DETERMINISTIC KALMAN FILTERING

Jan C. Willems
ESAT-SCD (SISTA)
University of Leuven, Belgium

MESSAGE

There is a deterministic interpretation of
the Kalman filter
that is as convincing as the stochastic one.

OUTLINE

1. The filtering principle
2. Derivation of the deterministic Kalman filter
3. The stochastic filter
4. Remarks

FILTERING

Two time-signals: an *observed* signal,

$$y : [0, \infty) \to \mathbb{R}^{\mathrm{y}},$$

and a *to-be-estimated* signal,

$$z : [0, \infty) \to \mathbb{R}^{\mathrm{z}}.$$

Problem: Find a map

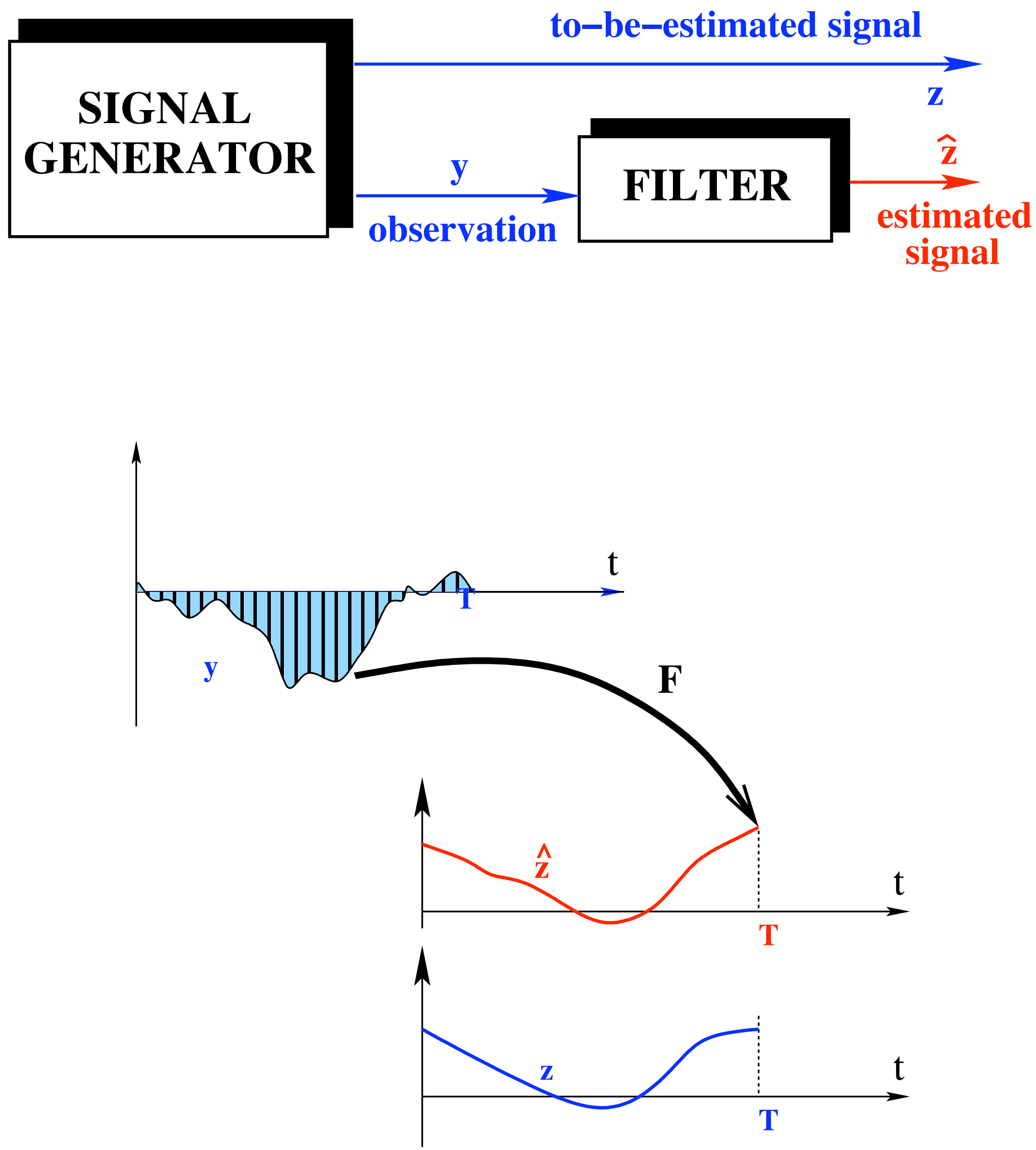
$$\boxed{\mathcal{F} : y \mapsto \hat{z}}$$

so that

$$\hat{z} : [0, \infty) \to \mathbb{R}^{\mathrm{z}}$$

is a 'good estimate' of z.

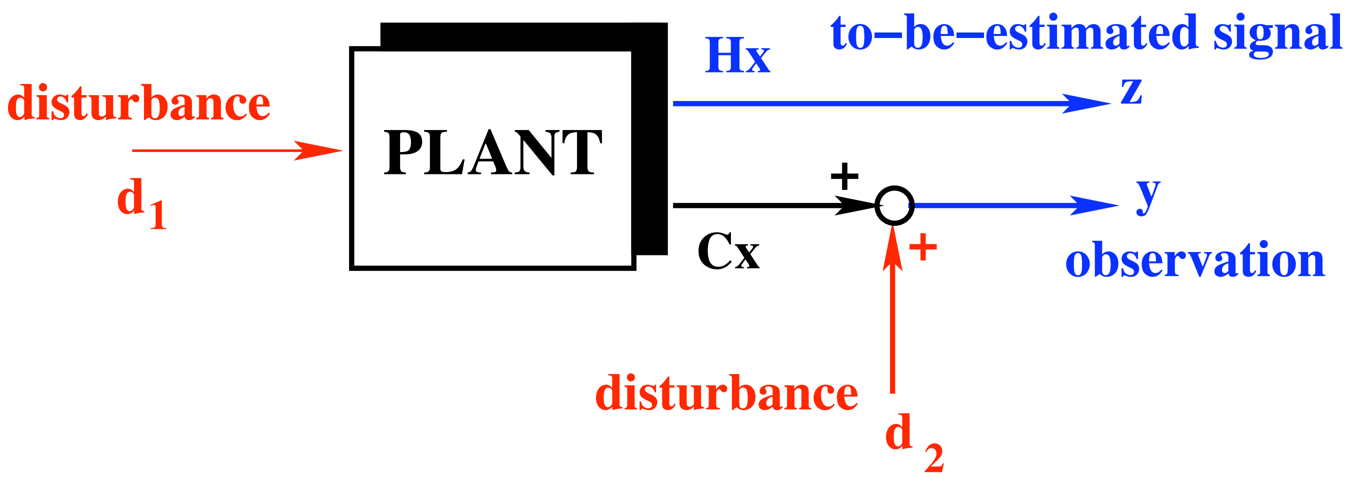
Requirement: $\hat{z}(T)$ at time T is allowed to depend only on the *past* of y: the filter map $\mathcal{F}$ should be *non-anticipating*.


SIGNAL GENERATOR
to−be−estimated signal
z
y
observation
FILTER
ẑ
estimated signal
t
T
y
F
ẑ
t
T
z
t
T

In order to turn this problem into a mathematical one, we need to:

1. **Model the relation between y and z mathematically**
2. **Formulate an estimation principle**
3. **Obtain an algorithm that computes $y \mapsto \hat{z}$,**
   i.e., an algorithm that implements the filter map $\mathcal{F}$

SIGNAL GENERATION MODEL



$$\frac{d}{dt}x = Ax + Gd_1, \quad y = Cx + d_2, \quad z = Hx.$$

Note

$$x(0), d_1(\cdot), d_2(\cdot) \quad \mapsto \quad y(\cdot)$$

$$x(0), d_1(\cdot), d_2(\cdot) \quad \mapsto \quad z(\cdot)$$

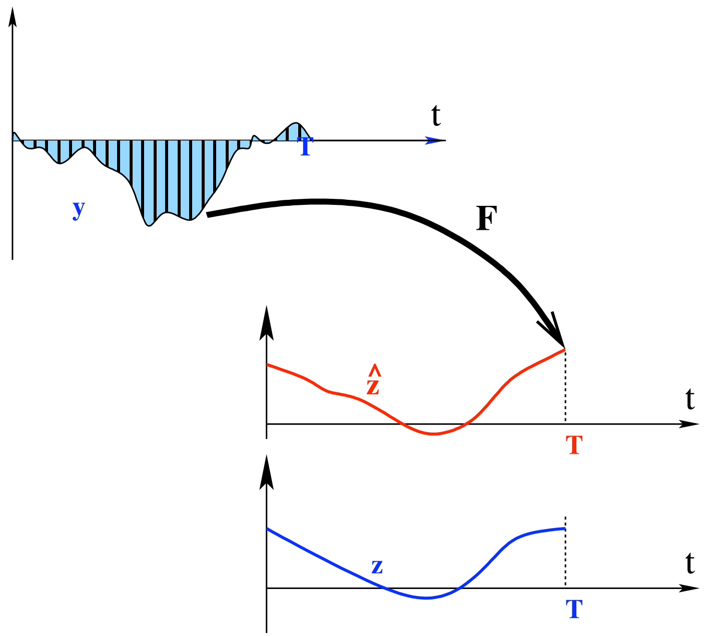
In fact,

$$\begin{aligned} y(t) &= Ce^{At}x(0) + \int_0^t Ce^{A(t-\tau)}Gd_1(\tau)\,d\tau + d_2(t), \\ z(t) &= He^{At}x(0) + \int_0^t He^{A(t-\tau)}Gd_1(\tau)\,d\tau. \end{aligned}$$

There is a 'hidden' vector signal (d_1, d_2) and a 'hidden' initial state $x(0)$ which together generate the observed signal y, and the to-be-estimated signal z.

ESTIMATION PRINCIPLE

What is a rational way of obtaining an estimate $\hat{z}(T)$ of $z(T)$ from $y(t)$ for $0 \le t \le T$?



The deterministic approach put is based on the following idea.

1. Among all the $d_1, d_2, x(0)$ that *'explain'* the observed y, compute the one that minimizes the *uncertainty measure*

$$\boxed{||x(0)||_\Gamma^2 + \int_0^T ||d_1(t)||^2\, dt + \int_0^T ||d_2(t)||^2\, dt.}$$

$\Gamma = \Gamma^\top \succ 0$ is a given matrix.

'Explain' means: consider only those $(d_1, d_2), x(0)$ are compatible with the observed signal y, i.e., for which

$$y(t) \quad = \quad Ce^{At}x(0) + \int_0^t Ce^{A(t-\tau)}Gd_1(\tau)\, d\tau + d_2(t).$$

2. Minimizing $(d_1, d_2), x(0) \quad \rightsquigarrow \quad (d_1^*, d_2^*), x(0)^*$.

Substitute in eq'n for z; resulting output: z^*.

3. Define the desired estimate of $z(T)$ by $\hat{z}(T) := z^*(T)$. Hence

$$\boxed{\hat{z}(T) = He^{AT}x(0)^* + \int_0^T He^{A(T-\tau)}Gd_1^*(\tau)\, d\tau.}$$

Note that $\hat{z}(T)$ depends only on $y(t)$ for $0 \leq t \leq T$:

$\Rightarrow$ non-anticipation.

This estimation principle is reasonable and intuitively acceptable:

Among all $(d_1, d_2), x(0)$ that explain the observations, choose the one that has *'smallest uncertainty measure'*, that is *'most likely'*, ...

FILTERING ALGORITHM

Note that $(d_1^*, d_2^*), x(0)^*$ depends not only on y, but also on T.

So, in order to compute z^* we need to solve,

at each time $T \in [0, \infty)$,

a dynamic optimization problem: minimize the uncertainty measure, subject to the dynamic eq'ns and fixed $y(t), 0 \leq t \leq T$.

But, we will obtain a recursive solution, yielding $\hat{z}(T)$ in a very efficient way, *and for all T at once!*

Key lemma: Let $\Sigma : [0, \infty) \to \mathbb{R}^{n \times n}$ be the unique sol'n to the Riccati differential equation

$$\frac{d}{dt}\Sigma = GG^\top + A\Sigma + \Sigma A^\top - \Sigma C^\top C\Sigma, \quad \Sigma(0) = \Gamma^{-1}.$$

Then $\Sigma(t) = \Sigma(t)^\top \succ 0$ for $t \in [0, \infty)$.

Consider the differential equation involving d_1, d_2, x and y

$$\frac{d}{dt}x = Ax + Gd_1, \quad y = Cx + d_2.$$

Define $\hat{x} : \mathbb{R} \to \mathbb{R}^n$ in terms of Σ and y by

$$\frac{d}{dt}\hat{x} = A\hat{x} + \Sigma C^\top(y - C\hat{x}), \quad \hat{x}(0) = 0.$$

Then

$$||x(0)||_{\Gamma}^{2} + \int_0^T ||d_1(t)||^2 \, dt + \int_0^T ||d_2(t)||^2 \, dt$$

$$= ||x(T) - \hat{x}(T)||^2_{\Sigma(T)^{-1}} + \int_0^T ||(d_1 - G^\top \Sigma^{-1}(x - \hat{x}))(t)||^2 \, dt + \int_0^T ||(y - C\hat{x})(t)||^2 \, dt$$

<u>Proof</u>: Verify the following straightforward calculation

$$\frac{d}{dt}[(x - \hat{x})^\top \Sigma^{-1}(x - \hat{x})]$$
$$= ||d_1||^2 + ||d_2||^2 - ||d_1 - G^\top \Sigma^{-1}(x - \hat{x})||^2 - ||y - C\hat{x}||^2,$$

and integrate.

The optimal filter readily follows. Indeed, whenever $d_1, d_2, x(0)$ leads to y, there holds

$$
\begin{aligned}
&||x(0)||_{\Gamma}^2 + \int_0^T ||d_1(t)||^2\, dt + \int_0^T ||d_2(t)||^2\, dt \\
&= ||x(T) - \hat{x}(T)||^2_{\Sigma(T)^{-1}} + \int_0^T ||(d_1 - G^\top \Sigma^{-1}(x - \hat{x}))(t)||^2\, dt \\
&\qquad + \int_0^T ||(y - C\hat{x})(t)||^2\, dt \\
&\qquad\qquad \geq \int_0^T ||(y - C\hat{x})(t)||^2\, dt.
\end{aligned}
$$

with

$$
\frac{d}{dt}\hat{x} = A\hat{x} + \Sigma C^\top (y - C\hat{x}), \quad \hat{x}(0) = 0.
$$

Observe that $\hat{x}$ is a function of y, but not of the specific $(d_1, d_2), x(0)$ that generated y.

Therefore

$$||x(0)||_{\Gamma}^{2} + \int_{0}^{T} ||d_1(t)||^2 \, dt + \int_{0}^{T} ||d_2(t)||^2 \, dt$$

is minimized if, among the $(d_1, d_2), x(0)$ that generate y, we can choose one such that

1. $x(T) = \hat{x}(T)$, and
2. $d_1(t) = G^{\top}\Sigma(t)^{-1}(x(t) - \hat{x}(t))$ for $0 \leq t \leq T$.

Such a choice clearly exists!

This implies that the optimal $(d_1^*, d_2^*), x(0)^*$ yields

$$\boxed{x(T) = \hat{x}(T),} \quad \text{and hence} \quad \boxed{\hat{z}(T) = H\hat{x}(T).}$$

The optimal $(d_1^*, d_2^*), x(0)^*$ is not needed.

Summarizing.

The least squares filter

Let y be the observed output.

Let $\Sigma : [0, \infty) \to \mathbb{R}^{n \times n}$ be the (unique) solution of the RDE

$$\frac{d}{dt}\Sigma = GG^\top + A\Sigma + \Sigma A^\top - \Sigma C^\top C\Sigma, \quad \Sigma(0) = \Gamma^{-1}.$$

The least squares filter is given by

$$\frac{d}{dt}\hat{x} = A\hat{x} + \Sigma C^\top(y - C\hat{x}), \quad \hat{x}(0) = 0, \quad \hat{z} = H\hat{x}.$$

Input: y; output: $\hat{z}$; Σ: filter parameters computed 'off-line'.

THE STOCHASTIC FILTER

The stochastic model that leads to the (classical Kalman) filter is:

$$\frac{d}{dt}x = Ax + Gd_1, \;\; y = Cx + d_2, \;\; z + Hx, \;\; x(0) = x_0,$$

with d_1, d_2 zero mean gaussian white noises, x_0 gaussian, all mutually independent, and

$$\mathcal{E}\{\begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix} \begin{bmatrix} d_1(t+t') \\ d_2(t+t') \end{bmatrix}^\top\} = I\delta(t'), \quad \mathfrak{L}(x_0) = \mathfrak{N}(0, \Gamma^{-1}).$$

The conditional mean ($\hat{z}(T) = \mathcal{E}\{z(T) \mid y(t), 0 \leq t \leq T\}$) $\cong$ maximum likelihood $\cong$ stochastic least squares filter is also

$$\boxed{\frac{d}{dt}\hat{x} = A\hat{x} + \Sigma C^\top(y - C\hat{x}), \quad \hat{x}(0) = 0, \quad \hat{z} = H\hat{x}.}$$

One can write model + filter also in Itô notation.

This filter, and its discrete-time counterpart, is one of the most important algorithms in existence. Exceedingly important in aerospace applications.

<u>Features</u>:

- **Explicit calculation of filter gains by the RDE**
- **Recursivity**
- **Automatic data reduction**
- **Computational complexity**
- **Generality and generalizability**

GENERAL REMARKS

STATIC ESTIMATION

Let d be a $\mathtt{d}$-dimensional real random vector, $\mathfrak{L}(d) = \mathfrak{N}(0, I)$.

!! Estimate $z = Hd$ from observing $y = Cd$.

Well-known:

$$\hat{z} = HC^\top (CC^\top)^{-1} y$$

is the conditional expectation / maximum likelihood estimate.

Deterministic interpretation of this formula:

'Explain' the observed y as generated by the d

of least Euclidean norm that satisfies $y = Cd$,

denote this least squares d by d^*,

define the resulting estimate $\hat{z}$ of z by $\hat{z} = Hd^*$.

Which interpretation is to be preferred, the probabilistic conditional mean/maximum likelihood interpretation, or the deterministic least squares one?

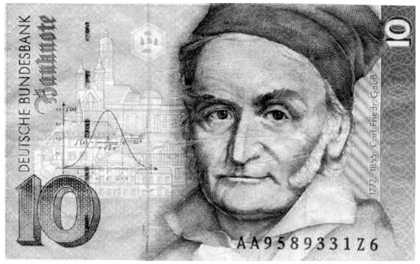
This has been a matter of debate at least since Gauss justified Legendre's least squares, as a method of computing the most probable, maximum likelihood, outcome.

Legendre (least squares)

$\rightsquigarrow$ **Gauss (probability)**

$\rightsquigarrow$ **Wiener & Kolmogorov (time-series, probability)**

$\rightsquigarrow$ **Kalman (probabilistic, state, recursive filter)**



DEUTSCHE BUNDESBANK
Banknote
10
AA9589331Z6

Professor Rudolf E.
Special guest
EL3 August-9

The uncertainty in models is very often due to such things as model approximation and simplification, neglected dynamics of sensors, quantization in time and space, unknown deterministic inputs, etc.

It is hard to conceive situations in which precise stochastic knowledge about real uncertainty can be justified, as a description of reality.

What does probability mean anyway, in this context

- **Relative frequency?**
- **Degree of belief?**
- **Plausibility?**

Cloudy and fuzzy ...., and, in filtering, needlessly so.

Isn't simple deterministic least squares more satisfactory?
It is more pragmatic, and lays its strengths and weaknesses bare.

There is more

Deterministic static least squares $\cong$ max. likelihood driver.

This equivalence of deterministic least squares generation and a stochastic (cond. mean / max. likelihood) interpretation persists in discrete-time dynamical systems over a finite horizon...

But not for continuous-time dynamic systems, (or estimation over an infinite horizon). Indeed, if d is white noise, then

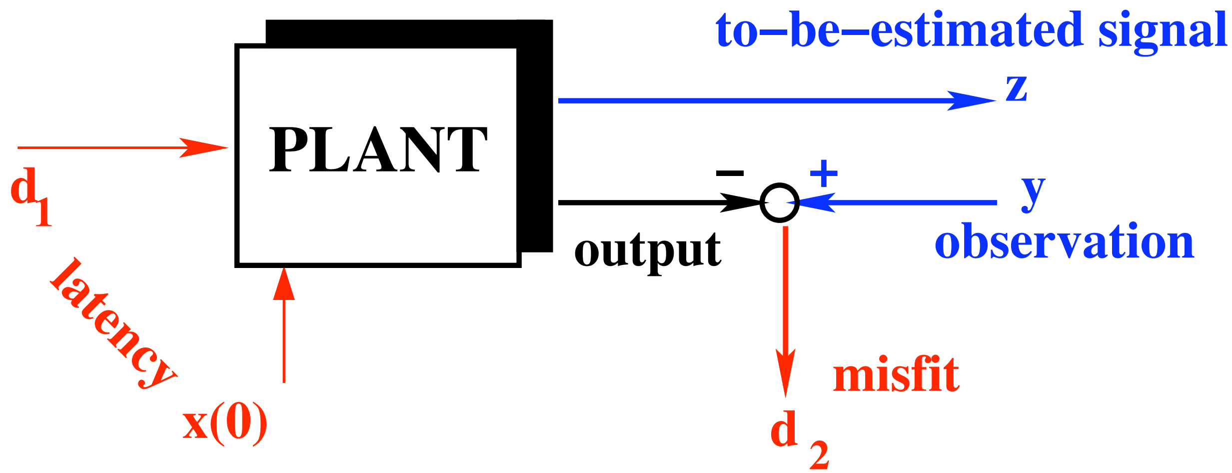
$$\mathcal{E}\{\int_{t_0}^{t_1} |d|^2\, dt\} = \infty \qquad \text{w.p. } 1$$

Hence d's with small $\mathcal{L}_2$-norm are now not 'more likely' than d's with large $\mathcal{L}_2$-norm.

The stochastic Kalman filter does not have an interpretation in terms of the 'most likely $\cong$ least squares driving signal' ...

MY PREFERRED POINT OF VIEW

Two undesirable elements for explaining the observation y:



1. **The ‘latency’ $(x(0), d_1)$ explaining the output Cx**
2. **The ‘misfit’ $y - Cx = d_2$**

yields the uncertainty measure $\cong$

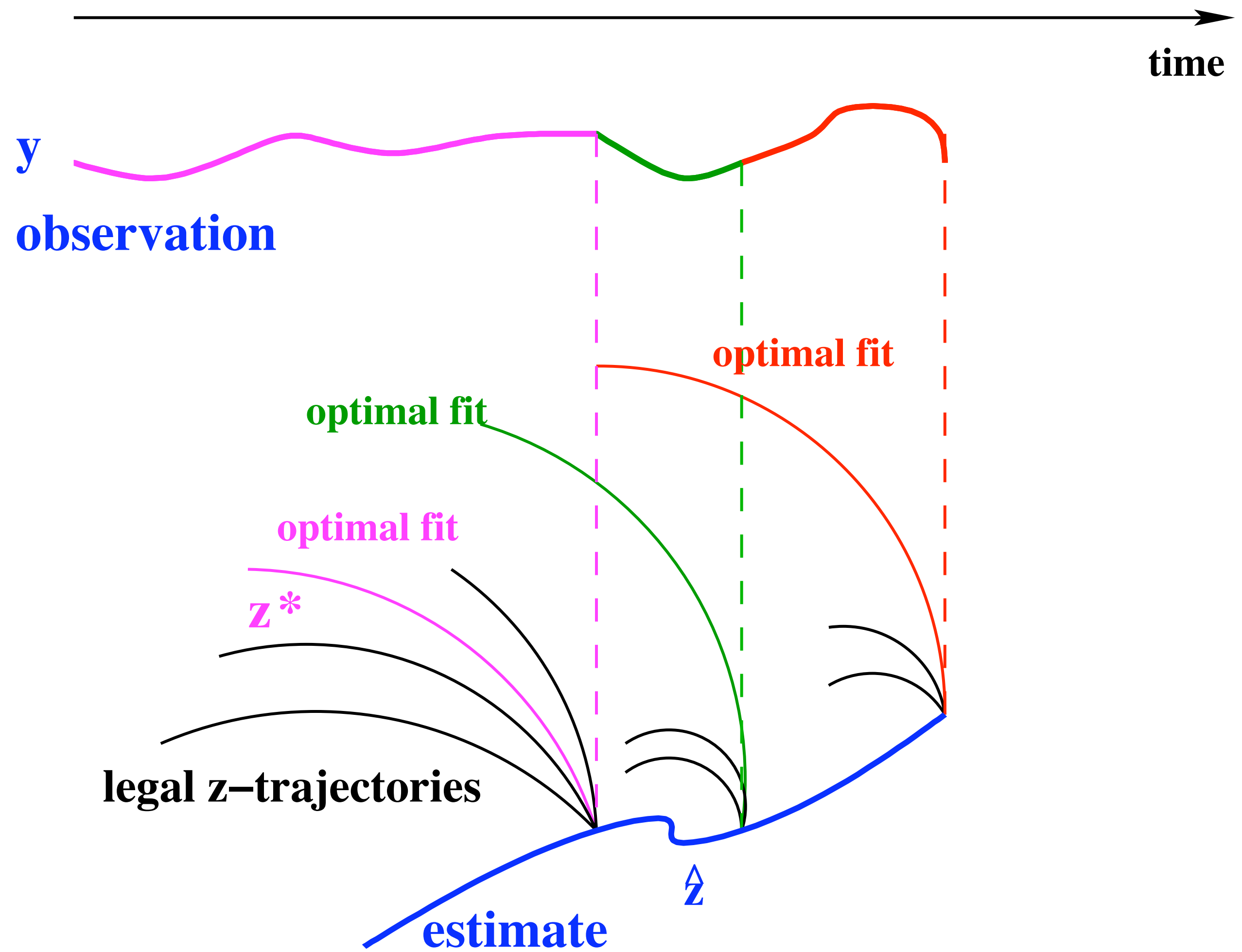
$$\text{the latency measure} = ||x(0)||^2_\Gamma + \int_0^T ||d_1(t)||^2 \, dt$$

$$\text{the misfit measure} = \int_0^T ||d_2(t)||^2 \, dt.$$

Filtering, prediction, etc. $\rightsquigarrow$ **Minimize their (weighted) sum!**

Don't explain the latency $x(0)$, d_1 and the misfit $y - Cx$ stochastically, as driving and sensor 'noise'!
There is no need for it, and often the sensor inaccuracy, if at all significant, is not the dominating difficulty in filtering and system ID.

A graphical interpretation of the deterministic KF:



RECAPITULATION

- Filtering: estimate a signal from the past of an observed one.
- Deterministic least squares: explain the observations by the variables of least norm that generate them; substitute in equations of to-be-estimated signal.
- This leads to the deterministic Kalman filter with the RDE.
- Generalizable in many directions, including least squares control.
- Strictly speaking this result is not of the type:
  deterministic least squares
  $\cong$ stochastic maximum likelihood driving signal.
- Pedagogical advantages of the deterministic derivation are beyond debate.

CONCLUSION

There is a deterministic interpretation of

the Kalman filter

that is as convincing as the stochastic one.

<u>Reference</u>: JCW, Deterministic Kalman filtering, accepted for publication in the *Journal of Econometrics.* Available via

`Jan.Willems@esat.kuleuven.ac.be`

Thank you