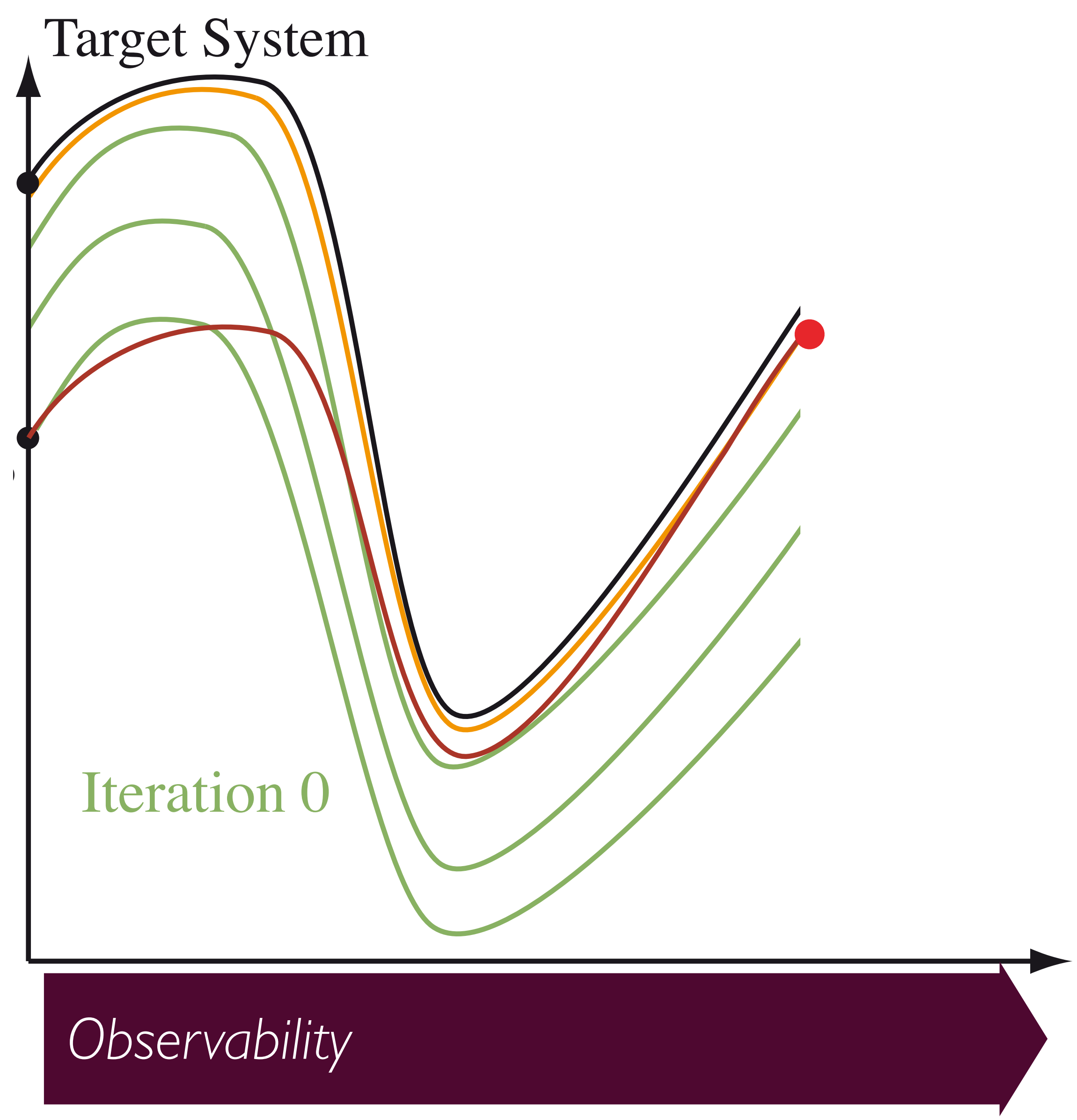
Illustration



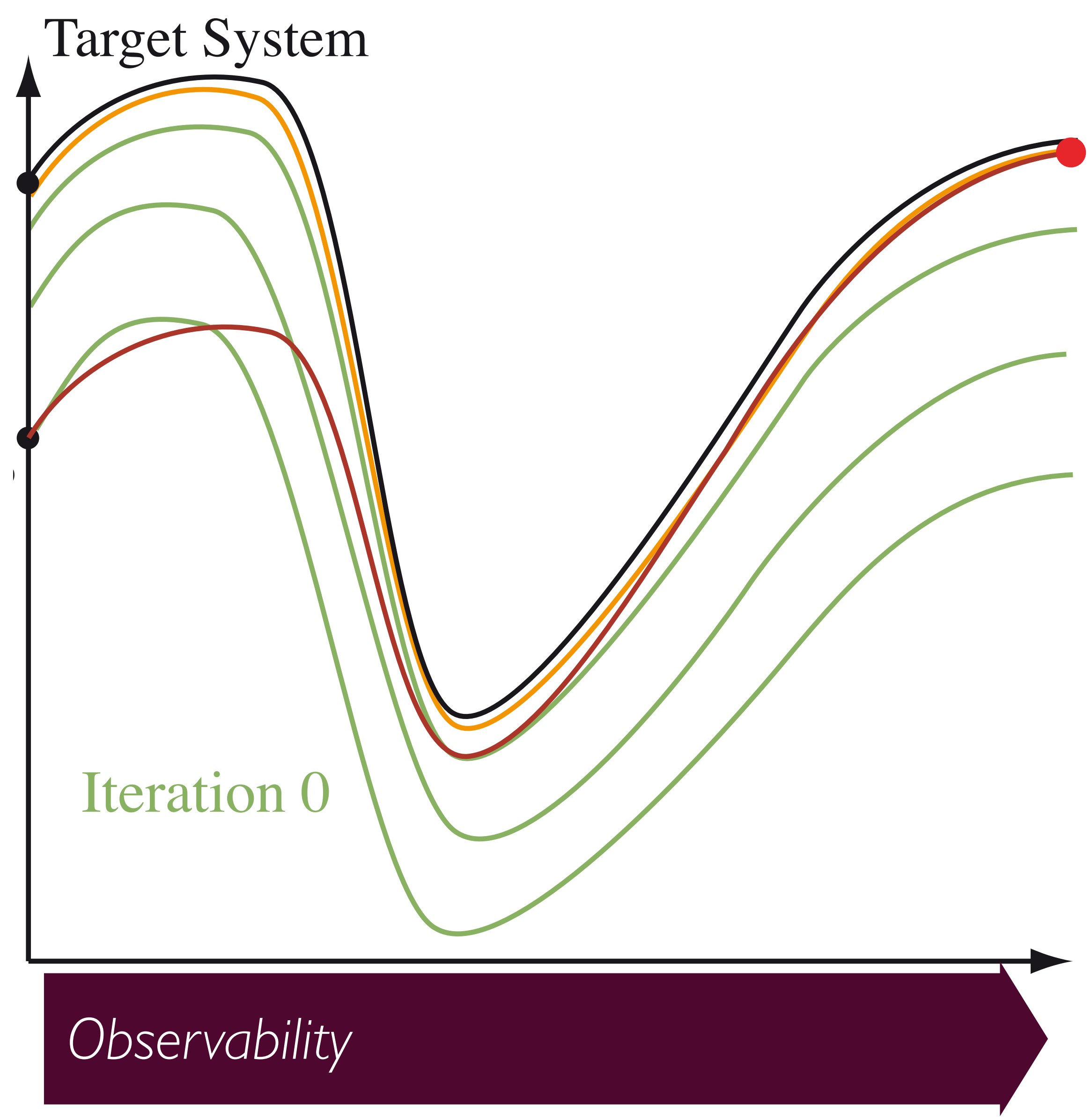
Illustration



Illustration



Illustration



Riccati

Definition (Riccati dynamics)

We call *Covariance* operator, the solution of the following Riccati dynamics

$$\begin{cases} \dot{\Pi} = A\Pi + \Pi A^* - \Pi C^* R C \Pi + B^* Q B \\ \Pi(0) = \Pi_0. \end{cases} \tag{1}$$

Theorem

There exists a unique Covariance operator $\Pi \in C^a([0,T], \mathcal{S}^+(\mathcal{Y}))$ solution of (1).

Idea of the existence proof.

$\left(\begin{smallmatrix} \mathbb{1} & \Pi_\diamond \\ 0 & \mathbb{1} \end{smallmatrix}\right)$ diagonalizes the system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \bar{y} \\ \bar{q} \end{pmatrix} = \begin{pmatrix} A & BQB^* \\ C^*RC & -A \end{pmatrix}\begin{pmatrix} \bar{y} \\ \bar{q} \end{pmatrix} + c_{\mathrm{st}}, \qquad \begin{pmatrix} \bar{y}(0) \\ \bar{q}(0) \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \Pi_\diamond \\ 0 & \mathbb{1} \end{pmatrix}\begin{pmatrix} y_\diamond \\ q_0 \end{pmatrix}$$

□

Theorem

Kalman-Bucy The optimal sequential estimator is the unique solution to the following dynamics

$$\begin{cases} \dot{\hat{y}} = A\hat{y} + \Pi C^*(z^\delta - C\hat{y}), & \textit{in } (0, T) \\ \hat{y}(0) = y_\diamond \end{cases} \tag{2}$$

Moreover, we have the fundamental identity

$$\forall t \in [0, T], \quad \bar{y}_{|T}(t) = \hat{y}(t) + \Pi(t)\bar{q}_{|T}(t). \tag{3}$$

Proof.

There exists a unique solution in $C^a([0, T], \mathcal{Y})$ to the dynamics (2). By unicity of the solutions, we verify (3) as each side follow the same dynamics and initial condition. $\square$

Initial condition reconstruction, a back and forth approach

- Injecting $\bar{y}_{|T}(t) = \hat{y}(t) + \Pi(t)\bar{q}_{|T}(t)$ in the adjoint dynamics

$$\begin{cases} \dot{\bar{q}}_{|T} + (A^* - C^* C\Pi(t))\bar{q}_{|T} = -C^* R(z^\delta - C\hat{y}(t)), & \text{in } (0, T) \\ \bar{q}_{|T}(T) = 0 \end{cases}$$

- A simple configuration: $A = -A^*$, $C^* = B$ hence $\forall t \in [0, T], \quad \Pi = \Pi_\diamond = \frac{1}{\alpha^2}\mathbb{1}$. Here, we can prove that we retrieve with $\alpha = 1$ and

$$\forall t \in [0, T], \quad \hat{y}^b(t) = \hat{y}(t) - 2q(t)$$

the back and forth nudging algorithm for conservative systems (Auroux and Blum, 2005; Ramdani et al., 2010), namely

$$\begin{cases} \dot{\hat{y}}_b = A^* \hat{y}_b - C^* R(z^\delta - C\hat{y}_b(t)), & \text{in } (0, T) \\ \hat{y}_b(T) = \hat{y}(T) \end{cases}$$

Initial condition reconstruction, an equivalent sequential approach

- Augmented form ${}^\flat y = (y, \zeta)^\intercal$ with $\dot\zeta = 0$.

$$ {}^\flat A = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, {}^\flat B = \begin{pmatrix} B \\ 0 \end{pmatrix}, {}^\flat C = \begin{pmatrix} C & 0 \end{pmatrix}, {}^\flat \Pi_\diamond = \begin{pmatrix} \Pi_\diamond & \Pi_\diamond \\ \Pi_\diamond & \Pi_\diamond \end{pmatrix} $$

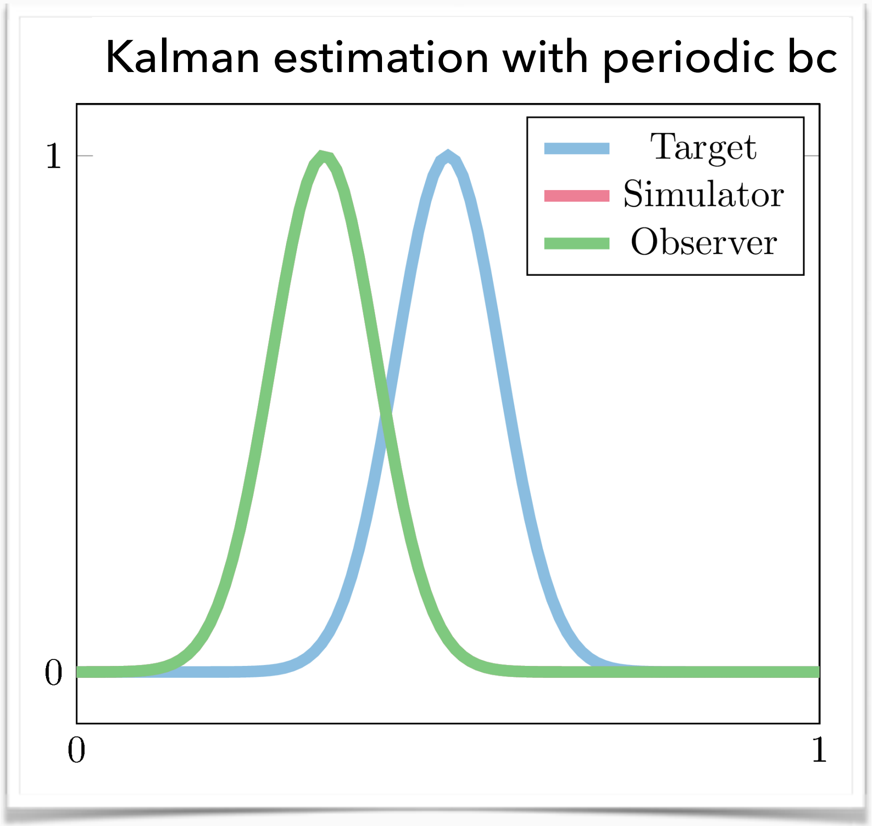
- Let ${}^\flat q = (q, \lambda)^\intercal$, We get $\dot\lambda = 0$, $\lambda(T) = 0 \Rightarrow \lambda \equiv 0$ hence for all $t \in [0, T]$

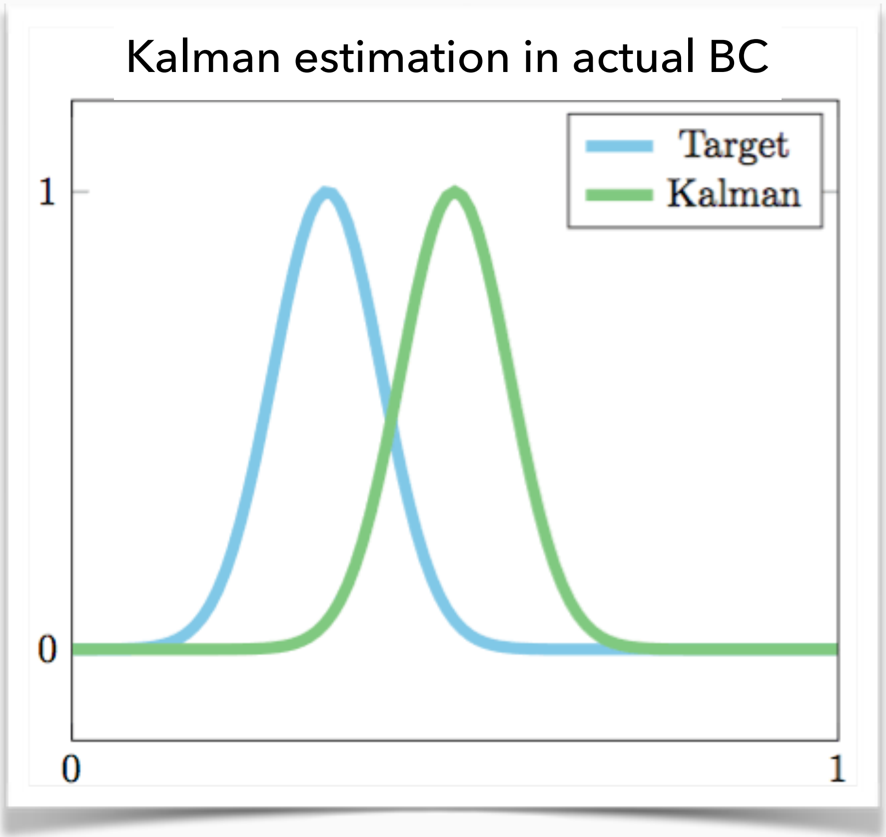
$$ \begin{pmatrix} \bar y \\ \bar\zeta \end{pmatrix} = \begin{pmatrix} \hat y \\ \hat\zeta \end{pmatrix} + \begin{pmatrix} \Pi_{yy} & \Pi_{y\zeta} \\ \Pi_{\zeta y} & \Pi_{\zeta\zeta} \end{pmatrix} \begin{pmatrix} \bar q \\ \bar\lambda \end{pmatrix} = \begin{pmatrix} \hat y + \Pi_{yy}\bar q \\ \hat\zeta + \Pi_{\zeta y}\bar q \end{pmatrix} $$

- Same equations for $\hat y$ and Π_{yy}, and we additionally retrieve in $(0, T)$

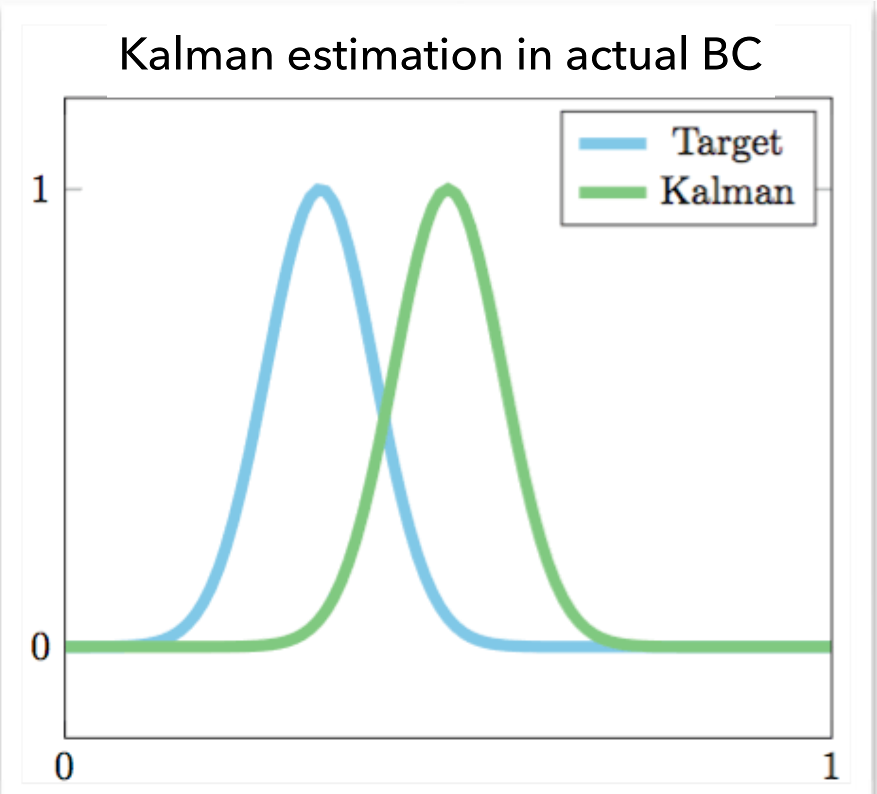
$$ \dot{\zeta} = \Pi_{\zeta y} C^* R(z^\delta - C\hat y), $$
$$ \dot\Pi_{\zeta y} = \Pi_{\zeta y} A^* - \Pi_{\zeta y} C^* R C \Pi_{yy} $$

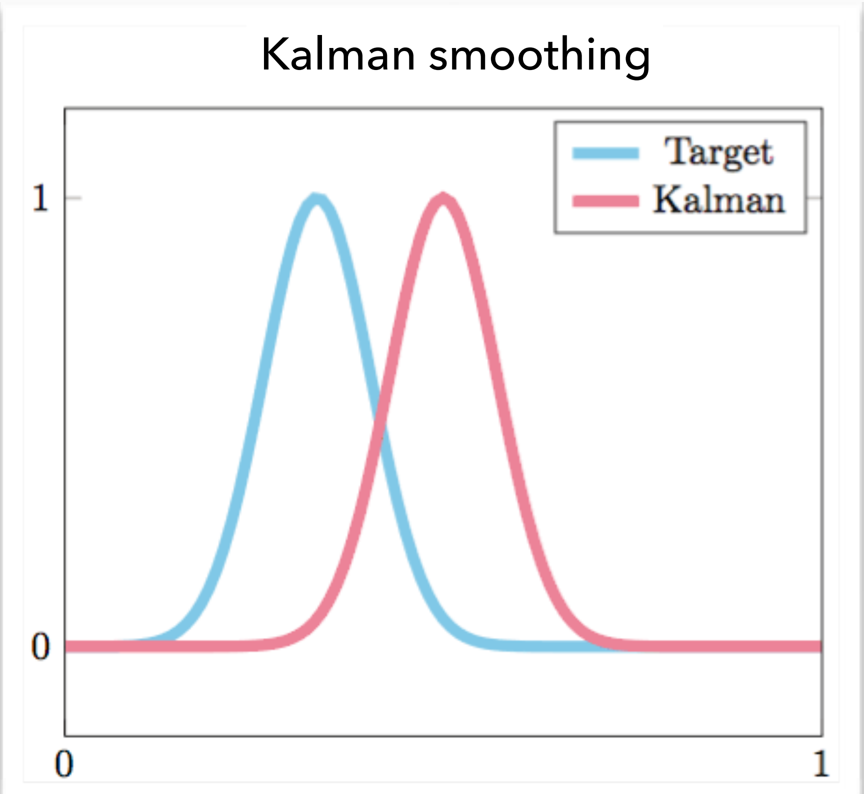
Example


Kalman estimation with periodic bc
Target
Simulator
Observer
1
0
0
1


Kalman estimation in actual BC
Target
Kalman
1
0
0
1

Example


Kalman estimation in actual BC
Target
Kalman
1
0
0
1


Kalman smoothing
Target
Kalman
1
0
0
1

Kalman in an infinite dimension context

- What sense to give to
$$\dot{y} = Ay + B\nu$$
namely How to generalize e^{tA} ...
- Notion of mild solution via Duhamel formulae:
$$\forall t \in [0, T], \quad y(t) = \Phi(t)y_0 + \int_0^t \Phi(t-s)B\nu(s)\,\mathrm{d}s.$$
- Ex: If $y_0 \in \mathcal{Y}$ and $\nu \in L^2(0, T; \mathcal{U})$, we have a unique mild solution y that satisfies
$$y \in W^{1,2}(0, T; \mathcal{D}(A)') \cap C^0([0, T]; \mathcal{Y});$$
- Notion of weak solution (see for instance (Bensoussan et al., 2007)):

  $y \in L^p(0, T; \mathcal{Y})$,
  for all $q \in D(A^*)$, $\langle q, y(\cdot)\rangle$ belongs to $W^{1,p}(0, T)$

for almost all $t \in (0, T)$ and $q \in D(A^*)$,

$$\frac{d}{dt}\langle q, y(t)\rangle = \langle A^* q, y(t)\rangle + \langle q, B\nu(t)\rangle.$$

- The variational case: We can consider $A(t)$ associated to a $a(u, v, t)$ elliptic bilinear form. The evolution operator becomes $\Phi(t, s)$. We have additional regularity properties.

Covariance operator from Riccati dynamics $\dot{\Pi} = A\Pi + \Pi A^* - \Pi C^* R C \Pi + B^* Q B$

- A mild solution to (1) is a function $\Pi \in C([0,T], \mathcal{S}(\mathcal{Y}))$ that satisfies for all $y \in \mathcal{Y}$, $t \in [0,T]$

$$\Pi(t)y = \Phi(t)\Pi_0\Phi^*(t) - \int_0^t \Phi(t-s)\Pi(s)C^*RC\Pi(s)\Phi^*(t-s)\,\mathrm{d}s + \int_0^t \Phi(s)BQB^*\Phi^*(s)\,\mathrm{d}s; \quad (4)$$

- A weak solution to (1) is a function $\Pi \in C([0,T], \mathcal{S}(\mathcal{Y}))$ such that for all $(y_1, y_2) \in \mathcal{D}(A)$, $(\Pi(\cdot)y_1, y_2)_{\mathcal{Y}}$ is differentiable and verifies

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Pi(t)y_1, y_2)_{\mathcal{Y}} = (\Pi(t)y_1, A^*y_2)_{\mathcal{Y}} + (\Pi(t)A^*y_1, y_2)_{\mathcal{Y}} - (C\Pi(t)y_1, RC\Pi(t)y_2)_{\mathcal{Z}} + (B^*y_1, QB^*y_2)_{\mathcal{U}}, \quad t \in [0,T] \quad (5)$$

- Let's $t_1 > t_0$ and introduce a similar criterion

$$\mathscr{J}_{|t_0,t_1,\lambda}(\zeta,\nu) = \frac{1}{2}(\zeta,\Lambda_\diamond\zeta)_{\mathcal{Y}} - (\lambda, y_{|\zeta,\nu}(t_1))_{\mathcal{Y}} + \frac{1}{2}\int_{t_0}^{t_1}(\nu(t), S\nu(t))_{\mathcal{U}}\,\mathrm{d}t + \frac{1}{2}\int_{t_0}^{t_1}(C(t)y_{|\zeta,\nu}(t), RC(t)y_{\zeta,\nu}(t))_{\mathcal{Z}}\,\mathrm{d}t,$$

- It admits a unique optimal solution leading to the both-end dynamics with $\Pi_\diamond = \Lambda_\diamond^{-1}$, $Q = S^{-1}$:

$$\begin{cases} \dot{\bar{y}}_{|t_0,t_1,\lambda}(t) = A(t)\bar{y}_{|t_0,t_1,\lambda}(t) + B(t)QB(t)^*\bar{q}_{|t_0,t_1,\lambda}(t), & t \in (t_0,t_1) \\ \dot{\bar{q}}_{|t_0,t_1,\lambda}(t) + A(t)^*\bar{q}_{|t_0,t_1,\lambda}(t) = C(t)^*RC(t)\bar{y}_{|t_0,t_1,\lambda}(t), & t \in (t_0,t_1) \\ \bar{y}_{|t_0,t_1,\lambda}(t_0) = \Pi_\diamond\bar{q}_{|t_0,t_1,\lambda}(t_0), \\ \bar{q}_{|t_0,t_1,\lambda}(t_1) = \lambda. \end{cases}$$

Inherent definition ii

Lemma (Covariance operator inherent definition)

- *The mapping* $\lambda \in \mathcal{Y} \mapsto (\bar{\zeta}_{|t_0,t_1,\lambda}, \bar{\nu}_{|t_0,t_1,\lambda}) \in \mathcal{Y} \times L^2(t_0, t_1; \mathcal{U})$ *is linear continuous.*
- *The mapping* $q \mapsto \bar{y}_{|t_0,t_1,\lambda}(t_1)$ *is a bounded operator on* $\mathcal{Y}$*. We denote by* $\Pi(t_1) \in \mathcal{L}(\mathcal{Y})$ *this operator:*

$$\bar{y}_{|t_0,t_1,\lambda}(t_1) = \Pi(t_1)\lambda, \quad \forall \lambda \in \mathcal{Y}.$$

- *For every* $t_1 > t_0$, $\sup_{t\in[t_0,t_1]} \|\Pi(t)\|_{\mathcal{L}(\mathcal{Y})} < +\infty$.

- We even have $\forall t \in [t_0, t_1], \quad \bar{y}_{|t_0,t_1,\lambda}(t) = \Pi(t)\bar{q}_{|t_0,t_1,\lambda}(t)$
- which implies $\forall \lambda \in \mathcal{Y}, \quad (\lambda, \Pi(t)\lambda)_{\mathcal{Y}} \geq 0$,
- and $\min_{(\zeta,\nu)\in\mathcal{Y}\times L^2(t_0,t;\mathcal{U})} \mathscr{J}_{t_0,t,\lambda}(\zeta, \nu) = -\frac{1}{2}(\lambda, \Pi(t)\lambda)_{\mathcal{Y}}$.

Trace class and Hilbert Schmidt operators i

- Let's $P \in \mathcal{J}_1(\mathcal{Y})$, the spaces of trace class (nuclear) operator.

$$\forall (e_n)_{n\geq 0} \text{ b.o.n of } \mathcal{Y}, \quad \mathrm{tr}(|P|) = \sum_{n\geq 0} (e_n, (P^*P)^{\frac{1}{2}} e_n)_{\mathcal{Y}} = \|P\|_1.$$

- Let's $P \in \mathcal{J}_2(\mathcal{Y})$, the spaces of Hilbert-Schmidt operator.

$$\|P\|_2 = \sqrt{\mathrm{tr}(P^*P)} = \Big(\sum_{n\geq 0} (e_n, P^*Pe_n)_{\mathcal{Y}}\Big)^{\frac{1}{2}}.$$

- When additionally $P \in \mathcal{S}^+(\mathcal{Y})$, then $\mathrm{tr}(P) = \sum_{n\geq 0} \langle Pe_n, e_n \rangle$.
- With $\mathcal{K}(\mathcal{Y})$ the space of compact operators, we have $\mathcal{J}_1(\mathcal{Y}) \subset \mathcal{J}_2(\mathcal{Y}) \subset \mathcal{K}(\mathcal{Y}) \subset \mathcal{L}(\mathcal{Y})$.

Theorem

Let $p \in \{1, 2\}$ and assume that

- $\mathit{\Pi}_0 \in \mathcal{S}^+(\mathcal{Y}) \cap \mathcal{J}_p(\mathcal{Y})$,
- $BB^* \in \mathcal{S}^+(\mathcal{Y}) \cap \mathcal{J}_p(\mathcal{Y})$
- $CC^* \in \mathcal{S}^+(\mathcal{Y})$.

Then, the Riccati dynamics (1) *admits one and only one mild solution $\mathit{\Pi}$ and $\mathit{\Pi} \in C([0, T], \mathcal{S}^+(\mathcal{Y})) \cap C([0, T], \mathcal{J}_p(\mathcal{Y}))$.*

Theorem (Kernel Theorem in L^2)

An operator Π from $L^2(\Omega)$ to $L^2(\Omega)$ is a Hilbert-Schmidt operator if and only if it is associated with a kernel $\pi \in L^2(\Omega \times \Omega)$ such that

$$\forall \varphi \in L^2(\Omega), \quad \forall x \in \Omega, \quad (\Pi \varphi)(x) = \int_\Omega \pi(x', x)\varphi(x')\ \mathrm{d}x'.$$

Theorem (Kernel Theorem in Sobolev Spaces)

Let $(m, p) \in \mathbb{N}^2$. An operator Π from $H^m(\Omega)'$ to $H^p(\Omega)$ is a Hilbert-Schmidt operator if and only if it is associated with a kernel $\pi \in H^{m,p}(\Omega \times \Omega)$ such that

$$\forall \psi \in H^m(\Omega)', \quad (\Pi \psi)(x) = \langle \psi, \pi(\cdot, x)\rangle_{H^m}.$$

- Lets consider an illustrative example: the heat equation

$$\begin{cases} \partial_t u(x,t) - \Delta u(x,t) = \chi(x)\nu(t), & (x,t) \in \Omega \times (0,T) \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T) \\ u(x,0) = u_0(x) & x \in \Omega \end{cases} \quad (6)$$

- The state space $\mathcal{Y} = L^2(\Omega)$, $(e_n, \mu_n)_n$ the eigen-elements of $-\Delta^{-1}$ gives a b.o.n
- The model dynamics: $A = \Delta, \quad D(A) = H^2(\Omega) \cap H_0^1(\Omega)$.
- The model noise operator and observation operator:

$$B \,:\, \mathbb{R} \ni \nu \mapsto \chi(x)\nu \in L^2(\Omega), \quad B^* \,:\, L^2(\Omega) \ni \varphi(x) \mapsto \int_\Omega \chi(x)\varphi(x) \in \mathbb{R}.$$

Kernel representation: the heat equation example ii

Theorem

Let $p \in \{1,2\}$ and $\Pi_0 \in \mathcal{S}^+(\mathcal{Y}) \cap \mathcal{J}_p(L^2(\Omega)) \cap \mathcal{L}(H^{-1}(\Omega), H_0^1(\Omega))$. The Riccati dynamics associated with the heat equation admits one and only one mild solution Π, associated with a kernel $\pi \in W^{1,2}([0,T]; L^2(\Omega \times \Omega))$ solution to

$$\begin{cases} \partial_t \pi(x',x,t) - (\Delta_x + \Delta_{x'})\pi(x',x,t) \\ \qquad\qquad\qquad + \beta\,|\Omega|\,\chi(x')\chi(x) \\ \qquad = -\delta \displaystyle\int_\omega \pi(t,x,x'')\pi(t,x'',x')\,\mathrm{d}x'', & (x',x,t) \in \Omega \times \Omega \times (0,T) \\ \pi(x',x,t) = 0, & (x',x,t) \in \partial\Omega \times \Omega \times (0,T) \\ \pi(x',x,t) = 0, & (x',x,t) \in \Omega \times \partial\Omega \times (0,T) \\ \pi(x',x,0) = \pi_0(x',x), & (x',x) \in \Omega \times \Omega \end{cases}$$

where π_0 is the kernel associated with the initial covariance Π_0.

- Mild solution: $y \in C([0,T];\mathcal{Y})$

$$\begin{cases} \dot{\hat{y}} = A\hat{y} + \Pi C^*(z^\delta - C\hat{y}), & \text{in } (0,T) \\ \hat{y}(0) = y_\diamond \end{cases}$$

- Weak solution: $y \in L^p(0,T;\mathcal{Y})$ $(1 \leq p \leq \infty)$, $\langle q, y(\cdot)\rangle$ belongs to $W^{1,p}(0,T)$ and for almost all $t \in (0,T)$ and $q \in D(A^*)$

$$\frac{d}{dt}\langle q, \hat{y}(t)\rangle = \langle A^* q, \hat{y}(t)\rangle + \langle C\Pi q, z^\delta - C\hat{y}\rangle.$$

Estimator dynamics: the heat equation example

- Weak form: For all $v \in L^2(0, T; H_0^1(\Omega))$

$$(\partial_t \hat{u}(x,t), v)_{L^2(\Omega)} - (\nabla \hat{u}(x,t), \nabla v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)} + \int_{\omega} \int_{\Omega} \delta\pi(x', x, t)\hat{v}(x', t)\big(z^{\delta}(x,t) - \hat{u}(x,t)\big)\ \mathrm{d}x'\ \mathrm{d}x.$$

- Strong form (by symetry of π):

$$\begin{cases} \partial_t \hat{u}(x,t) - \Delta \hat{u}(x,t) = f + \displaystyle\int_{\omega} \delta\pi(x', x, t)\big(z^{\delta}(x', t) - \hat{u}(x', t)\big)\ \mathrm{d}x', & x \text{ in } \Omega, \\ u(x,t) = 0, & x \text{ on } \partial\Omega. \end{cases}$$

Space discretization: Formalism in the context of the heat equation

A Ritz-Galerkin method $\mathcal{Y}^h \subset \mathcal{V}^h$

$$\forall (u^h, v^h) \in \mathcal{V}^h \times \mathcal{V}^h, \quad (u^h, v^h)_{\mathcal{Y}} = \mathrm{v}^\intercal \mathrm{M} \mathrm{u}, \quad (Au^h, v^h)_{\mathcal{Y}} = -\mathrm{v}^\intercal \mathrm{K} \mathrm{u},$$

Discrete-space Operator A^h s.t. $(-A^h u^h, v^h)_{\mathcal{Y}} = a(u^h, v^h) = (-Au, v)_{\mathcal{Y}}$.

Dynamics in $\mathcal{V}^h$

$$\begin{cases} \dot{y}^h = A^h y^h + B^{h,h''} \nu^{h''}, & \text{in } (0, T), \\ y^h(0) = P^h y_\diamond + P^h \zeta, \end{cases} \quad \Bigg| \quad \text{observed from } z^{\delta,h} = P^{h,h'} z^{\delta,h'}$$

Equivalently, the dynamics in $\mathbb{R}^{N^h}$

$$\begin{cases} \mathrm{M}\dot{\mathrm{y}} + \mathrm{K}\mathrm{y} = \mathrm{B}\nu, & \text{in } (0, T), \\ \mathrm{y}(0) = \mathrm{y}_\diamond + \zeta, \end{cases} \Bigg| \quad \text{observed from } \mathrm{z}^\delta \text{ the dof of } z^{\delta,h}$$

Finite dimensional criterion

- Criterion

$$\mathscr{J}_T^h(\zeta^h, \nu^{h''}) = \frac{1}{2}\|\zeta\|^2_{\mathcal{Y},\Lambda_\diamond} + \frac{1}{2}\int_0^T \left(\|z^{\delta,h}(s) - C^h y^h_{|\zeta^h,\nu^{h''}}(s)\|^2_{\mathcal{Z},R} + \|\nu^{h''}(s)\|^2_{\mathcal{U},R}\right) \mathrm{d}s$$

- Estimator

$$\begin{cases} \dot{\hat{y}} = A^h\hat{y} + \Pi^h C^{h*}(z^{\delta,h} - C^h\hat{y}^h), & \text{in } (0,T) \\ \hat{y}(0) = y_\diamond^h \end{cases}$$

- and Riccati

$$\begin{cases} \dot{\Pi}^h = A^h\Pi^h + \Pi^h A^{h*} - \Pi^h C^{h*} R C^h \Pi^h + B^{h*} Q B^h, & \text{in } (0,T) \\ \Pi(0) = P^{h*}\Pi_\diamond P^{h*}. \end{cases}$$

Resulting dynamical system in $\mathbb{R}^{N_y}$

- Criterion

$$\mathscr{J}_T^h(\zeta,\nu) = \mathscr{J}_T(\zeta^h,\nu) \;=\; \frac{1}{2}\zeta^\intercal \Lambda_\diamond \zeta + \frac{1}{2}\int_0^T \nu(t)^\intercal \mathrm{S}\nu(t)\,\mathrm{d}t$$
$$+\frac{1}{2}\int_0^T \left(\mathrm{z}^\delta(t) - \mathrm{C}\mathrm{y}(t)\right)^\intercal \mathrm{R}\left(\mathrm{z}^\delta(t) - \mathrm{C}\mathrm{y}(t)\right)\,\mathrm{d}t,$$

- Estimator

$$\begin{cases} \mathrm{M}\dot{\hat{\mathrm{y}}} = -\mathrm{K}\hat{\mathrm{y}} + \mathrm{M}\Pi\mathrm{M}^{-1}\mathrm{C}^\intercal\mathrm{R}\left(\mathrm{z}^\delta - \mathrm{C}\hat{\mathrm{y}}\right), \quad t>0, \\ \hat{\mathrm{y}}(0) = \mathrm{y}_\diamond, \end{cases}$$

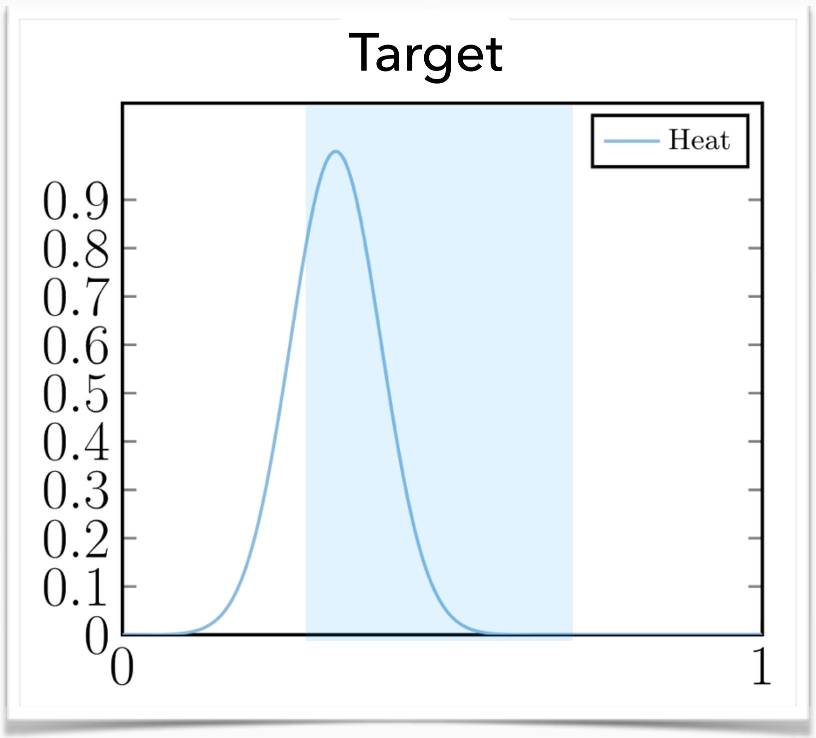
- and Riccati

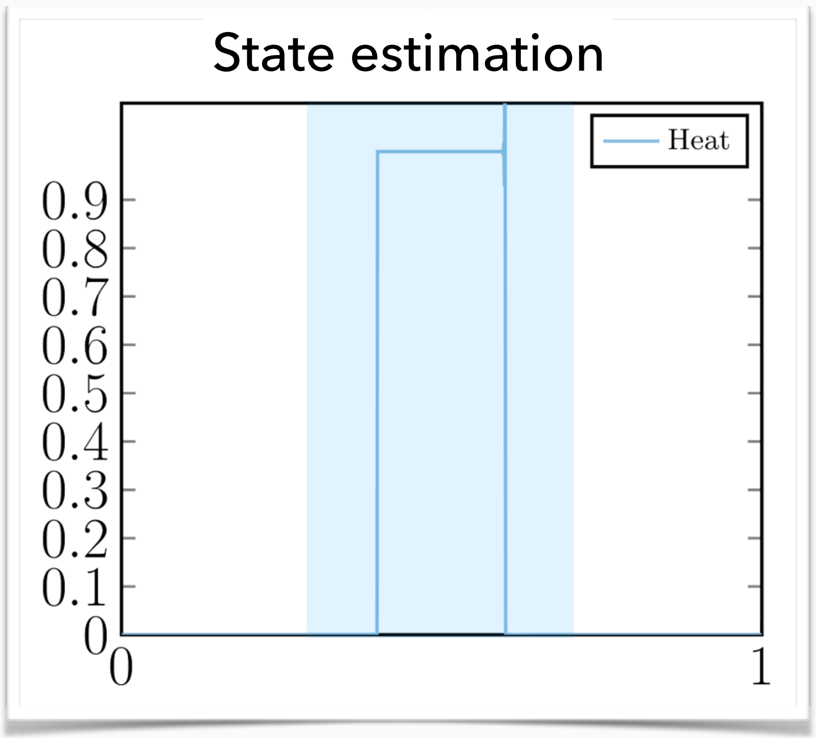
$$\begin{cases} \mathrm{M}\dot{\Pi} = \; -\mathrm{K}\Pi - \mathrm{M}\Pi\mathrm{M}^{-1}\mathrm{K}^\intercal - \mathrm{M}\Pi\mathrm{M}^{-1}\mathrm{C}^\intercal\mathrm{R}^{-1}\mathrm{C}\Pi + \mathrm{M}\mathrm{B}\mathrm{Q}\mathrm{B}^\intercal\mathrm{M}, \quad t>0, \\ \Pi(0) = \Pi_\diamond. \end{cases}$$

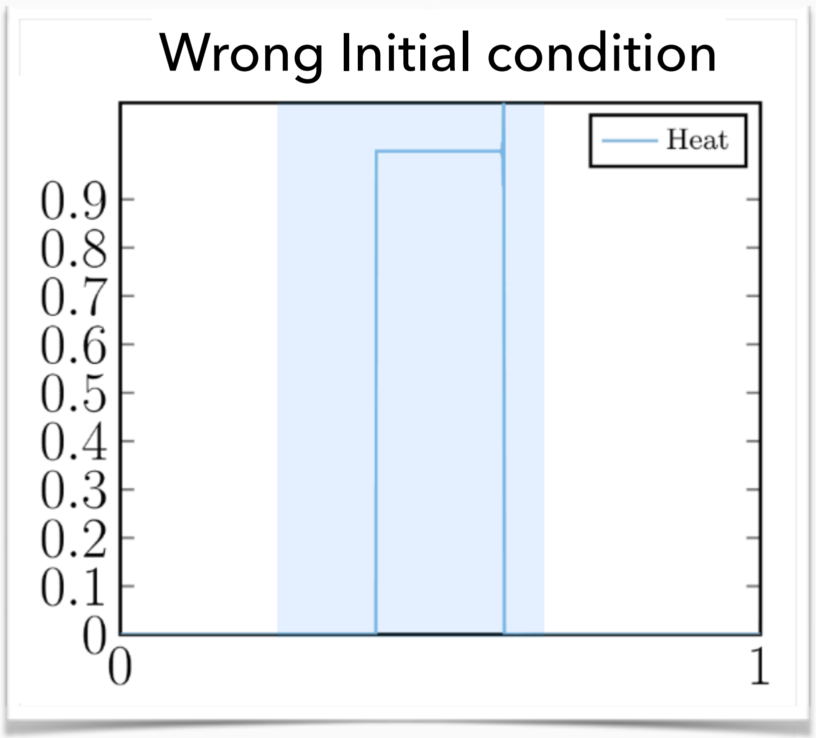
- The resulting matrix is dense $\Rightarrow$ use H-matrix to represent the matrix as it is associated with a kernel.

Example

1D problem $N_h = 10^3$

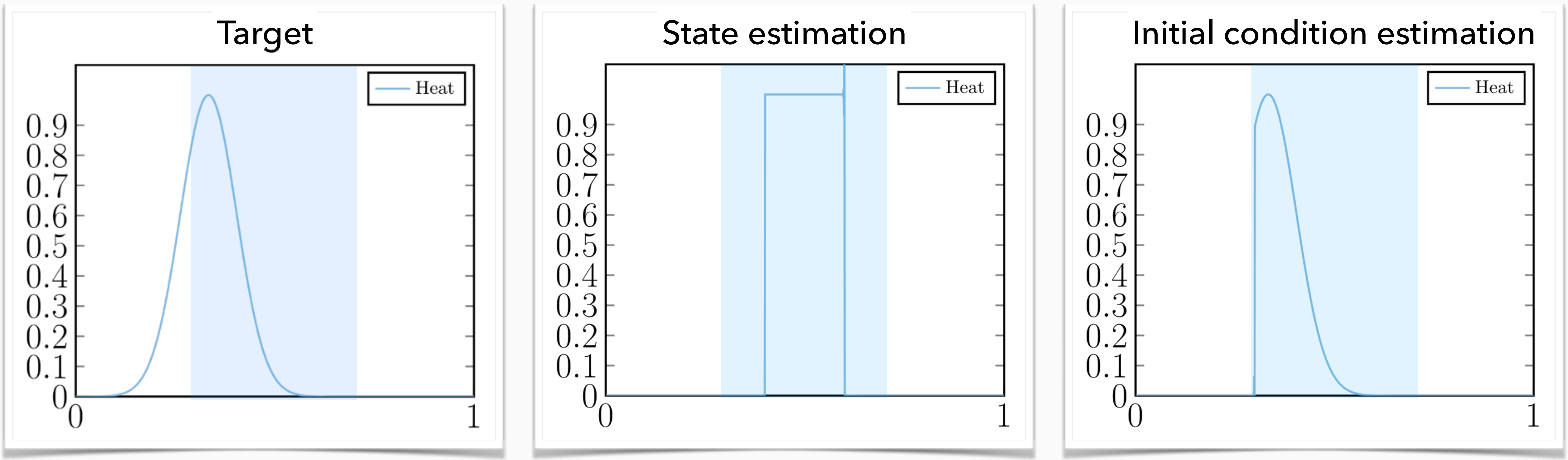






Example

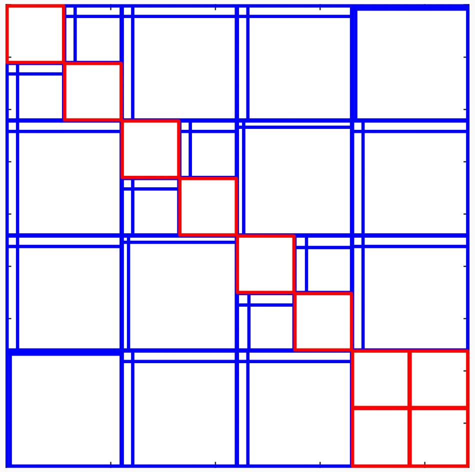
1D problem $N_h = 10^3$

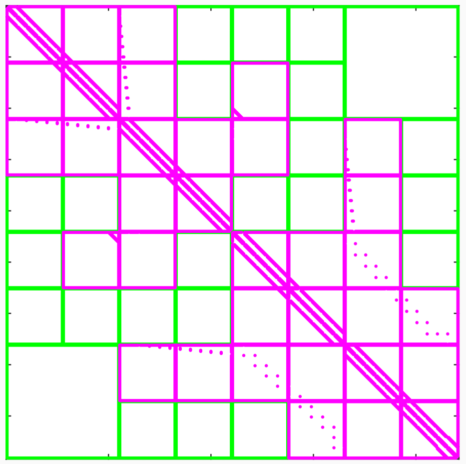


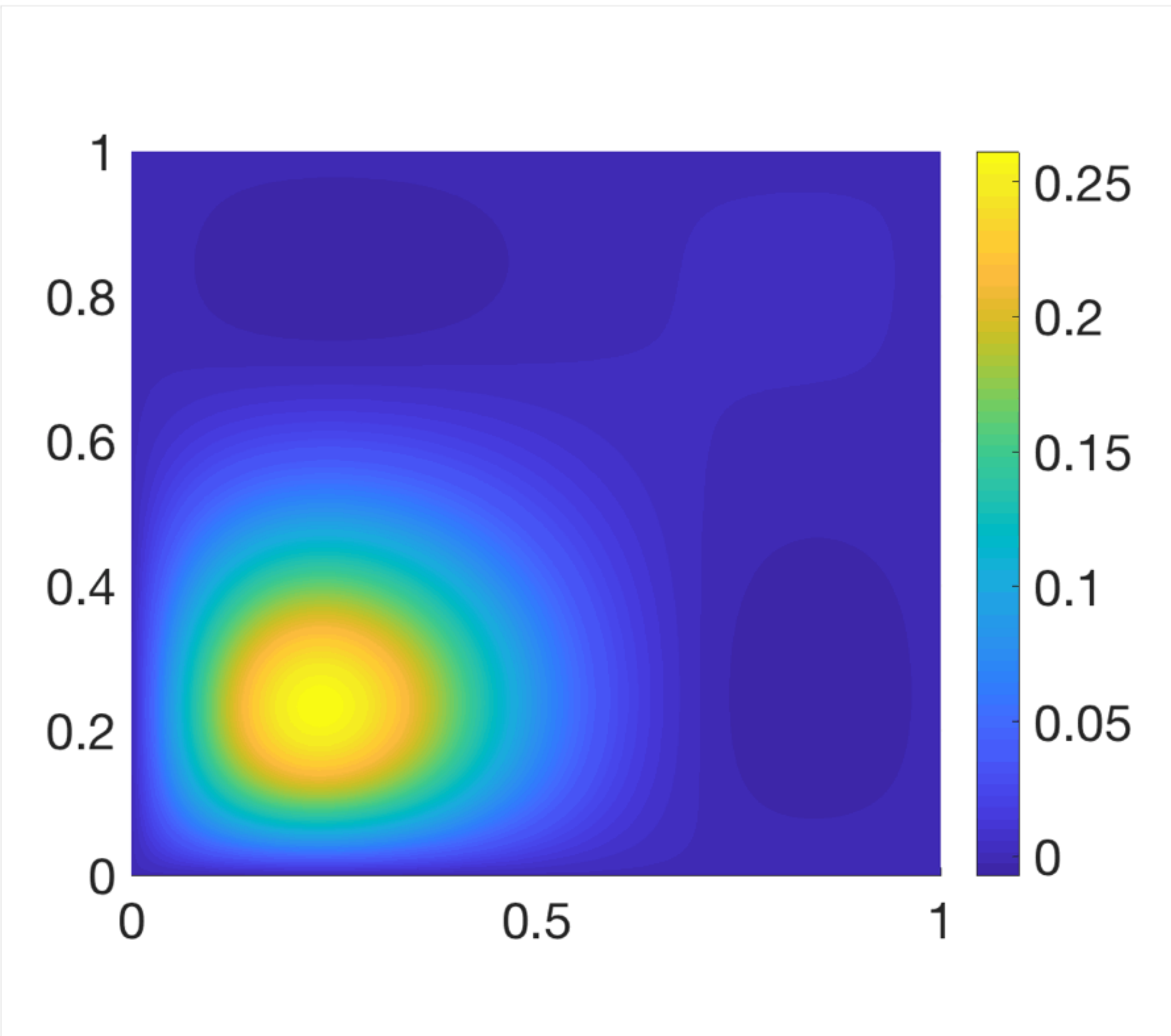
Example

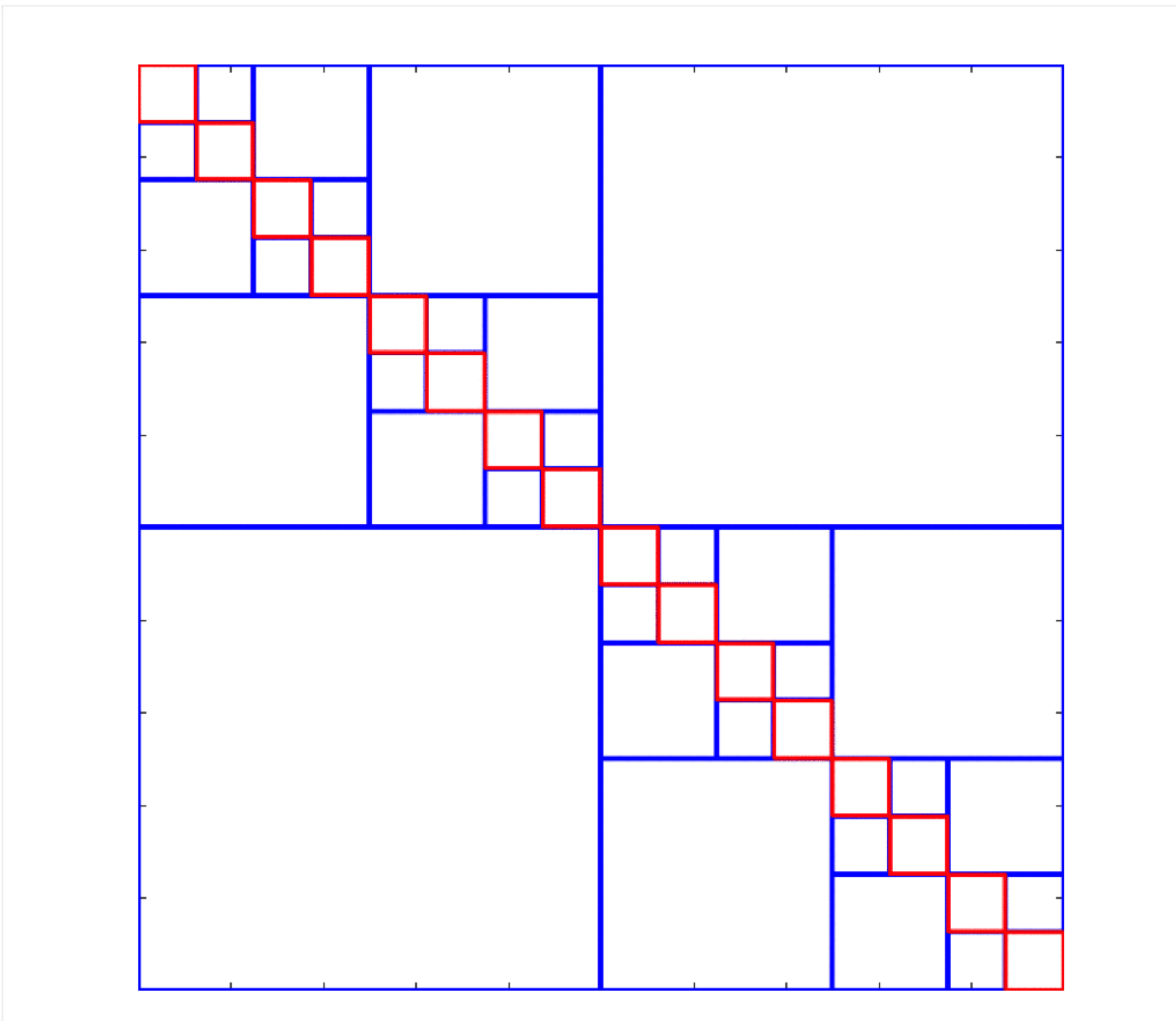
1D problem $N_h = 10^3$

H-matrix representation:





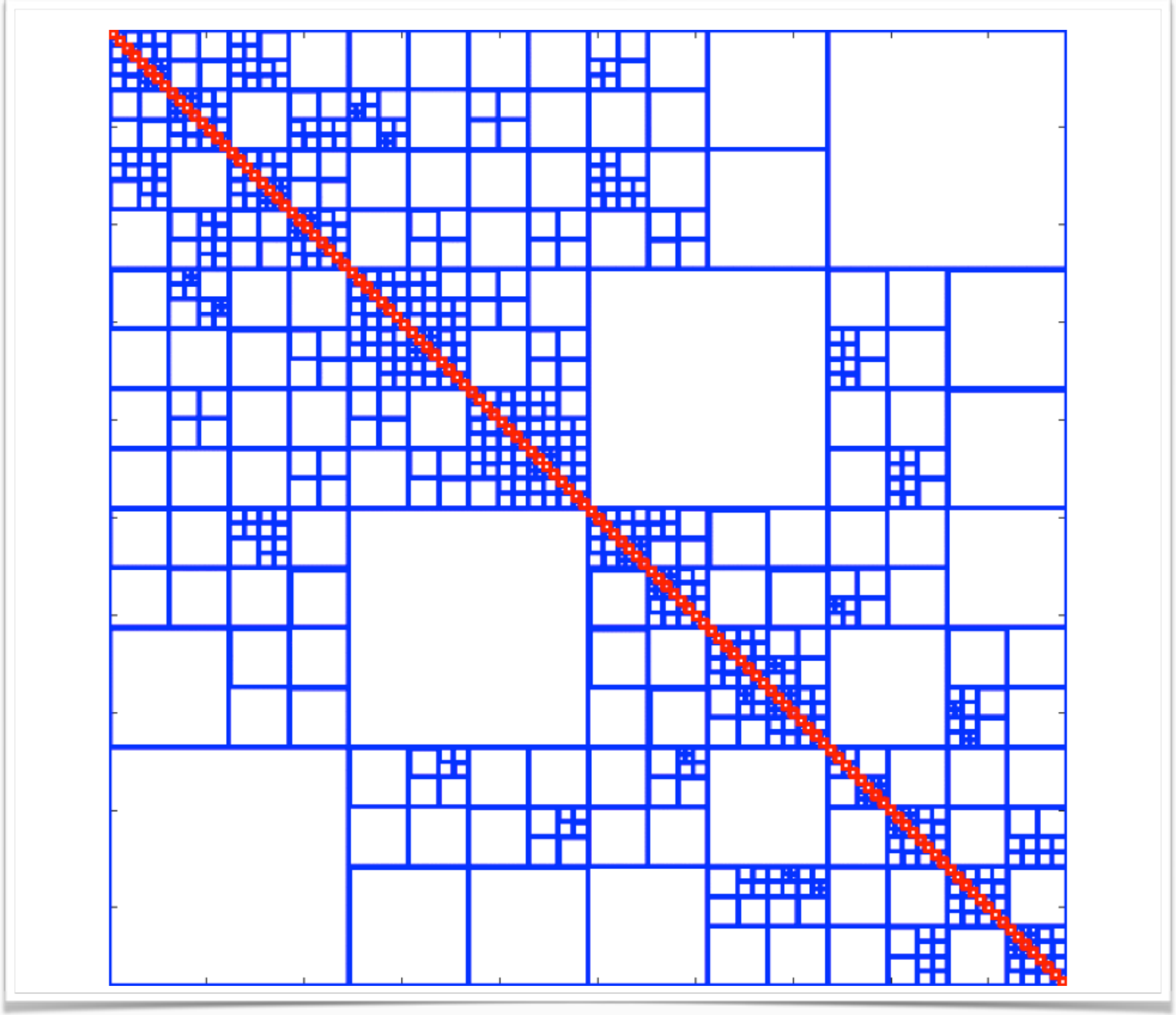




(Alouges et al 2019 ?)

Example

2D problem $N_h = 10^6$!!!!!!!!!



(Alouges et al 2019 ?)

Convergence analysis (case h',h" = 0)

- Based on the criterion approximation (Bensoussan, 1971)

$$\mathscr{J}_T^h(\zeta,\nu) \xrightarrow[h\to 0]{} \mathscr{J}_T(\zeta,\nu); \quad (\bar{\zeta}^h,\bar{\nu}^h) \xrightarrow[h\to 0]{\mathcal{Y}\times L^2(0,T;\mathcal{U})} (\bar{\zeta},\bar{\nu})$$

$$(\bar{y}^h_{|T},\bar{q}^h_{|T}) \xrightarrow[h\to 0]{L^2(0,T;\mathcal{V})} (\bar{y}_{|T},\bar{q}_{|T}); \quad \hat{y}^h \xrightarrow[h\to 0]{L^2(0,T;\mathcal{Y})} \hat{y}$$

$$(\Pi^h y^h, y_2) \xrightarrow[h\to 0]{L^\infty(0,T)} (\Pi y, y_2), \quad \forall (y,y^h,y_2)\in\mathcal{Y}^3 \mid y^h \xrightarrow[h\to 0]{\mathcal{Y}} y$$

- Or directly on the Riccati and estimator dynamics (Germani et al., 1988)

$$\lim_{h\to 0}\sup_{t\in[0,T]} \|\Pi^h(t) - P_h\Pi(t)P_h^*\|_{HS} = 0$$

$$\lim_{h\to 0}\sup_{0\leq s\leq t\leq T} \|\Phi^h(t,s)P_h y - P_h\Phi(t,s)y\| = 0, \quad \forall y\in\mathcal{Y}$$

The discrete-time Kalman filter as a time discretization of the continuous-time Kalman estimator

Discrete-time dynamical systems

- Recursive relation

$$\begin{cases} \mathrm{y}_{n+1} = \Phi_{n+1|n}\mathrm{y}_n + \mathrm{B}_{n+1}\nu_{n+1}, \quad n \geq 0, \\ \mathrm{y}_0 = \mathrm{y}_\diamond + \zeta, \end{cases}$$

- Observations $\mathrm{z}_n^\delta = \mathrm{C}_n y_n + \chi_n$ with possibly a different sampling, for instance

$$\mathrm{C}_n = \delta_{m,n} C \text{ or } \chi_n = \chi(t_n) + \epsilon_n.$$

- Criterion

$$\begin{aligned} \mathscr{J}_n(\zeta, (\nu_k)_{0\leq k\leq n}) = {} & \frac{1}{2}\zeta^\intercal \Lambda_\diamond \zeta + \frac{1}{2}\sum_{k=1}^{n} \nu_k^\intercal \mathrm{S}_k \nu_k \\ & + \frac{1}{2}\sum_{k=0}^{n} \left(\mathrm{z}_k^\delta - \mathrm{C}\mathrm{y}_{k|\zeta,(\nu_k)_{0\leq k\leq n}}\right)^\intercal \mathrm{R}_k \left(\mathrm{z}_k^\delta - \mathrm{C}\mathrm{y}_{k|\zeta,(\nu_k)_{0\leq k\leq n}}\right). \end{aligned}$$

Optimality system

- Criterion minimizers

$$\bar{\zeta}^{n+1} = \Pi_0(\bar{q}^{n+1})_{k=0}, \qquad (\bar{\nu}^{n+1})_k = \mathrm{Q}_k \mathrm{B}_k^\intercal \bar{\mathrm{q}}_{k|n},\ \forall k = 1, \cdots, n+1,$$

- Both ends system

$$\begin{cases} \bar{\mathrm{y}}_{k+1|n} = \Phi_{k+1|k}\bar{\mathrm{y}}_{k+1|n} + \mathrm{B}_{k+1}\mathrm{Q}_{k+1}\mathrm{B}_{k+1}^\intercal \bar{\mathrm{q}}_{k+1|n}, & 0 \leq k \leq n, \\ \bar{y}_{0|n} = \mathrm{y}_\diamond + \Pi_0 \bar{q}_{0|n}, & \\ \bar{\mathrm{q}}_{k|n} - \Phi_{k+1|k}^\intercal \bar{\mathrm{q}}_{k+1|n} = \mathrm{C}^\intercal \mathrm{R}_k \left(\mathrm{z}_k^\delta - \mathrm{C}\bar{\mathrm{y}}_{k|n}\right), & 0 \leq k \leq n, \\ \bar{\mathrm{q}}_{n+1|n} = 0. & \end{cases} \tag{7}$$

- Proof: Same as for the continuous-time case with a Lagrangian.

Discrete-time Kalman estimator (Kalman, 1960)

Definition (Optimal estimator)

The time-discrete optimal sequential estimator is defined by

$\forall n \in \mathbb{N}, \quad \hat{\mathrm{y}}_n = \bar{\mathrm{y}}_{n|n}.$

Theorem (Discrete-time Kalman estimator dynamics)

We have the following identity: $\bar{\mathrm{y}}_{k|n} = \hat{y}_k^- + \Pi_k^- \bar{\mathrm{q}}_{k|n}, \quad \forall 0 \leq k \leq n+1,$
where $(\hat{y}_k^-, \Pi_k^-)_{k\geq 0}$ *is defined by*

$$\begin{cases} \hat{\mathrm{y}}_0^- = \mathrm{y}_\diamond, \\ \hat{\mathrm{y}}_n^+ = \hat{\mathrm{y}}_n^- + \Pi_n^- \mathrm{C}^\intercal \left(\mathrm{C}\Pi_n^- \mathrm{C}^\intercal + \mathrm{W}_n\right)^{-1} \left(z_n - \mathrm{C}\hat{y}_n^-\right), \quad n \geq 0, \\ \hat{\mathrm{y}}_{n+1}^- = \Phi_{n+1|n} \hat{\mathrm{y}}_n^+, \quad n \geq 0. \end{cases}$$

$$\begin{cases} \Pi_0^- = \Pi_0, \\ \Pi_n^+ = \Pi_n^- - \Pi_n^- \mathrm{C}^\intercal \left(\mathrm{C}\Pi_n^- \mathrm{C}^\intercal + \mathrm{W}_n\right)^{-1} \mathrm{C}\Pi_n^-, \quad n \geq 0, \\ \Pi_{n+1}^- = \Phi_{n+1|n} \Pi_n^+ \Phi_{n+1|n}^\intercal + \mathrm{B}_{n+1} \mathrm{Q}_{n+1} \mathrm{B}_{n+1}^\intercal, \quad n \geq 0. \end{cases}$$

Convergence aspects – The heat equation case

- Choose adequately the weights, $(m_k)_{k=0}^{N+1}$ and $(\tilde{m}_k)_{k=1}^{N+1}$ quadrature rules on [0,T]

$$\Lambda_\diamond = \frac{1}{\alpha}\mathrm{M}, \quad \mathrm{S}_k = \frac{\tilde{m}_k}{\beta \Delta t}\mathrm{M}, \quad \mathrm{R}_k = \frac{m_k}{\delta \Delta t}\mathrm{M}_\omega.$$

- Note that the time-discrete adjoint variable $\mathrm{q}_{|n}$ is not a discretization of $q_{|T}$

$$\mathrm{q}_k - \mathrm{M}(\mathrm{M} + \delta t\mathrm{K})^{-1}\mathrm{q}_{k+1} = \mathrm{C}^\intercal \mathrm{R}(\mathrm{z}^\delta - \mathrm{C}\mathrm{y}_k).$$

  However, $\tilde{\mathrm{q}}_{|n} = \mathrm{M}^{-1}\mathrm{q}_{|n}$ is a stable and consistent approximation of $q_{|T}$ following

$$(\mathrm{M} + \delta t\mathrm{K})\tilde{\mathrm{q}}_k - \mathrm{M}\tilde{\mathrm{q}}_{k+1} = (\mathbb{1} + \delta t\mathrm{K}\mathrm{M}^{-1})\mathrm{C}^\intercal \mathrm{R}(\mathrm{z}^\delta - \mathrm{C}\mathrm{y}_k).$$

- Claim $\lim_{h\to 0, \Delta t \to 0} \| \Pi_n^h - P_h \Pi(t_n) P_h^* \|_{HS} = 0$
- Again we should consider the discretization δt and the data sampling ΔT.

Non-linear optimal estimator

Context

- Model: A finite dimensional dynamical systems, $A \in C^1(\mathbb{R}^{\mathrm{N}_y})$, B, $\mathrm{d}A$ bounded (possibly w.r.t time uniformly w.r.t time $A(y,t)$):

$$\begin{cases} \dot{y} = A(y,t) + B\nu, & \text{in } (0,T) \\ y(0) = y_\diamond + \zeta, \end{cases}$$

with $\|\zeta\|_{\mathcal{Y}} \leq \alpha$, $\|\nu\|_{\mathcal{U}_T} \leq \beta$.

- Observations $J(z^\delta, y) \in C^1(\mathbb{R}^{\mathrm{N}_{\mathrm{obs}}}, \mathbb{R}^{\mathrm{N}_y})$ replaces $z^\delta - C(y)$, with $\mathrm{d}_y J$ bounded uniformly (possibly w.r.t time $J(z^\delta, y, t)$):

$$J(z^\delta, y, t) = \eta,$$

with $\|\eta\|_{\mathcal{Z}_T} \leq \delta$.

The cost-to-come

- Cost-to-come:

$$\forall (y,t) \in \mathcal{Y} \times \mathbb{R}^+, \quad \mathscr{V}(y,t) = \min_{\substack{\omega \in L^2([0,T],\mathcal{U}),\\ \zeta \in \mathcal{Y} \mid y(t)=y}} \mathscr{J}(\zeta,\omega,t)$$

- Cost-to-dome dynamics : Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{cases} \partial_t \mathscr{V}(y,t) - \bar{\mathscr{H}}(y, \nabla \mathscr{V}(y,t), t) = 0, & (y,t) \in \mathcal{Y} \times \mathbb{R}^+ \\ \mathscr{V}(y,0) = \frac{1}{2}\|y - y_\diamond\|^2_{\mathcal{Y},\Lambda_\diamond} \end{cases}$$

with $\bar{\mathscr{H}}(y,q,t) = \frac{1}{2}\|J(y,t)\|_R^2 - \frac{1}{2} q^\intercal BQB^\intercal q - q^\intercal A(y,t)$

- Estimator dynamics when $\mathscr{V}$ is regular enough(Mortensen, 1968)

$$\begin{cases} \dot{\hat{y}}(t) = A(\hat{y},t) - (\nabla^2 \mathscr{V}(\hat{y}(t),t))^{-1} \, \mathrm{d}_y J(z^\delta, \hat{y}(t), t)^\intercal R J(z^\delta, \hat{y}(t), t), & t \in \mathbb{R}^+ \\ \hat{y}(0) = y_\diamond \end{cases}$$

- Linear case:

$$\mathscr{V}(y,t) = \frac{1}{2}(y - \hat{y}(t))^\intercal \Pi^{-1}(t)(y - \hat{y}(t)) + \frac{1}{2}\int_0^t \|z^\delta(s) - C(s)\hat{y}(s)\|_R^2 \, ds$$

The discrete-time non-linear case (Moireau, 2017): definitions

- For now: proofs are restricted to linear models and non-linear observation operator
- Two functionals

$$\min_{\substack{\zeta\in\mathcal{Y},\\(\nu_k)_{k<n}\in\mathcal{U}^n}} \left\{ \mathscr{J}_n^+(\zeta,(\nu_k)_{k<n}) = \frac{1}{2}\|\zeta\|^2_{\Lambda_\diamond} + \frac{1}{2}\sum_{k=0}^{n}\|J_k(y_k)\|^2_{R_k} + \frac{1}{2}\sum_{k=0}^{n-1}\|\nu_k\|^2_{S_k} \right\},$$

$$\min_{\substack{\zeta\in\mathcal{Y},\\(\nu_k)_{k\le n}\in\mathcal{U}^{n+1}}} \left\{ \mathscr{J}_{n+1}^-(\zeta,(\nu_k)_{k\le n}) = \frac{1}{2}\|\zeta\|^2_{\Lambda_\diamond} + \frac{1}{2}\sum_{k=0}^{n}\|J_k(y_k)\|^2_{R_k} + \frac{1}{2}\sum_{k=0}^{n}\|\nu_k\|^2_{S_k} \right\}$$

- Two costs-to-come

$$\begin{cases} \mathscr{V}_n^+(y) = \min\limits_{\substack{(\nu_k)_{k<n}\in\mathcal{U}^n\\ \zeta\in\mathcal{Y}\,|\,y_n=y}} \mathscr{J}_n^+(\zeta,(\nu_k)_{k<n}), \\ \mathscr{V}_{n+1}^-(y) = \min\limits_{\substack{(\nu_k)_{k\le n}\in\mathcal{U}^{n+1}\\ \zeta\in\mathcal{Y}\,|\,y_{n+1}=y}} \mathscr{J}_{n+1}^-(\zeta,(\nu_k)_{k\le n}), \quad \mathscr{V}_0^-(y) = \frac{1}{2}\|y-y_\diamond\|^2_{\Lambda_\diamond}. \end{cases}$$

The discrete-time non-linear case (Moireau, 2017): Recursive estimator

- Bellman eqns: $$\begin{cases} \mathscr{V}_0^-(y) = \frac{1}{2}\|y - y_\diamond\|^2_{\Lambda_\diamond}, \\ \mathscr{V}_n^+(y) = \mathscr{V}_n^-(y) + \frac{1}{2}\|J_n(y)\|^2_{R_n}, \\ \mathscr{V}_{n+1}^-(y) = \mathscr{V}_n^+(\lambda) + \frac{1}{2}\nabla\mathscr{V}_{n+1}^-(y)^\intercal B_n Q_n B_n^\intercal \nabla\mathscr{V}_{n+1}^-(y) \\ \qquad\qquad\qquad \text{with } y = A_{n+1|n}(\lambda) + B_n Q_n B_n^\intercal \nabla\mathscr{V}_{n+1}^-(y). \end{cases}$$
- Estimator: $$\begin{cases} \text{Initialization:} \quad \hat{y}_0^- = y_\diamond, \\ \text{Correction:} \quad \nabla\mathscr{V}_n^+(\hat{y}_n^+) = 0, \\ \text{Prediction:} \quad \hat{y}_{n+1}^- = A_{n+1|n}(\hat{y}_n^+) \text{ so that } \nabla\mathscr{V}_{n+1}^-(\hat{y}_{n+1}^-) = 0 \end{cases}$$
- With a Newton-Raphson algorithm

$$\begin{cases} \hat{y}_{0|n}^+ = \hat{y}_n^-, \quad n \in \mathbb{N} \\ \hat{y}_{k+1|n}^+ = \hat{y}_{k|n}^+ - (\nabla^2\mathscr{V}_n^+(\hat{y}_{k|n}^+))^{-1}\nabla\mathscr{V}_n^+(\hat{y}_{k|n}^+), \quad k \in \mathbb{N}. \end{cases}$$

Approximated Mortensen Filter: the Extended Kalman Filter (EFK)

- Continuous-time

$$\dot{\Pi} = \mathrm{d}A\Pi + \Pi\,\mathrm{d}A^* - \Pi\,\mathrm{d}J^* R\,\mathrm{d}J\Pi + \mathrm{d}B^* QB$$
$$\dot{\hat{y}} = A\hat{y} - \Pi\,\mathrm{d}J^* J$$

- Discrete-time

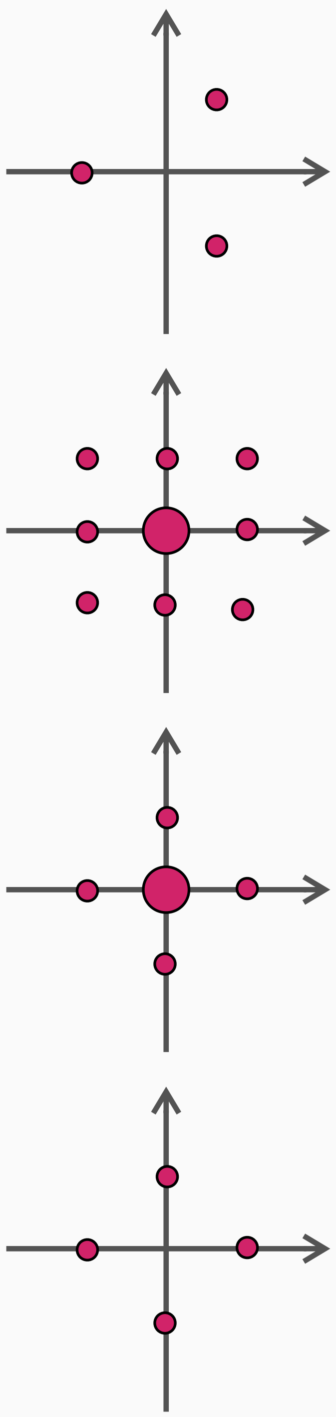
$$\hat{\mathrm{y}}_n^+ = \hat{\mathrm{y}}_n^- - \Pi_n^-\,\mathrm{dJ}_n^\intercal \left(\mathrm{dJ}_n\Pi_n^-\,\mathrm{dJ}_n^\intercal + \mathrm{W}_n\right)^{-1} \mathrm{J}_n$$
$$\Pi_n^+ = \Pi_n^- - \Pi_n^-\,\mathrm{dJ}_n^\intercal \left(\mathrm{dJ}_n\Pi_n^-\,\mathrm{dJ}_n^\intercal + \mathrm{W}_n\right)^{-1} \mathrm{dJ}_n\Pi_n^-$$
$$\hat{\mathrm{y}}_{n+1}^- = \Phi_{n+1|n}\hat{\mathrm{y}}_n^+$$
$$\Pi_{n+1}^- = \Phi_{n+1|n}\Pi_n^+\Phi_{n+1|n}^\intercal + \mathrm{B}_{n+1}\mathrm{Q}_{n+1}\mathrm{B}_{n+1}^\intercal$$

Approximated Mortensen Filter: the Uscented Transform

- Stencil $\mathbb{E}_\omega(e^{(\cdot)}) \overset{\text{def}}{=} \sum_{i=1}^{N_p} \omega^i e^{(i)} = 0, \quad \mathbb{C}\mathrm{ov}_\omega(e^{(\cdot)}) \overset{\text{def}}{=} \sum_{i=1}^{N_p} \omega^i e^{(i)} e^{(i)\intercal} = \frac{1}{\rho}\mathbb{1}$
- Consider "particles" propagated by a function φ and denote

$$\mathbb{y}_\varphi = (y_\varphi^{(i)})_{1\leq i\leq N_p} = (\varphi(y^{(i)}))_{1\leq i\leq N_p}$$

.



  - State approximation

$$\mathbb{E}_\omega(\mathbb{y}_\varphi) = \sum_{i=1}^{N_p} \omega_i y_\varphi^{(i)}$$

$$= \varphi(\mathbb{E}_\omega(\mathbb{y})) + \sum_{i=1}^{N_p} \omega_i \, \mathrm{d}\varphi(\mathbb{E}_\omega(\mathbb{y})) \cdot \tilde{y}^{(i)} + \dots$$

  - Covariance approximation

$$\mathbb{C}\mathrm{ov}_\omega(\mathbb{y}_\varphi) = \sum_{i=1}^{N_p} \omega_i (y_\varphi^{(i)} - \mathbb{E}_\omega(\mathbb{y}_\varphi))(y_\varphi^{(i)} - \mathbb{E}_\omega(\mathbb{y}_\varphi))^\intercal$$

$$= \mathrm{d}\varphi(\mathbb{E}_\omega(\mathbb{y})) \cdot \Big(\sum_{i=1}^{N_p} \omega_i \mathbb{C}\mathrm{ov}_\omega(\mathbb{y})\Big) \cdot \mathrm{d}\varphi(\mathbb{E}_\omega(\mathbb{y})) + \dots$$

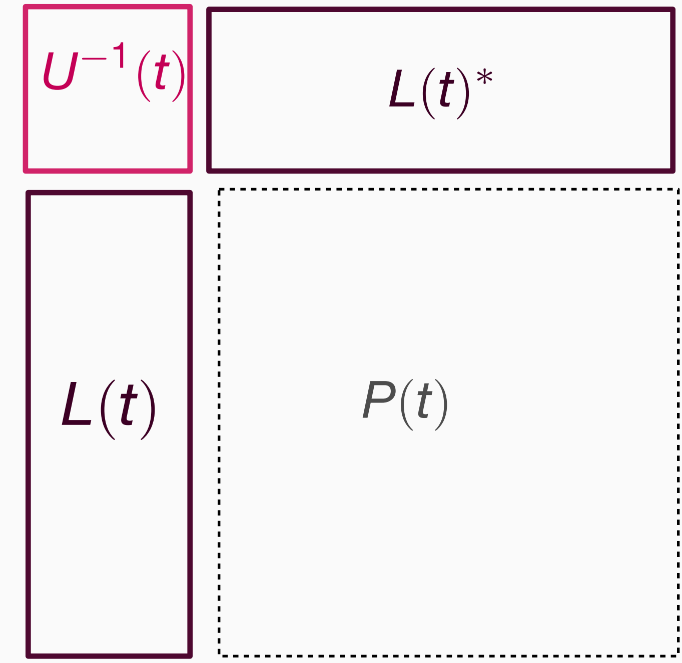
Approximated Mortensen Filter: the Uscented Kalman Filter (UFK)

- Stencil $\mathbb{E}_\omega(e^{(\cdot)}) \stackrel{\text{def}}{=} \sum_{i=1}^{\mathrm{N}_p} \omega^i e^{(i)} = 0, \quad \mathbb{C}\mathrm{ov}_\omega(e^{(\cdot)}) \stackrel{\text{def}}{=} \sum_{i=1}^{\mathrm{N}_p} \omega^i e^{(i)} e^{(i)\intercal} = \frac{1}{\rho}\mathbb{1}$
- Algorithm:

$$
\begin{aligned}
\text{Initialization: } & \hat{y}_0^- = y_\diamond \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \Pi_0^- = \Pi_\diamond \\
\text{Sampling: } & \zeta_n^{(i)} = J(z_n, \hat{y}_n^- + \sqrt{\rho \Pi_n^-} e^{(i)}), \ 1 \le i \le \mathrm{N}_p \\
\text{Correction: } & G_n = \sqrt{\Pi_0^n} \mathbb{C}\mathrm{ov}_\omega \big(e^{(\cdot)}, \zeta_n^{(\cdot)}\big) [\mathbb{C}\mathrm{ov}_\omega \big(\zeta_n^{(\cdot)}\big) + W_n]^{-1}, \\
& \hat{y}_n^+ = \hat{y}_n^- + G_n \big(z_n - C_n \hat{y}_n^-\big), \\
& \Pi_n^+ = \Pi_n^- - G_n [\mathbb{C}\mathrm{ov}_\omega \big(\tilde{z}_n^{(\cdot)}\big) + W_n] G_n^\intercal \\
\text{Sampling: } & \hat{y}_n^{(i)+} = \hat{y}_n^+ + \sqrt{\rho \Pi_n^+} e^{(i)}, \ 1 \le i \le \mathrm{N}_p \\
\text{Prediction: } & \hat{y}_{n+1}^{(i)-} = \Phi_{n+1|n} \big(\hat{y}_n^{(i)+}\big), \ 1 \le i \le \mathrm{N}_p \\
& \hat{y}_{n+1}^- = \mathbb{E}_\omega \big(\hat{y}_{n+1}^{(\cdot)-}\big), \\
& \Pi_{n+1}^- = \mathbb{C}\mathrm{ov}_\omega \big(\hat{y}_{n+1}^{(\cdot)-} - \hat{y}_{n+1}^-\big) + B_n Q_n B_n^\intercal,
\end{aligned}
$$

Reduced Order Optimal Filtering approaches

Reduced Order Decomposition



- Introduce the ansatz $P = LU^{-1}L^*$ where $L : \mathcal{Y}_r \to \mathcal{Y}$
- And find the dynamics of L and U
- Several interesting cases:
  - No model error $B = 0$

$$\dot{L} = AL, \quad \dot{U}^{-1} = L^* C^* RCL$$

  - Reduction on the parameter space $L = (L_y \; L_\theta)^\intercal$

$$^\flat A = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow L_\theta = \mathbb{1}, \quad \dot{L}_y = AL_y$$

  - General case with model error projection

$$\dot{U} = L^* C^* RCL - UL(L^*L)^{-1}L^*BQB^*(L^*L)^{-1}L^*U$$

- Minimizes the functional
  $\mathscr{J}(\zeta_r, \nu) = \frac{1}{2}\|\zeta_r\|^2_{U_0,\mathcal{Y}_r} + \int_0^T \|z^\delta - Cy_{|\zeta_r,\nu}\|^2_{R,\mathcal{Z}} + \|\nu\|^2_{S,\mathcal{U}} \, \mathrm{d}t$
- The time-discrete optimal filter is a time discretization of the optimal filter

Beyond the deterministic approach

Formalism and difficulties

- the discrete-time formalism:

$$\mathrm{y}_{n+1} = \Phi_{n+1|n}(\mathbf{y}_n) + B_{n+1}\boldsymbol{\nu}_{n+1}, \quad \mathbf{z}_n = C(\mathbf{y}_n) + \boldsymbol{\eta}_n$$

  where $\boldsymbol{\nu}_n$ and $\boldsymbol{\eta}_n$ are white noises or 0 means and covariance Q and W
- the continuous-time formalism

$$\mathrm{d}y = A(y,t)\,\mathrm{d}t + B\,\mathrm{d}w_\nu \text{ with } \mathrm{d}z = C(y)\,\mathrm{d}t + \mathrm{d}w_\eta$$

- for PDEs: White noises in time or in space and time (Bensoussan, 1971)
- Note that when the discrete time correspond to a time discretization of order Δt,

$$W_n = \mathbb{C}\mathrm{ov}(\boldsymbol{\eta}_n) = \frac{1}{\Delta t^2}\int_{t_n-\Delta t}^{t_n}\int_{t_n-\Delta t}^{t_n} \mathbb{E}(\boldsymbol{\eta}(t)\boldsymbol{\eta}(t'))\,\mathrm{d}t\,\mathrm{d}t' = \frac{1}{\Delta t}W(t_n)$$

The finite dimensional linear case equivalence

- Estimator dynamics

$$
\begin{cases}
\text{Initialization :} \\
\hat{y}_0^- = \mathbb{E}(\mathbf{y}_0) = \mathrm{y}_\diamond \\
\Pi_0^- = \mathbb{C}\mathrm{ov}(\mathbf{y}_0 - \hat{y}_0^-) = \Pi_\diamond \\
\text{Correction } (n \geq 0)\text{:} \\
\hat{y}_n^+ = \mathbb{E}(\mathbf{y}_n | z_0, \ldots, z_n) = \hat{\mathrm{y}}_n^- + \Pi_n^- \mathrm{C}^\intercal \left(\mathrm{C}\Pi_n^- \mathrm{C}^\intercal + \mathrm{W}_n\right)^{-1} \left(z_n - \mathrm{C}\hat{y}_n^-\right), \\
\Pi_n^+ = \mathbb{C}\mathrm{ov}(\mathbf{y}_n - \hat{y}_n^+), \\
\text{Prediction } (n \geq 0)\text{:} \\
\hat{y}_{n+1}^- = \mathbb{E}(\mathbf{y}_{n+1} | z_0, \ldots, z_n) = \Phi_{n+1|n} \hat{\mathrm{y}}_n^+, \\
\Pi_{n+1}^- = \mathbb{C}\mathrm{ov}(\mathbf{y}_{n+1} - \hat{y}_{n+1}^-) = \Phi_{n+1|n} \Pi_n^+ \Phi_{n+1|n}^\intercal + \mathrm{B}_{n+1} \mathrm{Q}_{n+1} \mathrm{B}_{n+1}^\intercal,
\end{cases}
$$

The UKF seen as an ensemble Kalman Filter (EnKF)

- Stencil $\mathbb{E}_\omega(e^{(\cdot)}) \overset{\text{def}}{=} \sum_{i=1}^{\mathrm{N}_p} \omega^i e^{(i)} = 0, \quad \mathbb{C}\mathrm{ov}_\omega(e^{(\cdot)}) \overset{\text{def}}{=} \sum_{i=1}^{\mathrm{N}_p} \omega^i e^{(i)} e^{(i)\intercal} = \frac{1}{\rho}\mathbb{1}$
- Algorithm:

$$\begin{aligned}
\text{Initialization:}\quad & \hat{y}_0^- = y_\diamond \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \varPi_0^- = \varPi_\diamond \\
\text{Sampling:}\quad & \zeta_n^{(i)} = J(z_n, \hat{y}_n^- + \sqrt{\rho \varPi_n^-} e^{(i)}),\ 1 \leq i \leq \mathrm{N}_p \\
\text{Correction:}\quad & G_n = \sqrt{\varPi_0^n} \mathbb{C}\mathrm{ov}_\omega\big(e^{(\cdot)}, \zeta_n^{(\cdot)}\big)[\mathbb{C}\mathrm{ov}_\omega\big(\zeta_n^{(\cdot)}\big) + W_n]^{-1}, \\
& \hat{y}_n^+ = \hat{y}_n^- + G_n\big(z_n - C_n \hat{y}_n^-\big), \\
& \varPi_n^+ = \varPi_n^- - G_n[\mathbb{C}\mathrm{ov}_\omega\big(\tilde{z}_n^{(\cdot)}\big) + W_n] G_n^\intercal \\
\text{Sampling:}\quad & \hat{y}_n^{(i)+} = \hat{y}_n^+ + \sqrt{\rho \varPi_n^+} e^{(i)},\ 1 \leq i \leq \mathrm{N}_p \\
\text{Prediction:}\quad & \hat{y}_{n+1}^{(i)-} = \varPhi_{n+1|n}\big(\hat{y}_n^{(i)+}\big),\ 1 \leq i \leq \mathrm{N}_p \\
& \hat{y}_{n+1}^- = \mathbb{E}_\omega\big(\hat{y}_{n+1}^{(\cdot)-}\big), \\
& \varPi_{n+1}^- = \mathbb{C}\mathrm{ov}_\omega\big(\hat{y}_{n+1}^{(\cdot)-} - \hat{y}_{n+1}^-\big) + B_n Q_n B_n^\intercal,
\end{aligned}$$

References

D. Auroux and J. Blum. Back and forth nudging algorithm for data assimilation problems. *Comptes Rendus Mathematique*, 340(12):873–878, June 2005.

A. Bensoussan. *Filtrage optimal des systèmes linéaires*. Dunod, 1971.

A. Bensoussan, M. C. Delfour, G. Da Prato, and S. K. Mitter. *Representation and Control of Infinite Dimensional Systems*. Birkhauser Verlag, second edition edition, 2007. ISBN 3-7643-3641-2.

A. Germani, L. Jetto, M Piccioni SIAM journal on control optimization, , and 1988. Galerkin approximation for optimal linear filtering of infinite-dimensional linear systems. 26(6):1287–1305, 1988.

R.E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng*, 82(1):35–45, 1960.

R.E. Kalman and R.S. Bucy. New Results in Linear Filtering and Prediction Theory. *J. Basic Eng*, 83(1):95–108, March 1961.

P. Moireau. A Discrete-time Optimal Filtering Approach for Non-linear Systems as a Stable Discretization of the Mortensen Observer. *ESAIM: Control, Optimisation and Calculus of Variations*, 2017. doi: 10.1051/cocv/2017077.

R.E. Mortensen. Maximum-likelihood recursive nonlinear filtering. *J. Optim. Theory Appl.*, 2(6):386–394, 1968.

K. Ramdani, M. Tucsnak, and G. Weiss. Recovering the initial state of an infinite-dimensional system using observers. *Automatica*, 46(10):1616–1625, 2010.