DETERMINISTIC KALMAN FILTERING

Jan C. Willems

ESAT-SCD (SISTA), University of Leuven, Belgium
&
Mathematics Department, University of Groningen, NL

MESSAGE

There is a deterministic interpretation of

the Kalman filter

that is as convincing as the stochastic one.

FILTERING

Two (vector-valued) time-signals: an *observed* signal,

$$y : [0, \infty) \longrightarrow \mathbb{R}^{\mathtt{y}},$$

and a *to-be-estimated* signal,

$$z : [0, \infty) \longrightarrow \mathbb{R}^{\mathtt{z}}.$$

<u>Problem</u>: Find a map

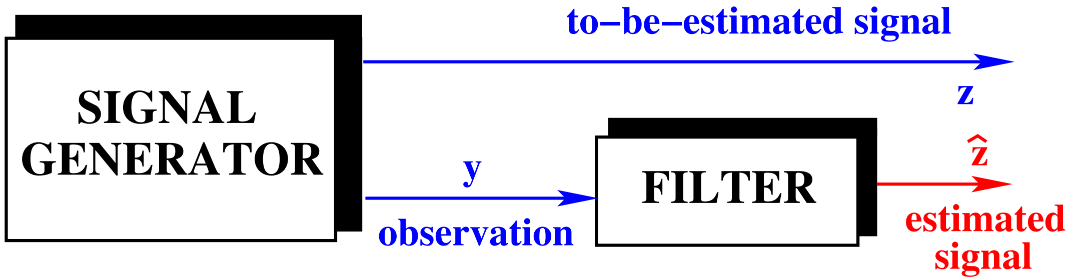
$$\boxed{\mathcal{F} : y \longmapsto \hat{z}}$$

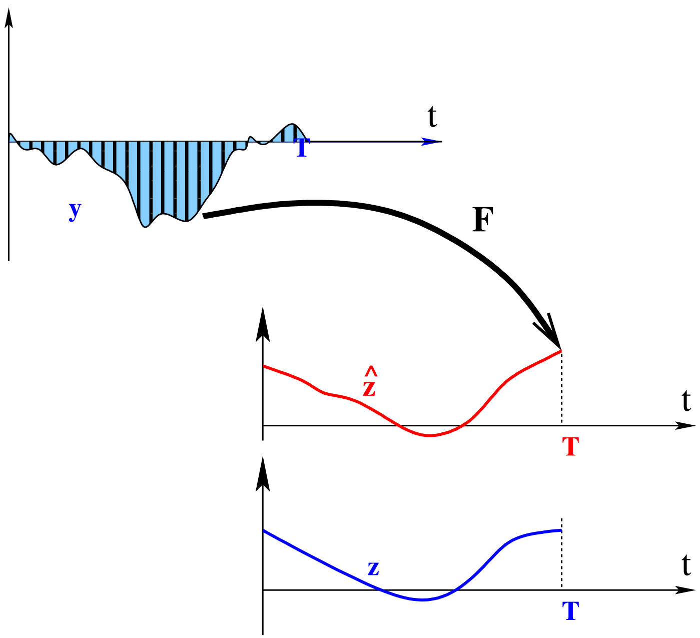
so that

$$\hat{z} : [0, \infty) \longrightarrow \mathbb{R}^{\mathtt{z}}$$

is a 'good estimate' of z.

<u>Requirement:</u> $\hat{z}(T)$ 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

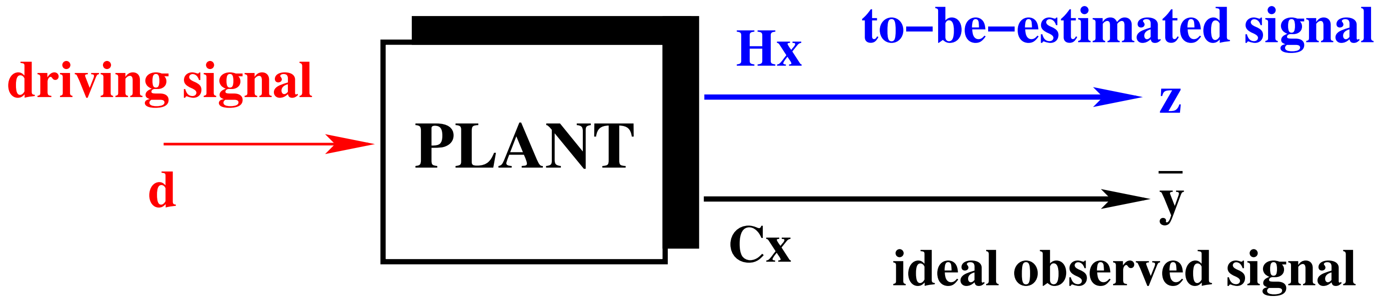

t
T
y
F
ẑ
t
T
z
t
T

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

1. **Model the relation between the to-be-estimated signal z and the observed signal y 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 $\mathcal{F}$**

The most natural setting is in terms of behaviors (see F. Fagnani & JCW, (*System & Control Letters*), but ...

SIGNAL GENERATION MODEL



Postulate the following behavioral equation relating the
to-be-estimated signal $z : [0, \infty) \longrightarrow \mathbb{R}^{\mathrm{z}}$
to an ideal observed signal $\bar{y} : [0, \infty) \longrightarrow \mathbb{R}^{\mathrm{y}}$

$$\boxed{\frac{d}{dt}x = Ax + Gd, \;\; z = Hx, \;\; \bar{y} = Cx.}$$

Note

$$x(0), d(\cdot) \quad \longmapsto \quad z(\cdot)$$

$$x(0), d(\cdot) \quad \longmapsto \quad \bar{y}(\cdot)$$

In order to fix the ideas, keep in mind the following example:

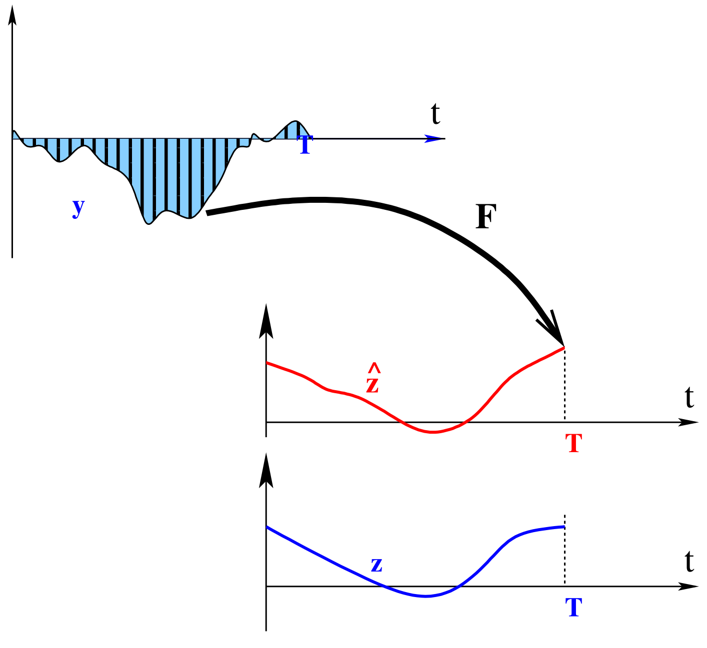
- **Observations y: the position of a moving vehicle, from a sensor that samples, quantizes, averages, ...**
- **To-be-estimated signal z: the velocity of the center of gravity of the vehicle.**
- **Driving input d: the accelaration (from an unknown force)**
- **Ideal observations $\bar{y}$: the actual position of the center of gravity.**

ESTIMATION PRINCIPLE

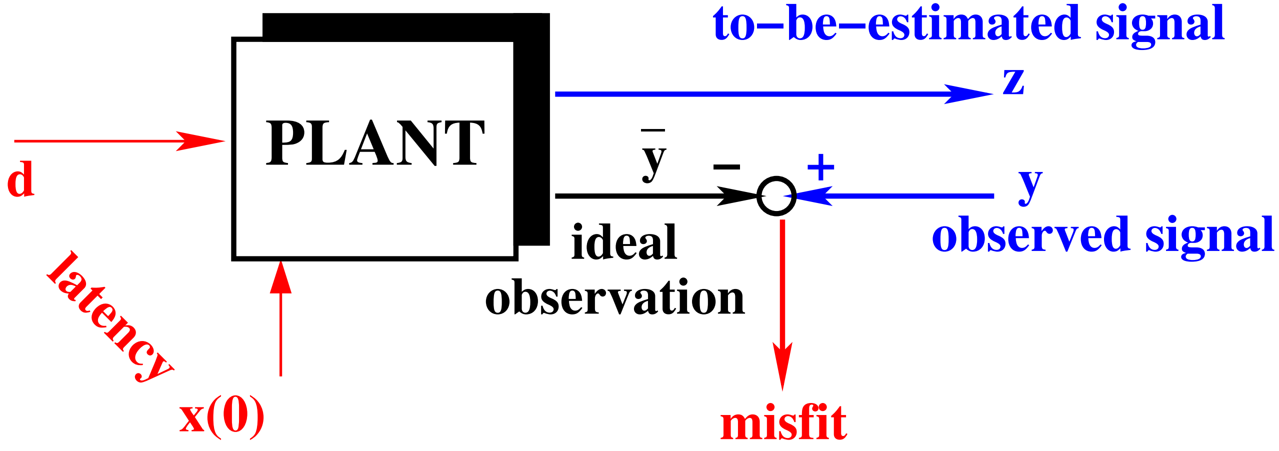
What is a rational way

of obtaining an estimate $\hat{z}(T)$ of $z(T)$

from $y(t)$ for $0 \leq t \leq T$?



Our deterministic approach is based on the following idea.



There are two undesirable elements in this model as an explanation for the observation y:

1. **The 'latency' $(x(0), d)$ explaining the output $\bar{y}$**

2. **The 'misfit' $y - \bar{y}$**

This leads to

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

$$\text{the misfit measure} = \int_{0}^{T} ||y(t) - \bar{y}(t)||^{2}\, dt.$$

$\Gamma = \Gamma^{\top} \succ 0$ a given weighting matrix.

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

$\Rightarrow$ Compute the $d, x(0)$ that minimizes the *uncertainty measure*

$$\boxed{||x(0)||_{\Gamma}^{2} + \int_{0}^{T} ||d(t)||^{2}\, dt + \int_{0}^{T} ||y(t) - \bar{y}(t)||^{2}\, dt.}$$

uncertainty $=$ latency $+$ misfit

Minimizing over $d, x(0) \quad \rightsquigarrow \quad d^*, x(0)^*$.

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

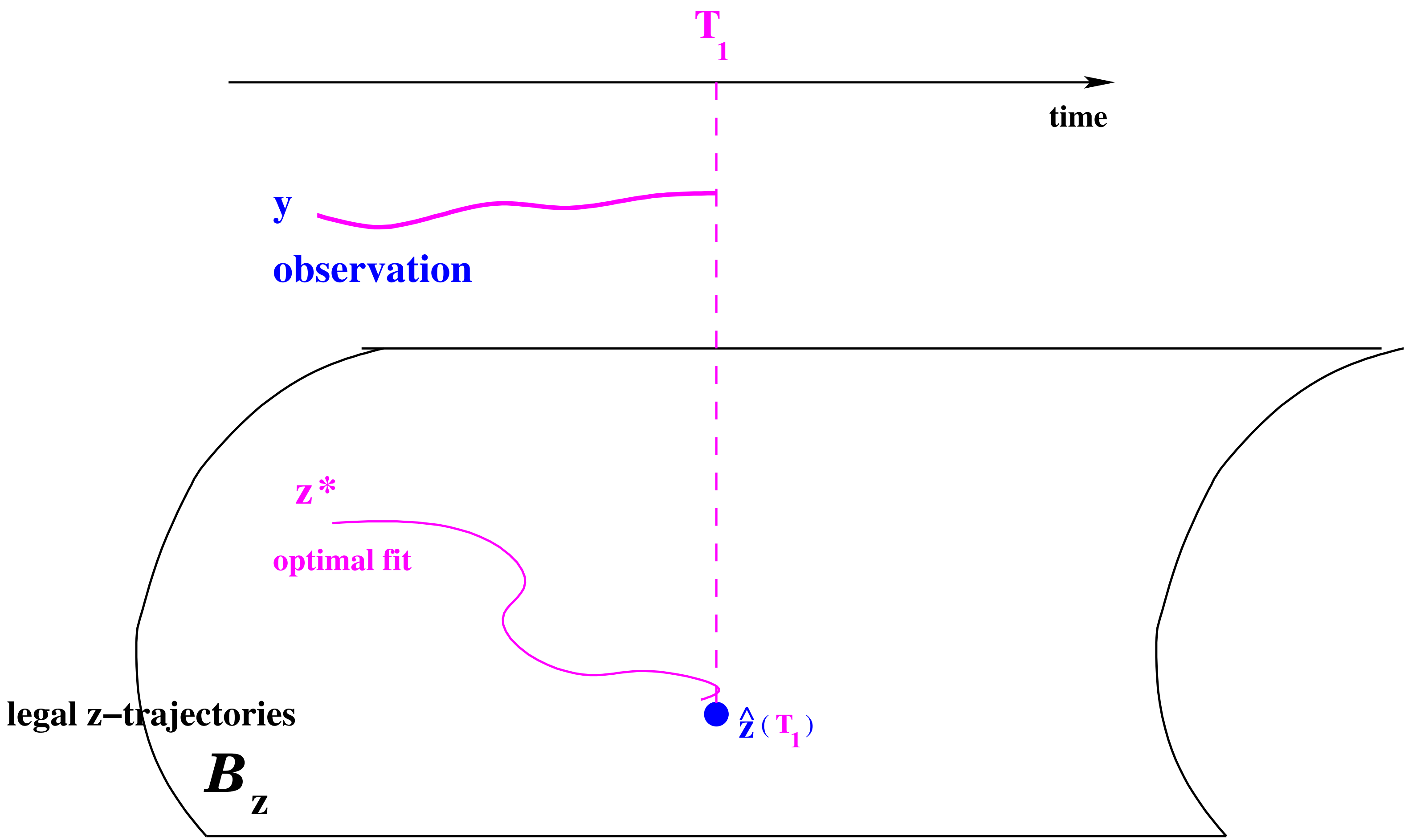
Define the desired estimate of $z(T)$ **by** $\hat{z}(T) := z^*(T)$. **Hence**

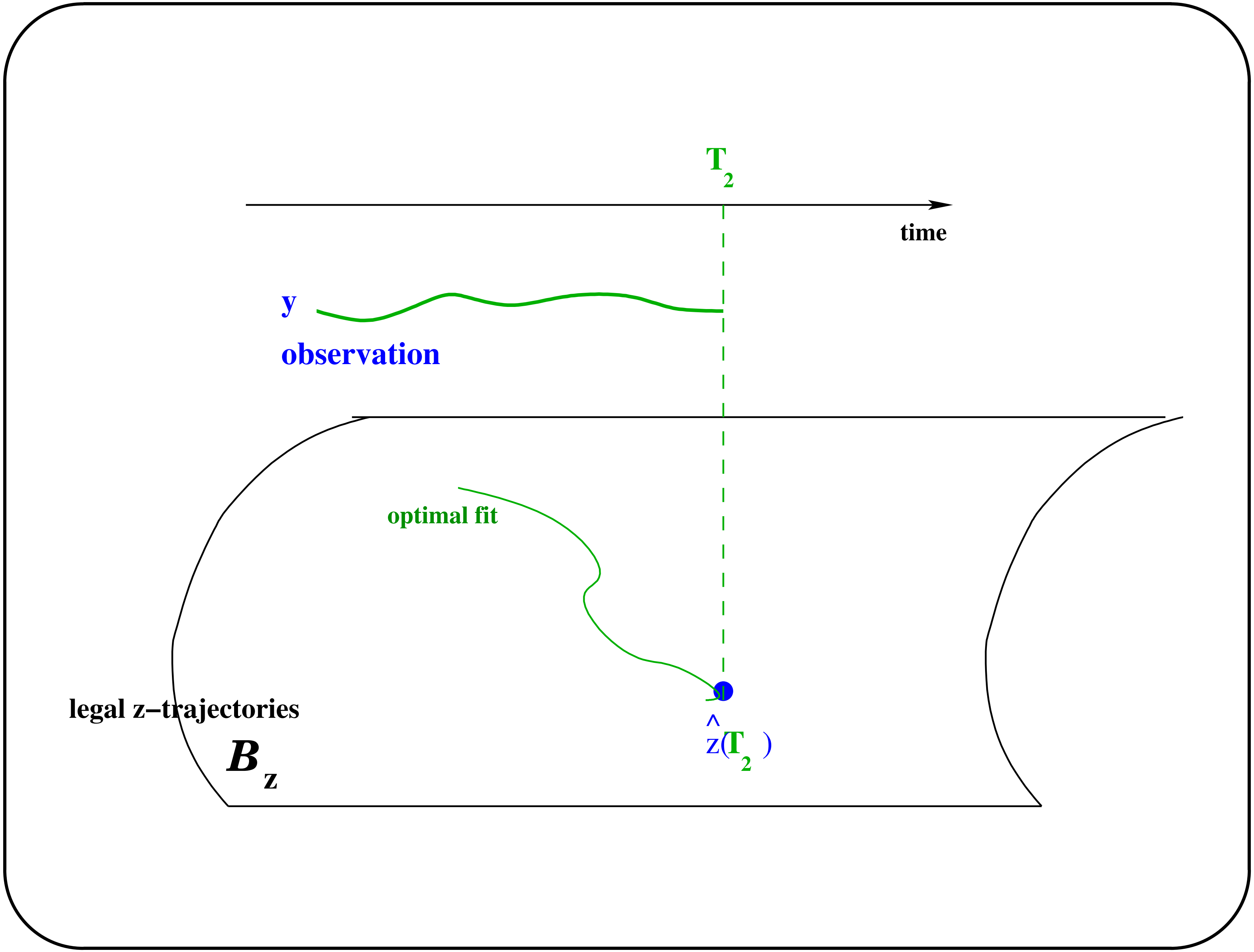
$$\boxed{\hat{z}(T) = He^{AT}x(0)^* + \int_0^T He^{A(T-\tau)}Gd^*(\tau)\, d\tau.}$$

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

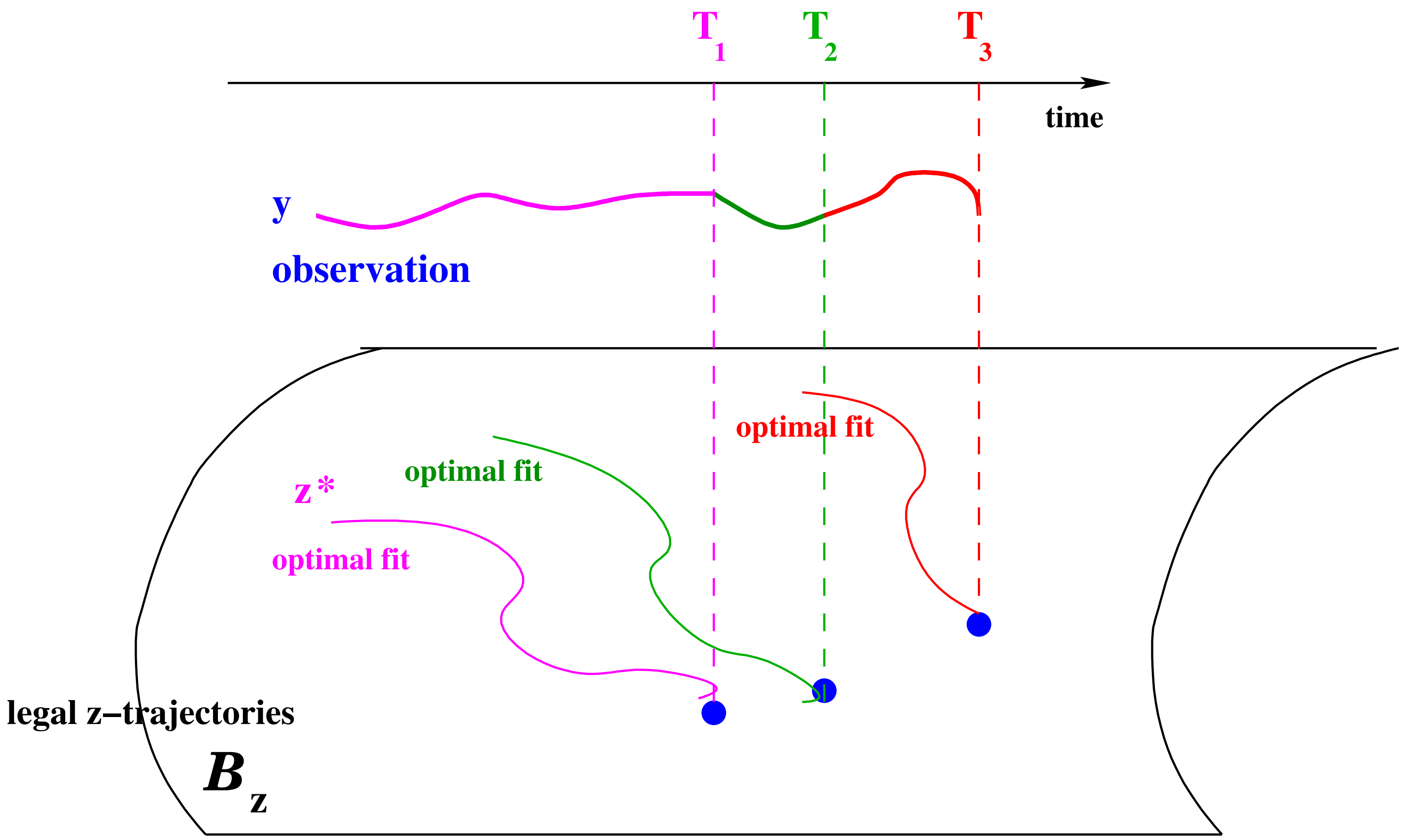
$\Rightarrow$ **non-anticipation.**

A graphical interpretation of this estimation principle:

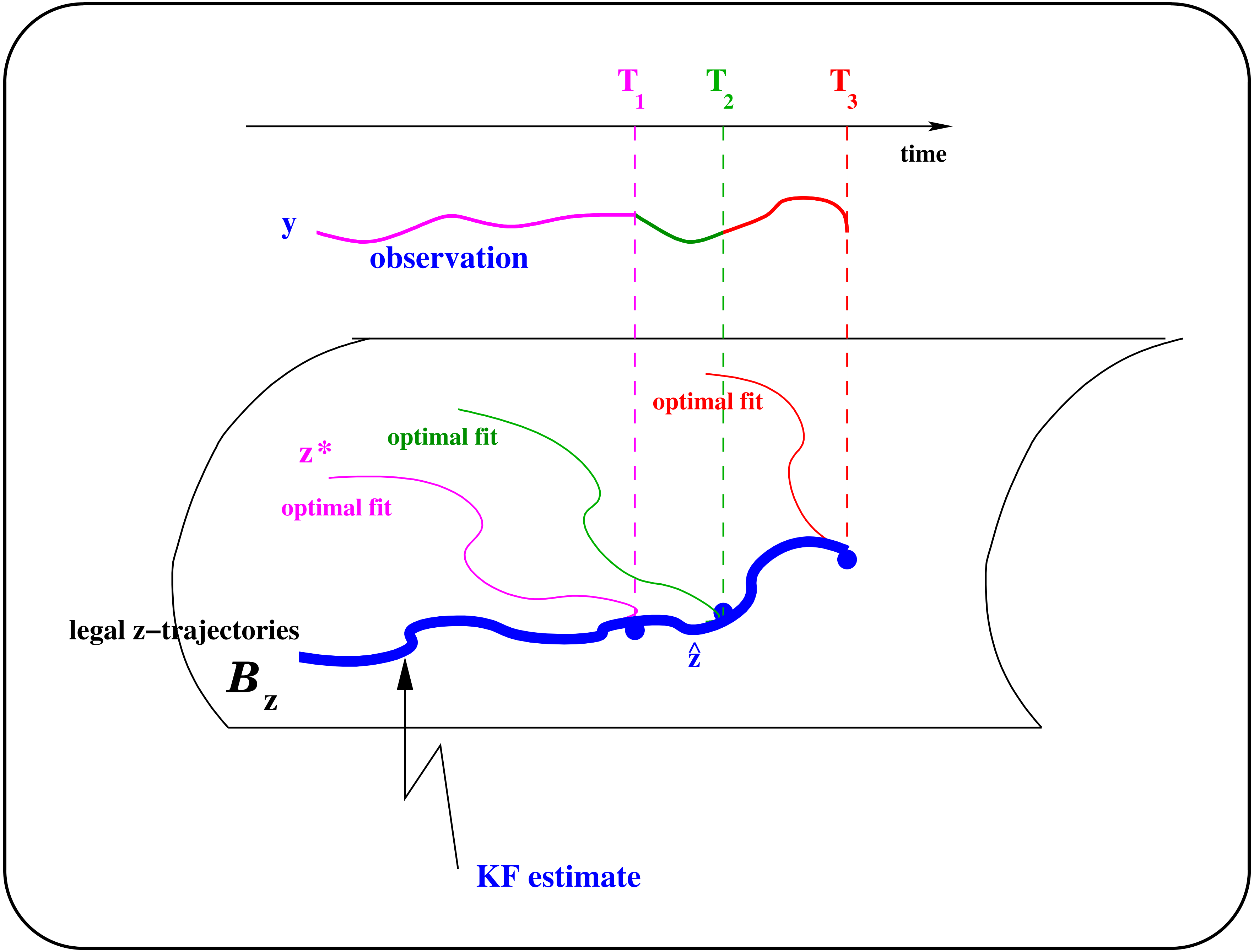




T2
time
y
observation
optimal fit
legal z−trajectories
B z
^
z(T2)



T1
T2
T3
time
y
observation
optimal fit
z*
optimal fit
optimal fit
legal z−trajectories
B z



T1
T2
T3
time
y
observation
optimal fit
optimal fit
z*
optimal fit
legal z−trajectories
B z
ẑ
KF estimate

FILTERING ALGORITHM

$(d^*, x(0)^*)$ **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 with $(y(t), 0 \leq t \leq T$ **fixed).**

It is possible to obtain, using a nice and very effective'completion of the squares' argument, a recursive solution, yielding $\hat{z}(T)$ **in a very efficient way,**

and for all T ***at once!***

Indeed, whenever $d, x(0)$ leads to $\bar{y}$, there holds

$$||x(0)||_\Gamma^2 + \int_0^T ||d(t)||^2\,dt + \int_0^T ||(y - \bar{y}(t)||^2\,dt$$

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

$$+ \int_0^T ||(y - C\hat{x})(t)||^2\,dt$$

$$\geq \int_0^T ||(y - C\hat{x})(t)||^2\,dt.$$

with

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

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

Therefore

$$||x(0)||_{\Gamma}^2 + \int_0^T ||d(t)||^2 \, dt + \int_0^T ||(y - \bar{y})(t)d(t)||^2 \, dt$$

is minimized if we can choose $d, x(0)$ such that

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

Such a choice clearly exists!

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

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

The optimal $d^*, x(0)^*$ 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 this (classical Kalman) filter is:

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

d, n gaussian white noises, x_0 gaussian, independent, and

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

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.

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

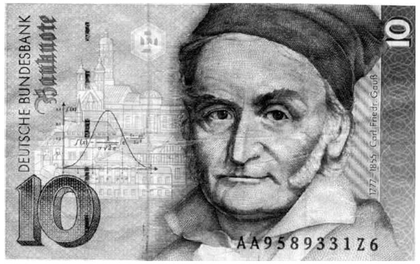
This has been a matter of debate at least since Gauss justified Legendre's least squares as a method of computing the most probable, maximum likelihood, outcome.

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 the present context

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

Cloudy and fuzzy ...,

and, in filtering, as we have shown, needlessly so.

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

But, there is more

Well-known: deterministic static least squares $\cong$ max. likelihood.

This equivalence of deterministic least squares 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' ...

The MORALE of the story

Don't explain

the latency $x(0), d$

and the misfit $y - \bar{y}$

stochastically, as driving and sensor 'noise'.

There is no need for it.

RECAPITULATION

- **Filtering: estimate a signal from the past of an observed one.**
- **Deterministic least squares: explain the observations by the variables of least uncertainty measure 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,
to appear in the *Journal of Econometrics*.

This ms., and copies of the lecture frames are available from/at

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

`http://www.esat.kuleuven.ac.be/~jwillems`

Thank you!