Set Membership Identification of Nonlinear Systems

Mario Milanese, Carlo Novara

Politecnico di Torino

Modern Computational Methods in Applied Mathematics
Banach International Mathematical Center

Bedlewo, Poland June 14-19, 2004

Making inferences from data

- Data y are generated by the nonlinear system f^o :

$$y^{t+1} = f^o(w^t)$$

$$w^t = [y^t \cdots y^{t-n_y} u^t \cdots u^{t-n_u}]$$

u^t : known variables

- The system f^o is unknown, but a finite number of noise corrupted measurements of y^t, w^t are available:

$$\tilde{y}^t = f^o(\tilde{w}^t) + d^t, \quad t = 1, \cdots, T$$

d^t accounts for errors in data $\tilde{y}^t, \tilde{w}^t$

Making inferences from data

- It is desired to make an inference on system f^o :
  - *prediction*
  - *identification*
  - *control, filtering, fault detection*

- The inference is described by the operator $I(f^o, w^T)$
  - *one-step prediction* ➡ $I(f^o, w^T) = f^o(w^T)$
  - *identification* ➡ $I(f^o, w^T) = f^o$

Making inferences from data

- Problems :
  - *for given estimates* $\hat{f} \simeq f^o, \hat{w}^T \simeq w^T$

    evaluate the inference error $\left\| I(f^o, w^T) - I(\hat{f}, \hat{w}^T) \right\|$
  - *find estimates* $\hat{f} \simeq f^o, \hat{w}^T \simeq w^T$

    "minimizing" the inference error

- The inference error cannot be exactly evaluated since f^o and w^T are not known

- Need of prior assumptions on f^o and d^t for deriving finite bounds on inference error

Making inferences from data

- Typical assumptions in literature:
  - on system: $f^o \in \mathcal{F}(\theta) = \left\{ f(w,\theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w,\beta_i) \right\}$
  - on noise: iid stochastic noise

- Functional form of $\mathcal{F}(\theta)$ required:
  - derived from physical laws
  - σ_i: ‘basis’ function (polynomial, sigmoid,..)

- Parameters θ are estimated by minimizing ML or LS functional, which are not convex wrt θ

Set Membership approach

- SM assumptions:
  - on system: $f^{o} \in \mathcal{F}(\gamma) = \left\{ f \in C^{1} : \left\| f^{'}(w) \right\|_{2} \leq \gamma, \forall w \in W \right\}$

    W: bounded set $\in R^{n}$
  - on noise: $\left| d^{t} \right| \leq \varepsilon^{t}, t = 1, .., T$

- Significant improvements obtained by:
  - **use of "local" bound** $\left\| f^{'}(w) \right\|_{2} \leq \gamma(w)$
  - **scaling of regressors *w* to adapt to data**

Set Membership approach

- All information (prior and data) are summarized in the Feasible Systems Set:

$$FSS^T = \left\{ f^o \in \mathcal{F}(\gamma) : |\tilde{y}^t - f^o(\tilde{w}^t)| \le \varepsilon^t, \quad t = 1, \cdots, T \right\}$$

- FSS^T is the set of all systems $\in \mathcal{F}(\gamma)$ that could have generated the data

- Inference algorithm Φ maps all information into estimated inference:

$$\hat{I} = \Phi(FSS^T) \simeq I(f^o, w^T)$$

Prior assumptions validation

- Prior assumptions are invalidated by data
  if FSS^T is empty

- Prior assumptions are considered validated
  if $FSS^T \neq \emptyset$

- The fact that the priors are validated by using the present data does not exclude that they may be invalidated by future data

  (Popper, “Conjectures and Refutations: the Growth of Scientific Knowledge”, 1969)

Prior assumptions validation

- Define:

$$\overline{f}(w) = \min_{t=1,..,T-1} (\overline{h}^t + \gamma \| w - \tilde{w}^t \|_2)$$

$$\underline{f}(w) = \max_{t=1,..,T-1} (\underline{h}^t - \gamma \| w - \tilde{w}^t \|_2)$$

$$\overline{h}^t = \tilde{y}^{t+1} + \varepsilon^t, \; \underline{h}^t = \tilde{y}^{t+1} - \varepsilon^t$$

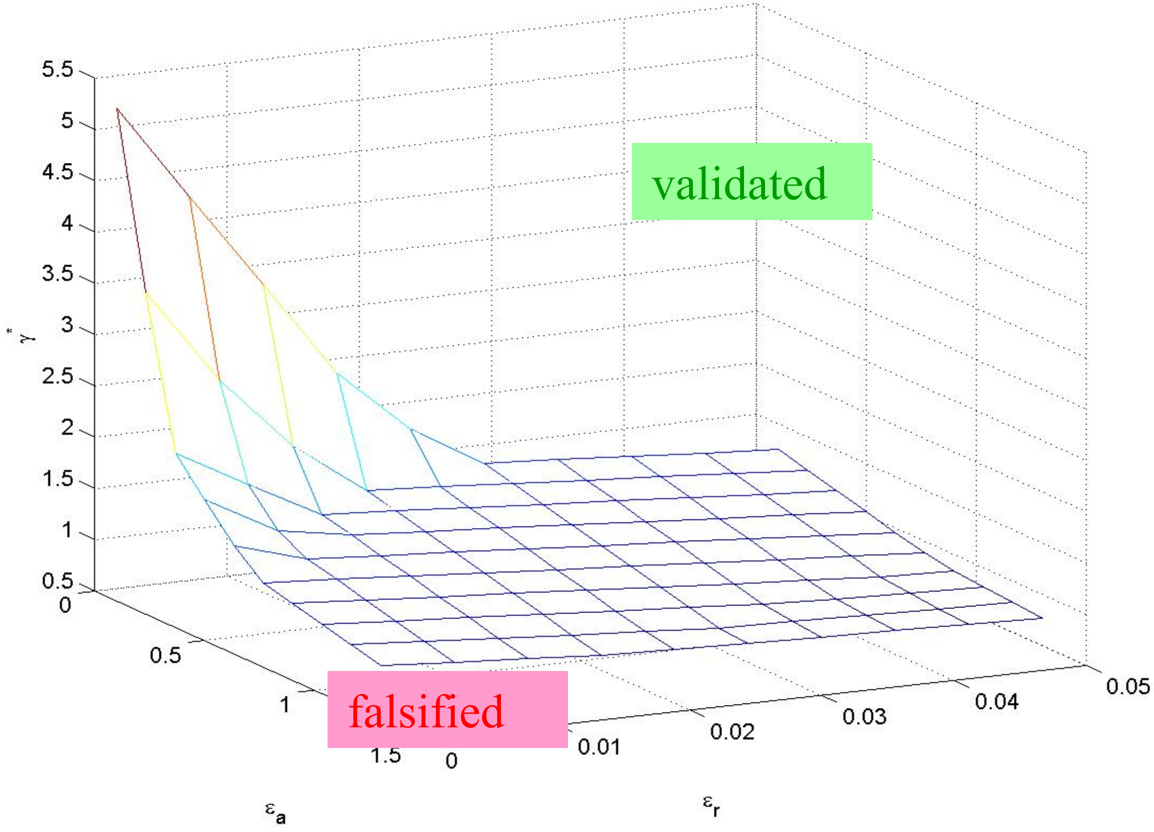
Theorem:

Conditions for $FSS^T \neq \varnothing$ are:

- necessary: $\overline{f}(\tilde{w}^t) \geq \underline{h}^t, t = 1,..,T$
- sufficient: $\overline{f}(\tilde{w}^t) > \underline{h}^t, t = 1,..,T$

Prior assumptions validation

- In space (γ,ε) the surface $\gamma^*(\varepsilon) = \inf_{FSS^T \neq \varnothing} \gamma$ separates falsified values from validated ones



Used for the choice of γ,ε values

Error and optimality concepts

- (Local) Identification $\begin{bmatrix}\text{prediction}\end{bmatrix}$ error:

$$E(\hat{I}) = E[\Phi(FSS^T)] = \sup_{f \in FSS^T} \left(\sup_{|w^T - \tilde{w}^T| \leq \delta^T} \right) \| \Phi(FSS^T) - I(f\ ,w^T) \|$$

- An algorithm Φ^* is optimal if:

$$E[\Phi^*(FSS^T)] = \inf_{\Phi} E[\Phi(FSS^T)] = r \quad \forall FSS^T$$

  - ***r***: (local) radius of information

- An algorithm Φ^α is α-optimal if:

$$E[\Phi^\alpha(FSS^T)] \leq \alpha \inf_{\Phi} E[\Phi(FSS^T)] \quad \forall FSS^T$$

Inference ⇒ Identification: $I(f, w^T) = f$

- Let $|| I(f,w^T) || = || f ||_p = [\int_W | f(w) |^p dw]^{1/p}$
- Define $f^c(w) = \frac{1}{2}[\underline{f}(w) + \overline{f}(w)]$

Theorem:

***i*) The identification algorithm $\Phi^c(FSS^T) = \boldsymbol{f}^c$**

is optimal for any L_p norm, $1 \leq p \leq \infty$

***ii*) The radius of information $\boldsymbol{r}$ is:**

$$E[f^c] = r = \frac{1}{2} \| \overline{f} - \underline{f} \|_p$$

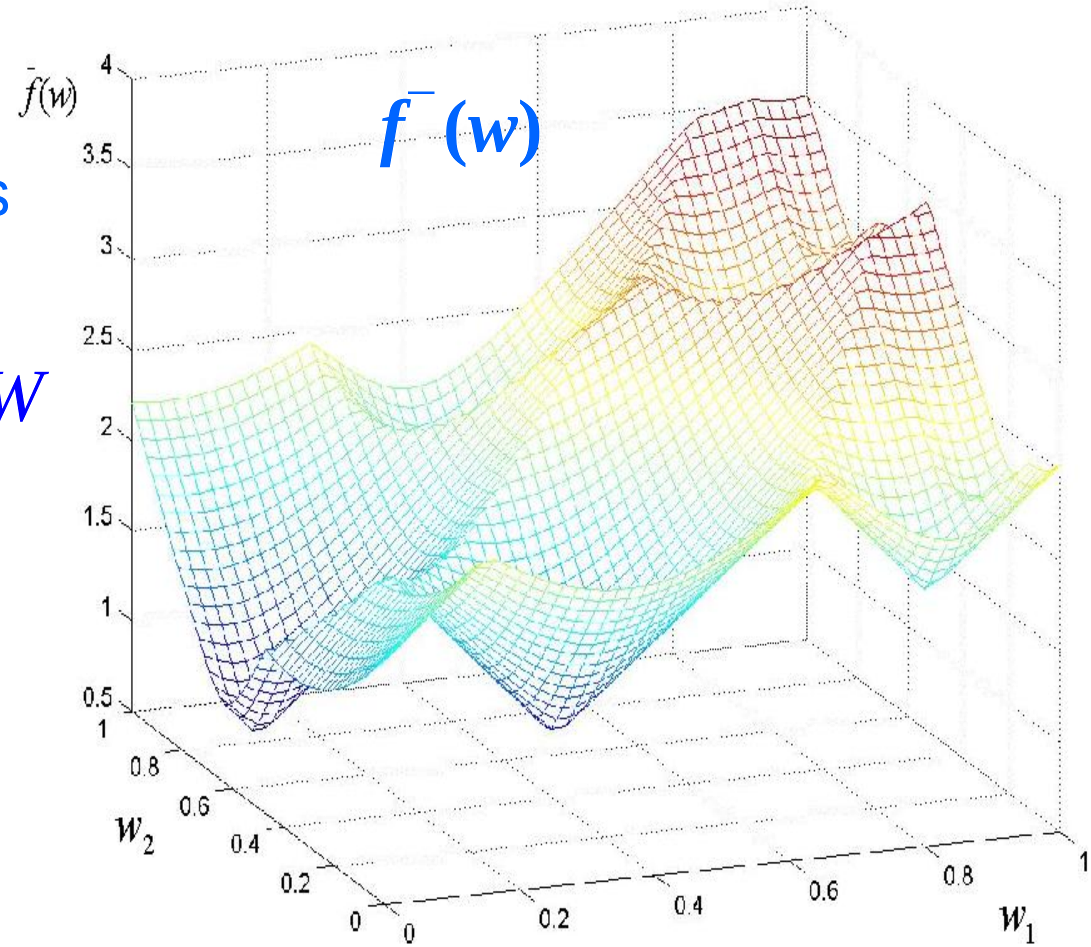
Properties of $\underline{f}$, $\bar{f}$ and f^c

i) $\underline{f}(w)$, $\bar{f}(w)$ are piece-wise conic Lipschitz continous functions with constant γ

ii) $\underline{f}(w)$, $(\bar{f}(w))$ is differentiable

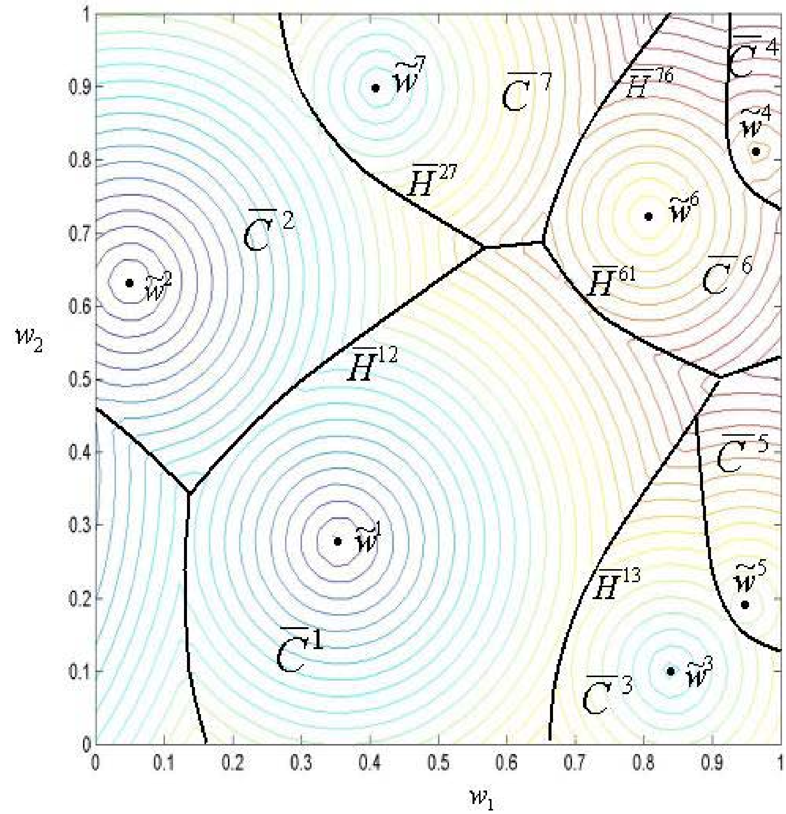
$\forall w \in \underline{M}\ (\bar{M}) \subset W$

$co\underline{M}\ (co\bar{M})$: sets of zero measure



Properties of f , f^{-} and f^{c}

- $co\underline{M}$ and $co\overline{M}$ are the faces of the Hyperbolic Voronoi Diagrams (HVD) generated by f , f^{-}
- HVD are generalization of standard Voronoi Diagrams Edelsbrunner, "*Combinatorial Geometry*", Springer 1987



HVD generated by f^{-}

Properties of $\underline{f}$, $\bar{f}$ and f^c

- Let $M = \underline{M} \cup \bar{M}$

Theorem:

i) The optimal estimate $f^c(w)$ is Lipschitz continous in W with constant γ

ii) $f^c(w)$ is differentiable $\forall w \in M \subset W$ and:

$$\left\| f^{c'}(w) \right\|_2 \leq \gamma, \forall w \in M \subset W$$

- Note: coM has zero measure

Inference ⇒ Prediction: $I(f, w^T) = f(w^T)$

- Let: $\| I(f, w^T) \| = | f(w^T) |$
- Assume: $\left| d^t \right| = \varepsilon^t + \gamma \delta^t$
- Let: $B_\delta(\tilde{w}^t) = \left\{ w \in W : \left\| w - \tilde{w}^t \right\|_2 \leq \delta^t \right\}$
- Since the HVD generated by $\underline{f}, \bar{f}$ give a complete partition of W in cells $\underline{C}^t$ and $\bar{C}^t$, then:

$$\tilde{w}^t \in \underline{C}^t \cap \bar{C}^t$$

Inference ⇨ **Prediction:** $I(f, w^T) = f(w^T)$

Theorem:

***i*)** The prediction algorithm $\Phi^c(FSS^T) = f^c(\tilde{w}^T)$

is 2-optimal, with prediction error bounded by:

$$E\left[\Phi^c\left(FSS^T\right)\right] \leq \frac{1}{2}\left[\overline{f}(\tilde{w}^T) - \underline{f}(\tilde{w}^T)\right] + \gamma\delta^T$$

***ii*)** If $B_\delta(\tilde{w}^T) \subset \underline{C}^T \cap \overline{C}^T$, then prediction $\hat{y}^{T+1} = f^c(\tilde{w}^T)$

is optimal and the radius of information is:

$$E\left[\Phi^c\right] = r = \frac{1}{2}\left[\overline{f}(\tilde{w}^T) - \underline{f}(\tilde{w}^T)\right] + \gamma\delta^T$$