EUROPEAN EMBEDDED CONTROL INSTITUTE (EECI) – INTERNATIONAL GRADUATE SCHOOL ON CONTROL

Learning-Based Predictive Control

Chapter 3
Introduction to Set Membership estimation

Prof. Lorenzo Fagiano
Politecnico di Milano
2023

Learning and control

A control problem is **primarily an estimation problem** (model “learning”).

“Good” model estimate → “good” control system

“Learning-based” control has a long-standing history in control engineering.

- Identification for control ([1],…)
- Dual control theory ([2],…)
- Adaptive control ([3],…)
- Iterative learning control ([4],…)
- …

So, what is new/challenging here?

Estimation in presence of constraints

Established industrial practice:

- Fully exploit **prior information** and collect **data** in ad-hoc experiments, try to reduce uncertainty "as much as possible" (why?);
- Rely on feedback control to deal with the remaining uncertainty;
- Use adaptation to improve performance over time;
- Deal with safety (= constraints) with ad-hoc solutions, e.g., override the control system with emergency procedures;
- Test, test, test

A challenge is to develop theoretically sound approaches that can manage **estimation** and **constraint satisfaction** together (why is it important?)

Control goal

Input ingredients

Prior information

Desired performance

Data

Design choices

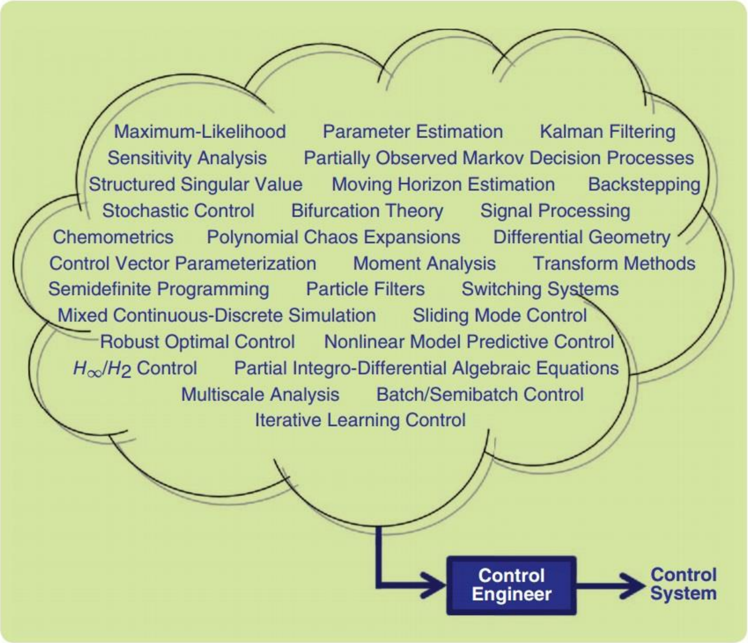
Working assumptions

Control technique

Estimation technique

Depending on the control goals and prior info, different assumptions, estimation and control techniques are adopted.

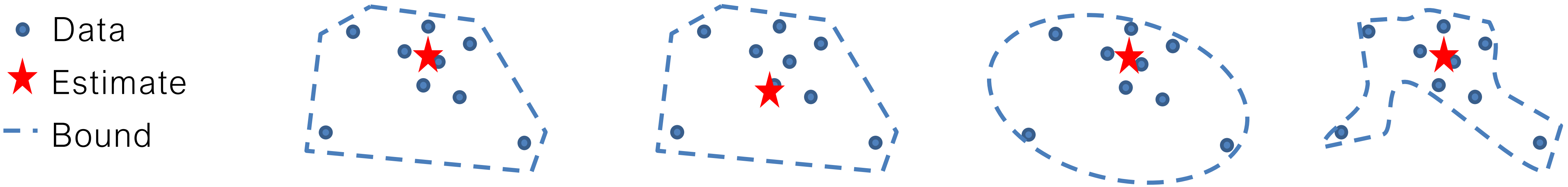
Here we consider the goal, among the other possible ones, of <u>robust constraint satisfaction</u>.



Source: [5]

Bounding the uncertainty

A prerequisite to robustly enforce constraints is to be able to compute bounds on the model uncertainty. Different estimation methods can lead to different uncertainty descriptions and "sizes".



- How can we measure uncertainty?
- How can we rank different estimation techniques?
- How can we evaluate whether an uncertainty bound is the smallest possible?
- How can we ensure that our uncertainty description is "valid"?

Plus: we want uncertainty descriptions that can be efficiently employed in the control computation

A theoretical framework

We want to have:

- Guaranteed uncertainty bounds
- A measure of optimality of the derived estimate

We need to deal with:

- Partial information
- Limited computation

nature

Explore Content ⌄ Journal Information ⌄ Publish With Us ⌄ Subscribe

nature › review articles › article

Published: 02 July 1987

Information-based complexity

Edward W. Packel & J. F. Traub

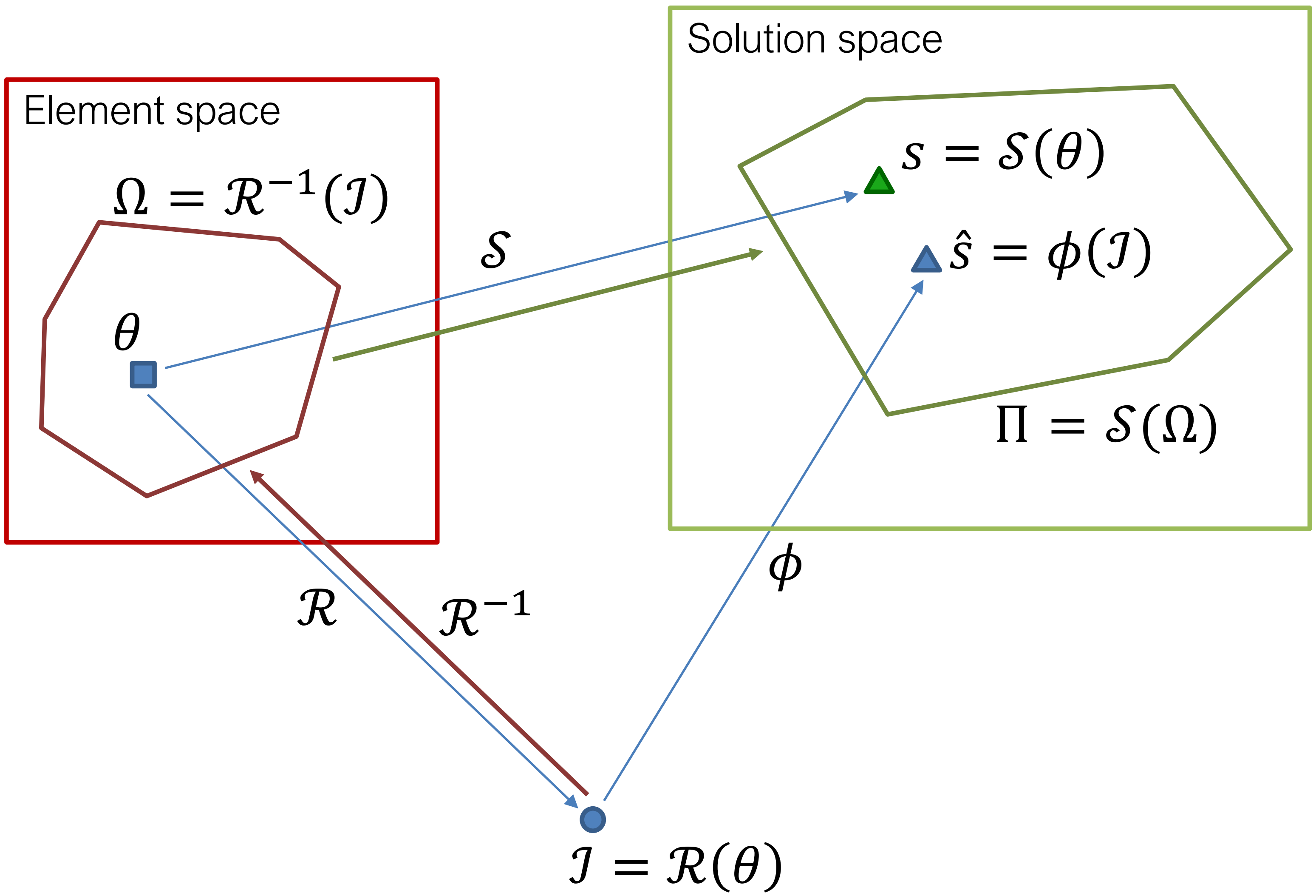
Nature **328**, 29–33(1987) | Cite this article

129 Accesses | **16** Citations | **0** Altmetric | Metrics

Abstract

Information-based complexity seeks to develop general results about the intrinsic difficulty of solving problems where available information is partial or approximate and to apply these results to specific problems. This allows one to determine what is meant by an optimal algorithm in many practical situations, and offers a variety of interesting and sometimes surprising theoretical results.

The big picture of Information-Based Complexity



θ : problem element

$\mathcal{S}$: solution operator

s : solution element

$\mathcal{R}$: information operator

$\mathcal{I}$: information

ϕ : estimation algorithm

$\hat{s}$: estimated solution

Ω : feasible element set

Π : solution uncertainty set

Set Membership (SM) estimation

SM estimation methods are rooted in Information-Based Complexity (see, e.g., [6]).

- **Problem elements** → model parameters, functions, variables,…
- **Solutions** → predictions, derived quantities, problem elements,…
- **Information** → prior assumptions on the elements and disturbance signals; data
- **Estimation algorithm** → mechanism employed to obtain an estimate
- **Feasible element set** → feasible (or unfalsified) parameter set, feasible function set, feasible variable set,…
- **Solution uncertainty set** → feasible predictions set, feasible parameter set,…

SM estimation: key ideas

1. Invert the information operator to compute the unfalsified element set Ω;
2. Apply the solution operator to all elements in Ω to compute the solution uncertainty set Π;
3. Employ the set Π to:
   - Compute uncertainty bounds on the solution values;
   - Compute uncertainty associated with a given estimate;
   - Evaluate the optimality of a given estimation algorithm;
   - Derive optimal estimation algorithms.

Optimality is typically measured in terms of worst-case error bound.

A 1D example: 1-step-ahead state prediction (order-1 system)

Bounded disturbance: $-0.5 \leq w \leq 1$

$$x(k+1) = \theta_0 x(k) + w(k)$$

Data: $\tilde{X} = \begin{bmatrix} 0.50 \\ 0.97 \\ 1.34 \end{bmatrix}$, $\tilde{X}^+ = \begin{bmatrix} 0.97 \\ 1.34 \\ 0.36 \end{bmatrix}$

Estimation algorithm n.1: ϕ_{LS} (least squares)

- Compute an estimate $\hat{\theta}_{LS} \approx \theta_0$ via least squares (LS) using the data;
- For given $x(k)$, estimate the solution as:

$$\hat{x}(k+1) = \hat{\theta}_{LS} x(k)$$

Result e.g. for $x(k) = 2$: $\hat{x}_{LS}^+ = 1.51$ ($\hat{\theta}_{LS} = 0.759$) → Uncertainty? Optimality?

A 1D example – SM estimation

Information operator

Bounded disturbance: $-0.5 \leq w \leq 1$

Data: $\tilde{X} = \begin{bmatrix} 0.50 \\ 0.97 \\ 1.34 \end{bmatrix}$, $\tilde{X}^+ = \begin{bmatrix} 0.97 \\ 1.34 \\ 0.36 \end{bmatrix}$

Solution Problem element

$$x(k+1) = \theta_0 x(k) + w(k)$$

Solution operator

Feasible parameter set: $\Omega = \{\theta | -0.5 \leq \tilde{x}_i^+ - \tilde{x}_i\theta \leq 1,\ i = 1,2,3\}$

That is: $\Omega = \left\{\theta \left| \begin{array}{c} -1.47 \leq -0.50\theta \leq 0.03 \\ -1.84 \leq -0.97\theta \leq -0.34 \\ -0.86 \leq -1.34\theta \leq 0.64 \end{array} \right.\right\} = \left\{\theta \left| \begin{array}{c} -0.06 \leq \theta \leq 2.94 \\ 0.35 \leq \theta \leq 1.89 \\ -0.47 \leq \theta \leq 0.64 \end{array} \right.\right\}$

A 1D example – SM estimation (2)

Feasible parameter set: $\Omega = [0.35, 0.64]$

Feasible solution set for a generic $x(k)$:

$$\Pi\big(x(k)\big) = \left[\min_{\theta\in\Omega} x(k)\theta - 0.5, \max_{\theta\in\Omega} x(k)\theta + 1\right]$$

Specifically, for $x(k) = 2$ we have $\Pi = [0.2, 2.28]$

Uncertainty bounds on the solution values

Uncertainty associated with the LS estimate $\hat{x}_{LS}^{+}$: $[-1.31, 0.76]$.

(Worst-case) error of LS estimate: $e(\phi_{LS}) = \max_{s\in\Pi} \|s - \hat{x}_{LS}^{+}\| = 1.31$

A 1D example – state of the art

Can we find a better estimation algorithm?

Estimation algorithm n.2: $\phi_{LS,c}$ (least squares + center of disturbance)

- Consider $x(k+1) = \theta_0 x(k) + \bar{w} + \Delta w(k)$, where $\bar{w} = \frac{1}{2}(\max w + \min w) = 0.25, |\Delta w| \leq 0.75$
- Correct the data as $\tilde{X}_c^+ = \tilde{X}^+ - \bar{w}$.
- Compute an estimate $\hat{\theta}_{LS,c} \approx \theta_0$ via least squares (LS) using the corrected data $\tilde{X}_c^+$, $\tilde{X}$;
- For given $x(k)$, estimate the solution as:

$$\hat{x}(k+1) = \hat{\theta}_{LS,c} x(k) + \bar{w}$$

Result for $x(k) = 2$: $\hat{x}_{LS,c}^+ = 1.3$ ($\hat{\theta}_{LS,c} = 0.52$), with uncertainty bounds $[-1.1, 0.98]$

(Worst-case) error: $e\left(\phi_{LS,c}\right) = \max_{s \in \Pi} \left\| s - \hat{x}_{LS,c}^+ \right\| = 1.1$

A 1D example – SM estimation (3)

Can we find a better estimation algorithm?

Estimation algorithm n.3: ϕ_c (central estimate)

$$\phi_c\big(x(k)\big) = \frac{\max\limits_{s\in\Pi(x(k))} s + \min\limits_{s\in\Pi(x(k))} s}{2}$$

Specifically, for $x(k) = 2$ we have $\hat{x}_c^+ = 1.24$, with uncertainty ± 1.04

(Worst-case) error: $e(\phi_c) = \max\limits_{x^+\in\Pi} \|s - \hat{x}_c^+\| = 1.04$

A 1D example – SM estimation (4)

Note that, in this case, the central estimate corresponds to the following estimation algorithm (show it):

- Compute an estimate $\hat{\theta}_c \approx \theta_0$ as: $\hat{\theta}_c = \frac{\max_{\theta \in \Omega} \theta + \min_{\theta \in \Omega} \theta}{2}$;
- For given $x(k)$, estimate the solution as:

$$\hat{x}(k+1) = \hat{\theta}_c x(k) + \bar{w}$$

i.e., to take the central parameter estimate inside Ω and consider the disturbance correction in the estimation.

In the numerical example we have $\hat{\theta}_c = 0.49$.

A 1D example – SM estimation (5)

In summary:

Feasible parameter set: $\Omega = [0.35,0.64]$ Feasible solution set for $x(k) = 2$: $\Pi = [0.2,2.28]$			
Estimation algorithm	Estimate	Worst-case error	Parameter estimate
ϕ_{LS}	$\hat{x}_{LS}^{+} = 1.51$	$e(\phi_{LS}) = 1.31$	$\hat{\theta}_{LS} = 0.759$
$\phi_{LS,c}$	$\hat{x}_{LS}^{+} = 1.30$	$e(\phi_{LS}) = 1.10$	$\hat{\theta}_{LS,c} = 0.52$
ϕ_c	$\hat{x}_{c}^{+} = 1.24$	$e(\phi_c) = 1.04$	$\hat{\theta}_c = 0.49$

Some notes:

- $\hat{\theta}_{LS} \notin \Omega$, yet we can compute the uncertainty bounds for any $x(k)$
- Guaranteed uncertainty bounds obtained with finite data
- No estimation algorithm can achieve smaller worst-case error than 1.04 (show it)
- The central estimate is optimal: it achieves minimum uncertainty

Measures of information

DEF Diameter of information: $D_{\mathcal{J},p}(\Pi) = \sup_{s_1,s_2 \in \Pi} \|s_1 - s_2\|_p$

(with a chosen p–norm)

DEF Radius of information: $R_{\mathcal{J},p}(\Pi) = \inf_{s_1 \in \Pi} \sup_{s_2 \in \Pi} \|s_1 - s_2\|_p$

In the example, $R_{\mathcal{J},p}(\Pi) = 1.04$, $D_{\mathcal{J},p}(\Pi) = 2.08$.

THM $R_{\mathcal{J},p}(\Pi) \leq D_{\mathcal{J},p}(\Pi) \leq 2R_{\mathcal{J},p}(\Pi)$ (prove it)

Error of an algorithm, optimal algorithm

DEF Error of an estimation algorithm ϕ:

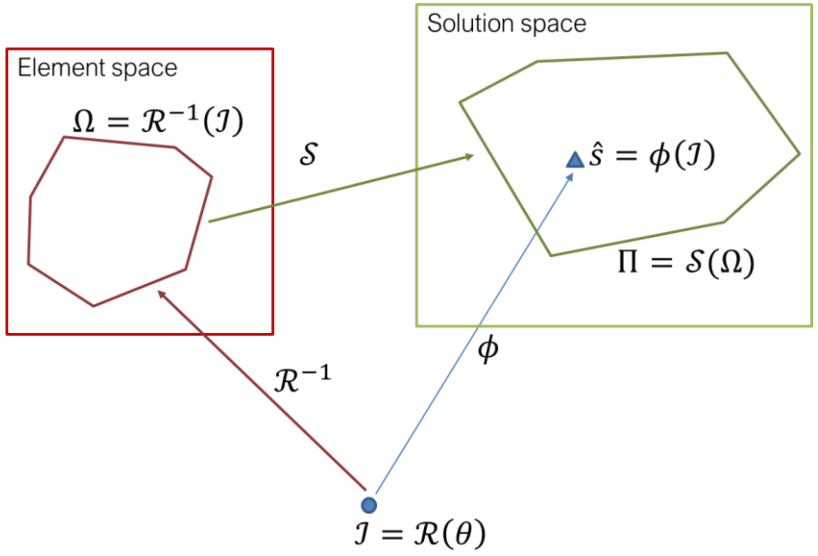
$$e_p(\phi) = \sup_{s\in\Pi} \|\hat{s} - s\|_p \text{ (where } \hat{s} = \phi(\mathcal{J}))$$

THM if Π is convex, $e_p(\phi) \geq R_{\mathcal{J},p}(\Pi), \forall\phi$ (prove it)

DEF Optimal error: $e_p^o = \inf_{\phi} e_p(\phi)$

DEF Optimal algorithm: $\phi^* | e_p(\phi^*) = e_p^o$

Interpolatory and central algorithms



DEF Interpolatory algorithm $\phi^I |\ \phi^I(\mathcal{J}) \in \Pi$

DEF Center of Π: solution element s_c such that if $s_c + g \in \Pi$ then also $s_c - g \in \Pi$, for any g in the solution space (in the example, $s = 1.24$)

DEF Central algorithm: $\phi_c | \phi_c(\mathcal{J})$ is a center of Π

Optimality of interpolatory and central algorithms

THM for any interpolatory algorithm ϕ^I, it holds: $R_{\mathcal{I},p}(\Pi) \leq e_p(\phi^I) \leq 2R_{\mathcal{I},p}(\Pi)$ (prove it)

THM Any central algorithm is optimal: $e_p(\phi_c) = e_p^o$ (prove it)

Note: central algorithms may be difficult to obtain in general. Typically, we can be happy with an interpolatory algorithm, which is "2-suboptimal".

Further aspects

- Bound-exploring property of data
- Assumption (in)validation and conservativeness
- Measurement noise
- Generalization to higher dimensions (affine parametrizations)
- Unbounded unfalsified element set

Bound-exploring property of data

Conceptually similar to persistence of excitation in stochastic settings. If the input is bound exploring, we can obtain minimal uncertainty.

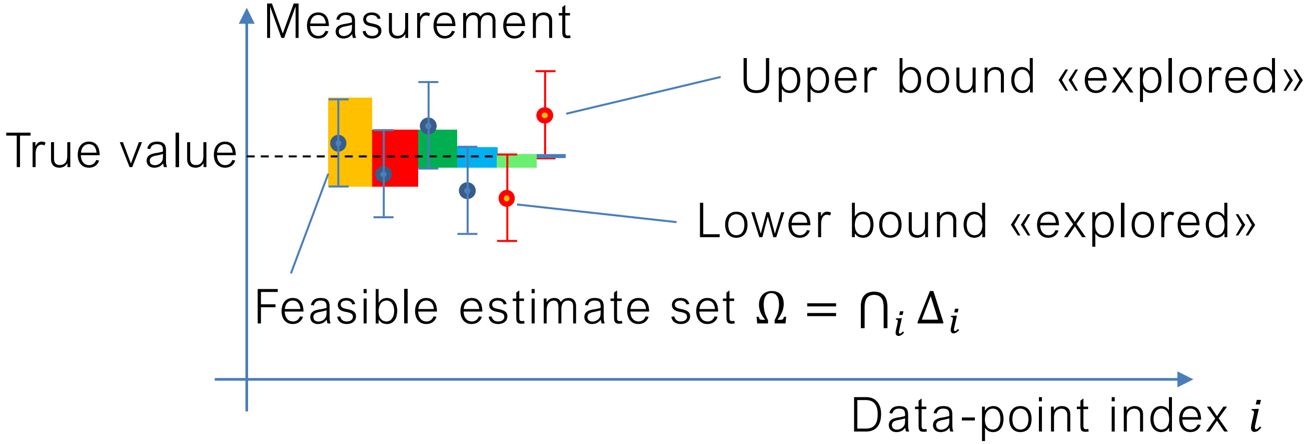
E.g., in our example assume to have:

$$\tilde{x}(0) = 0.5$$

$$\tilde{x}(1) = 1.25 \quad \rightarrow \quad 0.5 \leq \theta \leq 3.5 \qquad (-0.5 \leq 1.25 - 0.5\theta \leq 1)$$

$$\tilde{x}(2) = 0.125 \quad \rightarrow -0.7 \leq \theta \leq 0.5 \qquad (-0.5 \leq 0.125 - 1.25\theta \leq 1)$$

$\Rightarrow \Omega = \{0.5\}$ No uncertainty with just 2 samples!



Assumption (in)validation and conservativeness

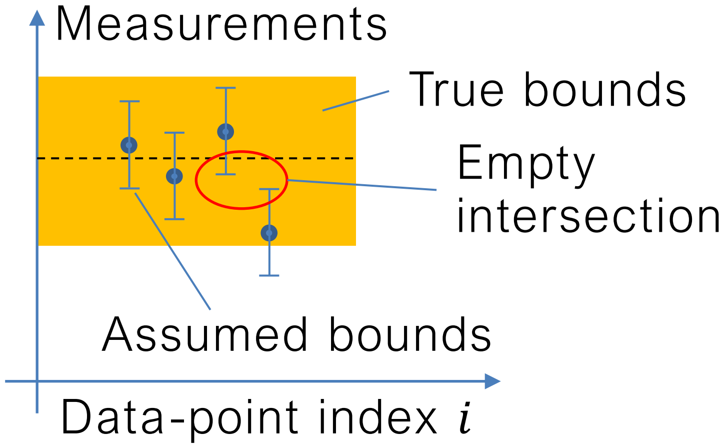
In practice, exact bounds on the inputs are seldom available.

Two cases can occur:

I. If bounds are under-estimated: **possible assumption invalidation**

In this case, it may happen that $\Omega = \emptyset$

- Can happen, e.g., with outliers
- Can be exploited for fault-detection
- Can be exploited to estimate the bounds



II. If bounds are over-estimated: **conservativeness**

In this case, Ω is never empty, it over-approximates the tightest feasible parameter set (which would be obtained with the exact dist. bounds)

Assumption (in)validation and conservativeness - remark

NOTE: prior assumptions can never be "validated" with finite data and computation!!

We can only verify that they are not falsified by the available information, hence the name "unfalsified element set".

Measurement noise

Measurement noise can be considered as well, for example:

$$x(k+1) = \theta_0 x(k) + w(k)$$

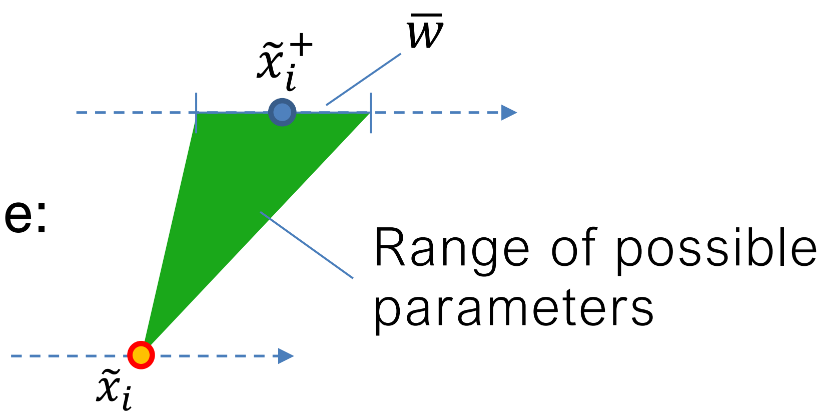
$$\tilde{x}(k) = x(k) + v(k)$$

Bounded disturbance: $|w| \leq \bar{w}$

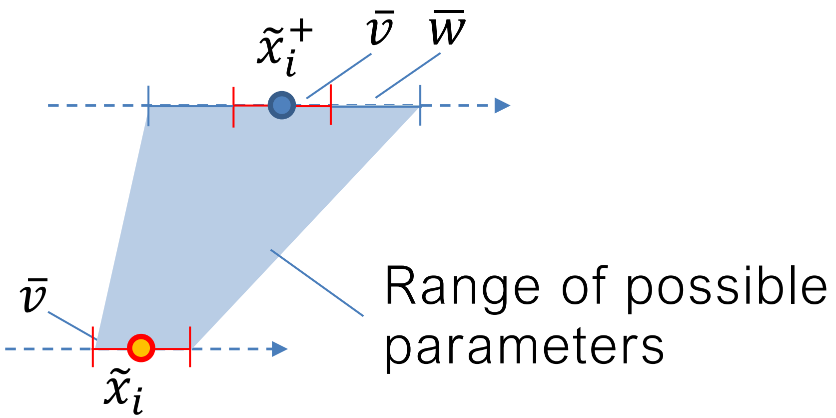
Bounded measurement noise: $|v| \leq \bar{v}$

Data pairs: $(\tilde{x}_i^+, \tilde{x}_i), i = 1, \dots, N$

No noise:



With noise:



Feasible parameter set: $\Omega = \cup_i \Delta_i$,

where $\Delta_i = \{\theta \,|\, \exists y_i : |y_i - \tilde{x}_i| \leq \bar{v} \wedge |\tilde{x}_i^+ - \theta y_i| \leq (\bar{w} + \bar{v})\}$

Measurement noise (continued)

The feasible solution set for given $\tilde{x}(k)$ can be computed as:

$$\Pi = \left[\underline{s} - \underline{w}\ , \overline{s} + \overline{w}\right],$$

where:

$$\underline{s} = \min_{\substack{\theta, y(k), y_i \\ i=1,\dots N}} \theta y(k) \qquad \overline{s} = \max_{\substack{\theta, y(k), y_i \\ i=1,\dots N}} \theta y(k)$$

$$\text{subject to} \qquad\qquad \text{subject to}$$

$$|\tilde{x}_i^+ - \theta y_i| \le (\overline{w} + \overline{v}) \qquad |\tilde{x}_i^+ - \theta y_i| \le (\overline{w} + \overline{v})$$

$$|y_i - \tilde{x}_i| \le \overline{v} \qquad |y_i - \tilde{x}_i| \le \overline{v}$$

$$|y(k) - \tilde{x}(k)| \le \overline{v} \qquad |y(k) - \tilde{x}(k)| \le \overline{v}$$

And the central estimate is $\hat{x}_c^+ = 0.5(\underline{s} + \overline{s})$ with radius of information $0.5(\underline{s} - \overline{s})$.

As a matter of fact, in most learning-based MPC approaches with SM estimation measurement noise is neglected, or the model parametrization is such that no bilinear term appears (e.g., FIR models with no input meas. noise).

Generalization to higher dimensions (affine parametrization)

All concepts generalize to $\theta \in \mathbb{R}^{n_\theta}$ with $n_\theta > 1$. Models with affine parametrization will be the most relevant for us. Some examples:

1. ARX model of order m: $y_j(k) = \varphi_j(k)^T \theta_j + w(k)\ ,\quad j = 1, \dots, n_y$

   where $\varphi_j(k) = \left[y_j(k-1), \dots, y_j(k-m), u_1(k-1), \dots, u_1(k-m), u_2(k-1), \dots, u_{n_u}(k-m)\right]$

   and $\theta_j \in \mathbb{R}^{(1+n_u)m}$ (system with n_y outputs and n_u inputs)

2. State-space model: $x(k+1) = \left(A + \sum_{i=1}^{n_\theta} A_i \theta_i\right) x(k) + \left(B + \sum_{i=1}^{n_\theta} B_i \theta_i\right) u(k) + w(k)$

   equivalent to $x(k+1) = Ax(k) + Bu(k) + \underbrace{\left[A_1 x(k) + B_1 u(k) \middle| \dots \middle| A_{n_\theta} x(k) + B_{n_\theta} u(k)\right]}_{M(x(k),u(k)) \in \mathbb{R}^{n \times n_\theta}} \theta + w(k)$

3. FIR model of order m: like ARX but without autoregressive part

Affine parametrization with polytopic disturbance bounds

When the model is affine in the parameters and the disturbance bounds are described by a convex polytope, the unfalsified parameter set Ω is a polytope.

- Let $y = M(\varphi)\theta + w$, where $M(\varphi) \in \mathbb{R}^{n_y \times n_\theta}$, $w\colon H_w w \leq h_w$, and $\varphi \in \mathbb{R}^{n_\varphi}$ is a measured regressor.
- Consider the i-th data pair $(\tilde{y}_i, \tilde{\varphi}_i)$. Compute the set Δ_i as:

$$\Delta_i = \{\theta \mid H_w \tilde{y}_i - H_w M(\tilde{\varphi}_i)\theta \leq h_w\} = \left\{\theta \,\middle|\, \underbrace{-H_w M(\tilde{\varphi}_i)}_{H_{\Delta_i}}\theta \leq \underbrace{h_w - H_w \tilde{y}_i}_{h_{\Delta_i}}\right\}$$

- With N data pairs, the set Ω is given by the intersection of the resulting convex polytopes Δ_i:

$$\Omega = \left\{\theta \,\middle|\, \begin{matrix} H_{\Delta_1}\theta \leq h_{\Delta_1} \\ \vdots \\ H_{\Delta_N}\theta \leq h_{\Delta_N} \end{matrix} \right\}$$

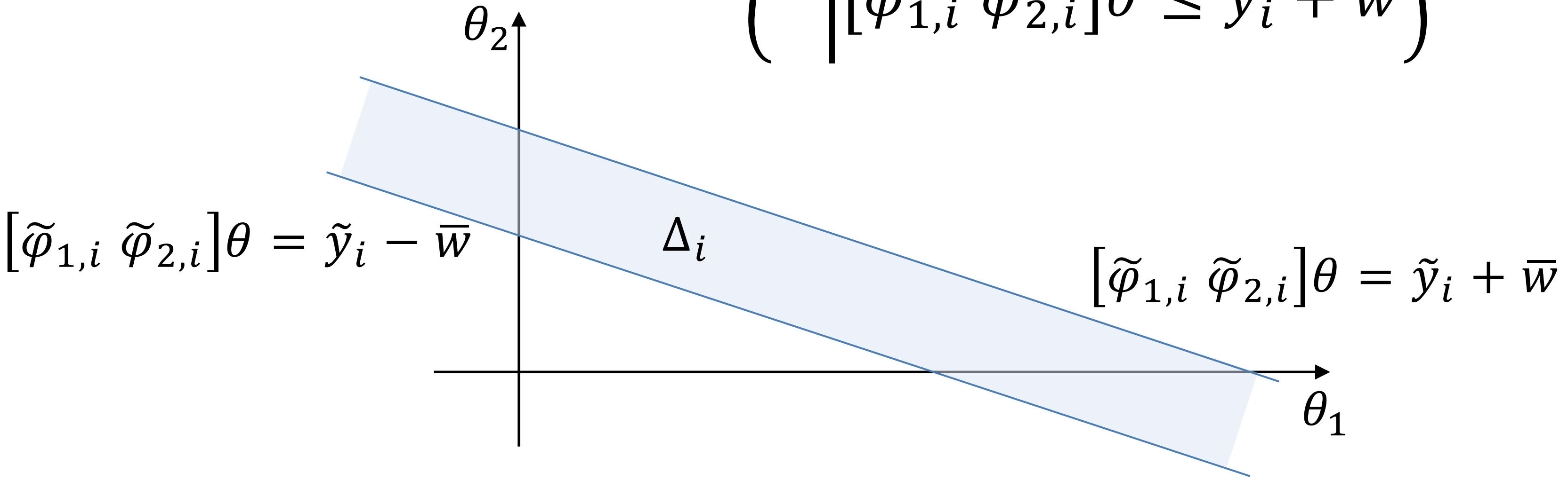
Polytopic unfalsified parameter set - example

Consider the model:

$$y = [\varphi_1 \ \varphi_2]\theta + w, \text{ where } \ w\colon \begin{bmatrix} 1 \\ -1 \end{bmatrix} w \leq \begin{bmatrix} \bar{w} \\ \bar{w} \end{bmatrix} \text{ (i.e. } |w| \leq \bar{w})$$

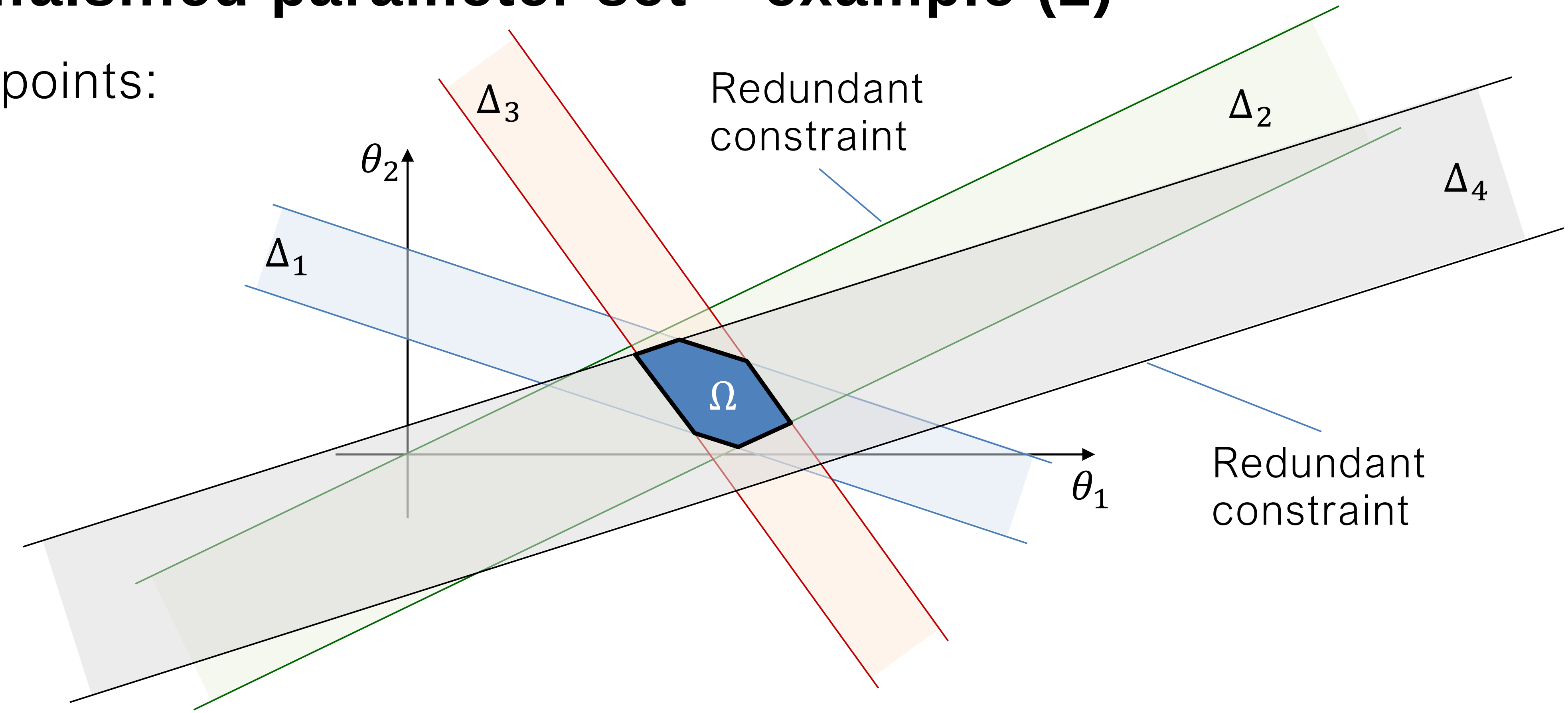
Each data pair results in a «slab», i.e. the intersection of two half-spaces with parallel separating hyper-planes (whose distance in this case is $2\bar{w}$):

$$\Delta_i = \left\{ \theta \ \middle| \begin{matrix} [\tilde{\varphi}_{1,i} \ \tilde{\varphi}_{2,i}]\theta \geq \tilde{y}_i - \bar{w} \\ [\tilde{\varphi}_{1,i} \ \tilde{\varphi}_{2,i}]\theta \leq \tilde{y}_i + \bar{w} \end{matrix} \right\}$$



Polytopic unfalsified parameter set – example (2)

With more data points:



- The unfalsified parameter set generally shrinks (up to the limit of point/minimum/empty, see discussion above on bound-exploring, conservativeness, and empty set)
- The unfalsified parameter set never grows
- We may have redundant constraints

Affine unfalsified parameter set with affine solution operator

When the unfalsified parameter set Ω is a polytope and the solution operator is affine, then the solution uncertainty set is also a polytope.

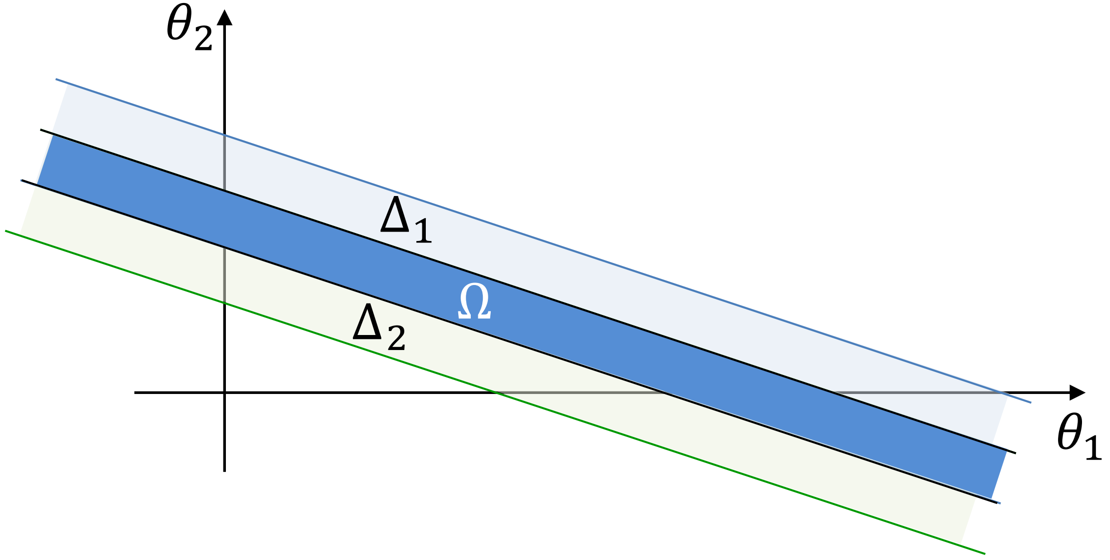
$$\Omega = \{\theta : H_\Omega \theta \leq h_\Omega\}\,,\ \mathcal{S}(\theta) = C\theta + d$$

$$\Rightarrow \Pi = \mathcal{S}(\Omega) = \{s = C\theta + d : H_\Omega \theta \leq h_\Omega\}$$

Unbounded unfalsified element set

It can happen that Ω is unbounded.

- This can be easily checked by linear programming, e.g. $\max_{\theta \in \Omega} \|\theta\|$
- This is a sign that the available prior assumption and/or data are not informative enough



Final considerations

SM approaches address the following questions:

- How can we measure uncertainty?
- How can we rank different estimation techniques?
- How can we evaluate whether an uncertainty model is the smallest possible?

We have seen that in case of affine parametrization we have good answers.

Plus: uncertainty description is polytopic, making it suitable for robust control.

The following question is not well-posed:

- How can we ensure that our uncertainty description is “valid”?

References

[1] M. Gevers, "Identification for Control: From the Early Achievements to the Revival of Experiment Design", *European Journal of Control*, Volume 11, Issues 4–5, 2005

[2] B. Wittenmark, "Adaptive dual control methods: An overview," IFAC Proceedings Volumes, vol. 28, no. 13, pp. 67–72, 5th IFAC Symposium on Adaptive Systems in Control and Signal Processing, Budapest, Hungary, 14-16 June, 1995

[3] G. Tao, "Multivariable adaptive control: A survey", *Automatica*, Volume 50, Issue 11, pp. 2737-2764, 2014

[4] D. A. Bristow, M. Tharayil and A. G. Alleyne, "A survey of iterative learning control", *IEEE Control Systems Magazine*, vol. 26, no. 3, pp. 96-114, June 2006

[5] R. D. Braatz, "Do You Have a Control Tool or a Control Toolbox?", *IEEE Control Systems Magazine* 34(3), pp. 6-7, 2014

[6] M. Milanese, A. Vicino, "Optimal Estimation Theory for Dynamic Systems with Set Membership Uncertainty: An Overview". In: Milanese M., Norton J., Piet-Lahanier H., Walter É. (eds) Bounding Approaches to System Identification. Springer, Boston, MA, 1996