SMIT: Theory of Inference Making From Data

Application to Identification, Prediction, Filtering & Control of Nonlinear Systems

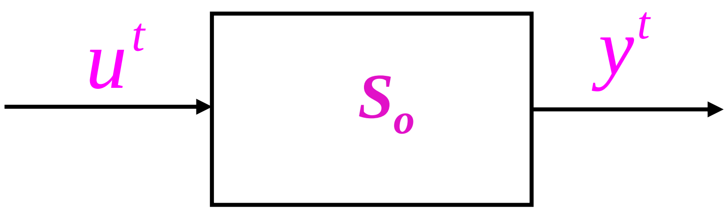
M. Milanese, Politecnico di Torino and Modelway

Sessione in ricordo di Roberto Tempo
Convegno SIDRA 2017, 13 Settembre, Milano

- In this talk the SM theory of inference making from data, named SMIT (Set Membership Inference Theory), is presented, to which Roberto dedicated his first years of scientific activity
- The theory allows to derive innovative methods for identification, prediction, filtering and control of linear and nonlinear systems
- Key feature of SMIT is to account that:
  - only finite number of noisy data is available
  - only approximate modeling is achievable

Theory of Inference Making from Data

- **It is of interest to make some inference about a system S_o**



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

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

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

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

accounts for noisy data $\tilde{y}^t, \tilde{w}^t$

Theory of Inference Making from Data

- **Desired inference described by** $I^o = I(f_o, w^t)$
  - **Identification:** ➡ $I(f_o, w^t) = f_o$
  - **1-step prediction:** ➡ $I(f_o, w^t) = y^{t+1} = f_o(w^t)$
  - **Filtering:** ➡ . . . . . .
  - **Control:** ➡ . . . . . .
- ***Problems* :**
  - ***for given inference*** $\hat{I} = I(\hat{f}, \hat{w}^t)$ ***evaluate the inference error*** $\| I^o - \hat{I} \| = \| I(f_o, w^t) - I(\hat{f}, \hat{w}^t) \|$
  - ***find*** $\hat{I}$ ***giving "minimal" inference error*** $\| I^o - \hat{I} \|$

MODELWAY
INTELLIGENCE TECHNOLOGY

Theory of Inference Making from Data

- The inference error $\left\| I(f_o, w^t) - I(\hat{f}, \hat{w}^t) \right\|$ cannot be exactly evaluated, being f_o, w^T unknown

- Prior assumptions on f_o and d^t are needed for deriving finite bounds on inference error

Parametric Stochastic (PS) approach

Set Membership (SM) approach

PS and SM approaches

- **PS assumptions:**
  - **on system:** $f_o = f(w, \theta_o), \theta_o \in R^p$
  - **on noise :** **stochastic noise**

- **SM assumptions:**
  - **on system:** $f_o \in \mathcal{F}(\gamma) = \left\{ f \in C^1 : \left\| f^{'}(w) \right\|_2 \leq \gamma, \forall w \in W \right\}$
  - **on noise:** $\left| d^t \right| \leq \varepsilon^t \doteq \eta^t + \gamma \delta^t, t = 1,..,T$

$$| y^t - \tilde{y}^t | \leq \eta^t , \| w^t - \tilde{w}^t \| \leq \delta^t$$

Initial Bibliography

Milanese, Tempo, Vicino: "Robust Time Series Prediction by Optimal Algorithms Theory" 23rd IEEE CDC, Las Vegas 1984

Genesio, Milanese, Tempo, Vicino: "Optimal error and GMDH predictors: a comparison with some statistical tecniques", 4th Int. Symp. on Forecasting, London 1984. Also in Int. J. Forecasting, 1987.

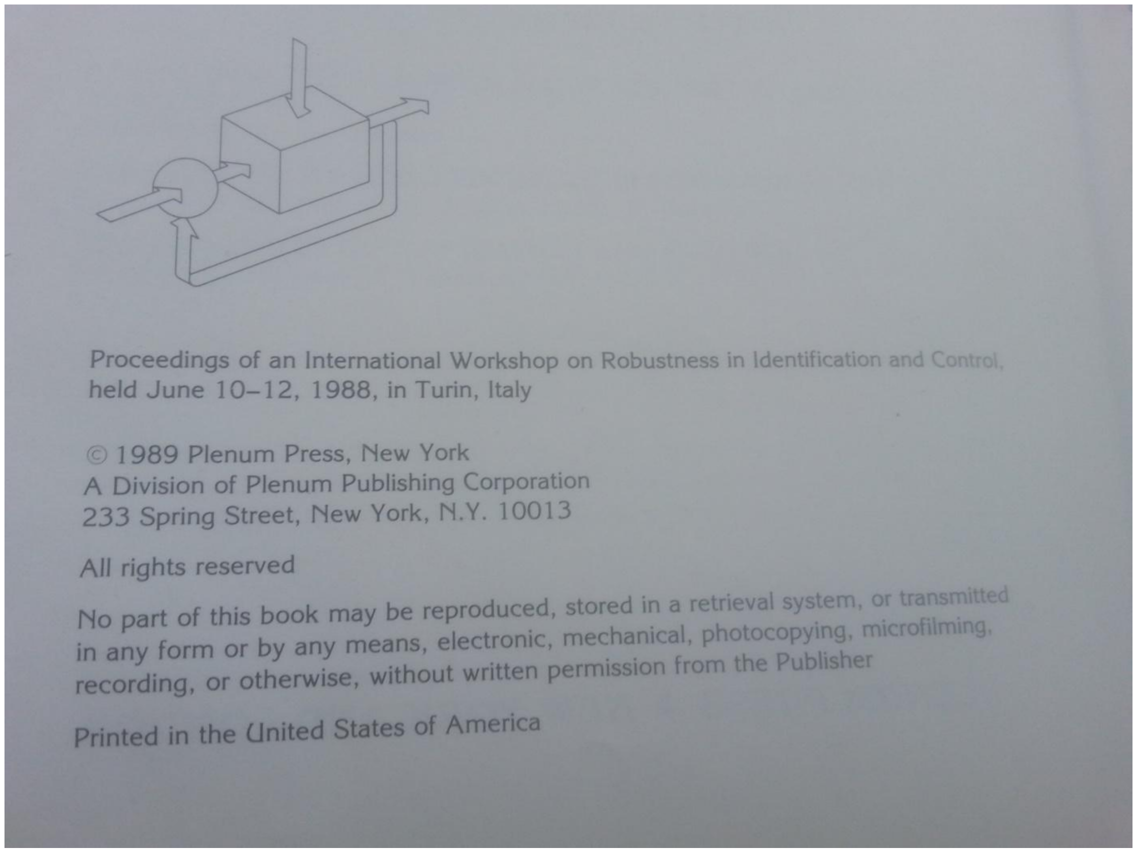
Milanese, Tempo: "Optimal Algorithms Theory for robust estimation and prediction" IEEE Trans. AC, 1985.

Milanese, Tempo, Vicino: "Strongly optimal algorithms and optimal information in estimation problems‟ Symp. on Complexity of Approx. Solved Problems, New York, 1985. Also in Jour. of Complexity, 1986.

M. Milanese, R. Tempo, A. Vicino: "Optimal error predictors for economic models", Int. Journal of Systems Science, 1988.

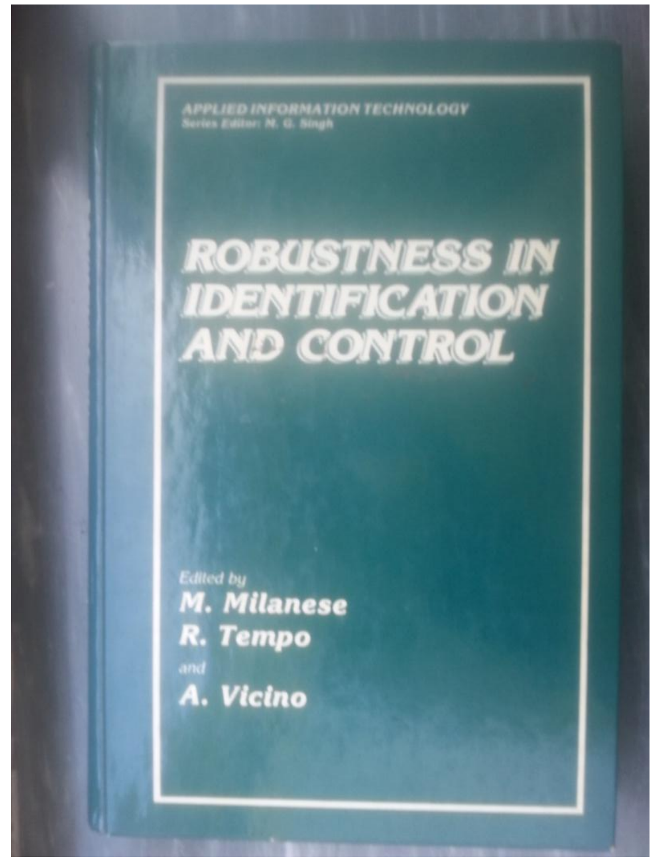
Milanese, Tempo, Vicino Eds., "Robustness in identification and control", Plenum Press 1989.

International Workshop on
“Robustness in Identification and Control”
Torino, June 1988



Proceedings of an International Workshop on Robustness in Identification and Control, held June 10–12, 1988, in Turin, Italy

Printed in the United States of America



Presentation outline

- A brief introduction to SMIT and its use for identification, prediction, filtering and control is presented
- Some basic questions are discussed:
  - what may be gained by using SMIT ?
  - why SMIT give strong results, using weak assumptions ?
  - can SMIT efficiently deal with complex applications ?
- The discussion is supported by presenting results achieved on complex real world applications

SMIT

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

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

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

- **An inference algorithm $\hat{\Phi}$ maps all information, described by FSS^T, into estimated inference $\hat{I} = \Phi(FSS^T)$**

Inference algorithm errors

- The inference error $e(\hat{I}) = \| I(f_o, w^t) - \hat{I} \|$ of estimated inference $\hat{I} = \Phi(FSS^T)$ cannot be exactly known since it is only known that:

$$f_o \in FSS^T, \;\; w^t \in B_\delta(\tilde{w}^t) = \left\{ w : \| w^t - \tilde{w}^t \| \le \delta^t \right\}$$

- The tightest bound on $e(\hat{I})$ can be computed as:

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

Optimality concepts

- **An algorithm** Φ^* **is optimal if:**

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

- **The "central" algorithm** $\Phi^c(FSS^T) = I^c = \text{center of } FSS^T$ **is optimal**

- **The "projection" algorithm** $\Phi^p(FSS^T) = I^p \in FSS^T$ **is almost optimal within a factor 2, i.e.:**

$$E[\Phi^p(FSS^T)] \leq 2 \inf_{\Phi} E[\Phi(FSS^T)] \quad \forall FSS^T$$

Inference: Identification

- **Inference operator:** $I(f_o, w^t) = f_o$
- **Inference algorithm:** $\Phi(FSS^T) = \hat{I} = \hat{f}$
- **Error norm:** $\| I(f_o, w^t) - \hat{I} \| = [\int_W | f_o(w) - \hat{f}(w) |^p \mathrm{d}w]^{1/p}$

- **The algorithm** $\Phi^c(FSS^T) = f^c = \frac{1}{2}[\overline{f} + \underline{f}]$

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

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

is central, i.e. optimal for any L_p norm, $1 \leq p \leq \infty$

Inference: k-step prediction

- **Inference operator:** $I(f_o, w^t) = y^{t+k} = f_o(w^t)$
- **Inference algorithm:** $\Phi(FSS^T) = \hat{I} = \hat{y}^{t+k}$
- **Error norm:** $\| I(f_o, w^t) - \hat{I} \| = | y^{t+k} - \hat{y}^{t+k} |$

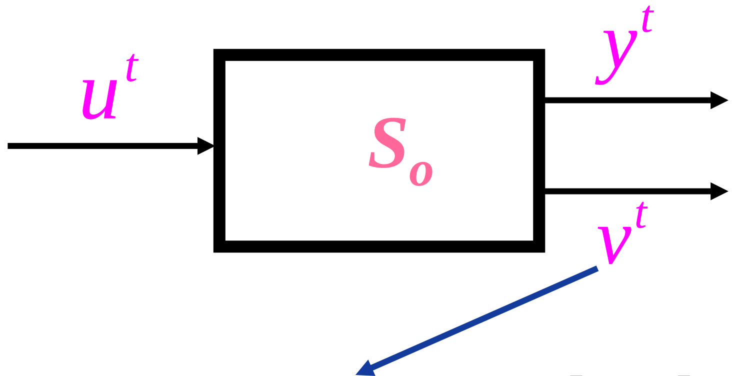
- **The algorithm** $\Phi^p(FSS^T) = \hat{y}^{t+k} = \frac{1}{2} | \overline{f}(\tilde{w}^t) + \underline{f}(\tilde{w}^t) |$

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

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

is almost optimal

Inference: filtering

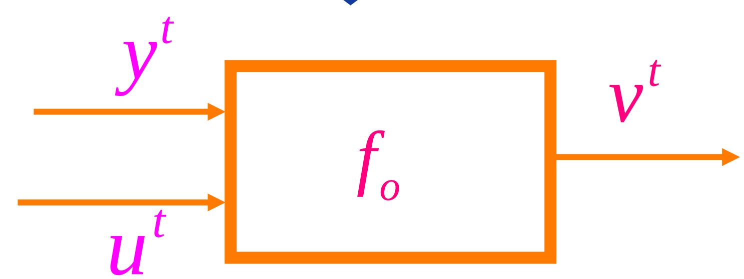


to be estimated from $u^\tau, y^\tau \quad \tau \le t$

- **If S_o is observable, then:**

$$v^t = f_o(w^t)$$

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



Filter Design:

from data $\tilde{u}^t, \tilde{y}^t, \tilde{v}^t, t = 1,..T$

estimate $\hat{f} \cong f_o$

giving $|v^t - \hat{v}^t|$ *'small'*

Inference: filtering

- **Inference operator:** $I(f_o, w^t) = v^t = f_o(w^t)$
- **Inference algorithm:** $\Phi(FSS^T) = \hat{I} = \hat{v}^t$
- **Error norm:** $\| I(f_o, w^t) - \hat{I} \| = | v^t - \hat{v}^t |$

- **The filtering algorithm:**

$$\Phi_{ao}(FSS^T) = \hat{v}^t = \frac{1}{2} | \overline{f}(\tilde{w}^t) + \underline{f}(\tilde{w}^t) |$$

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

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

is optimal

What may be gained by using the SMIT in *identification* and *prediction* ?

- The problems of searching reliable parametric families of models is avoided, which may be difficult in the identification of complex systems
- Optimal identification algorithms are obtained not using iterative optimization algorithms, at difference from what required by most PS methods
- Thus, the problems due to convergence of iterative algorithms to local optima is circumvented

What may be gained by using the SMIT in *identification* and *prediction* ?

- **Finite sample** optimal identification algorithms and error bounds are obtained
- From finite sample results, conditions are derived guaranteeing that the identified model has the Simulation Error Boundedness (SEB) property
- Obtaining similar results in the PS framework is a largely open problem

What may be gained by using the SMIT in *filter design*?

- An optimal filter is derived and boundedness of the filtering error is guaranteed.
- In the PS framework, only approximated filters (e.g. Extended Kalman Filter,..) can be actually computed, whose accuracy and even boundedness is difficult to evaluate
- The optimal filter is derived avoiding the model identification. In the PS framework, a model has to be identified, and a further problem is to evaluate the effects of modeling errors

What may be gained by using the SMIT instead of conventional modeling, filtering, control methods ?

SMIT allows to handle systems having very complex behaviour without requiring extensive first principle laws investigations

Reduction of time & cost in design and embedding up to 80%

Improvement in accuracy up to 10x

These benefits of SMIT approach are documented in several complex real world applications

More information:

mario.milanese@modelway.it

www.modelway.it

www.kitenergy.net

Thank you!