

A Novel Linear Recursive Estimator Based on the Frisch Scheme

Stefano Massaroli¹, Renato Miyagusuku², Federico Califano³,
Angela Faragasso¹, Atsushi Yamashita¹ and Hajime Asama¹

¹Department of Precision Engineering, **The University of Tokyo**

²Department of Mechanical and Intelligent Engineering, **Utsunomiya University**

³Department of Robotics and Mechatronics, **University of Twente**

12th Asian Control Conference
2019-06-11

Search for **connections** between observations (“**Laws of Nature**”)

⇒ At the basis of the development of **scientific knowledge**.

e.g. *Babylonian astronomical diaries* (≈ 747 BC)

Given:	Variables	Observations
	x_1, x_2, \dots, x_n	$x_{i1}, x_{i2}, \dots, x_{in}$

Search for a relation describing the process generating the data:

$$f(x_1, x_2, \dots, x_n) = 0$$

satisfied by **every** set of observations, i.e.

$$\forall i \quad f(x_{i1}, x_{i2}, \dots, x_{in}) = 0$$

Problem: In ALL practical situations, data will NEVER satisfy the relation.

⇒ Observations are affected by noise.

Search for **connections** between observations (“**Laws of Nature**”)

⇒ At the basis of the development of **scientific knowledge**.

e.g. *Babylonian astronomical diaries* (≈ 747 BC)

Given:	Variables	Observations
	x_1, x_2, \dots, x_n	$x_{i1}, x_{i2}, \dots, x_{in}$

Search for a relation describing the process generating the data:

$$f(x_1, x_2, \dots, x_n) = 0$$

satisfied by **every** set of observations, i.e.

$$\forall i \quad f(x_{i1}, x_{i2}, \dots, x_{in}) = 0$$

Problem: In ALL practical situations, data will NEVER satisfy the relation.
 ⇒ Observations are affected by noise.

Search for **connections** between observations (“**Laws of Nature**”)

⇒ At the basis of the development of **scientific knowledge**.

e.g. *Babylonian astronomical diaries* (≈ 747 BC)

Given:	Variables	Observations
	x_1, x_2, \dots, x_n	$x_{i1}, x_{i2}, \dots, x_{in}$

Search for a relation describing the process generating the data:

$$f(x_1, x_2, \dots, x_n) = 0$$

satisfied by **every** set of observations, i.e.

$$\forall i \quad f(x_{i1}, x_{i2}, \dots, x_{in}) = 0$$

Problem: In ALL practical situations, data will NEVER satisfy the relation.

⇒ **Observations are affected by noise.**

Problem setting:

Extract **linear relations** from data affected by **additive noise**.

Approach:

We introduce a **novel estimator**: a recursive version of the *Frisch Scheme*.

Contribution:

- Reduced size of the solutions space w.r.t. standard methods;
- Improved computational efficiency

Originality:

Approximate the intersection of n -dimensional simplexes by means of bounding boxes.

Problem setting:

Extract **linear relations** from data affected by **additive noise**.

Approach:

We introduce a **novel estimator**: a **recursive** version of the *Frisch Scheme*.

Contribution:

- Reduced size of the solutions space w.r.t. standard methods;
- Improved computational efficiency

Originality:

Approximate the intersection of n -dimensional simplexes by means of bounding boxes.

Problem setting:

Extract **linear relations** from data affected by **additive noise**.

Approach:

We introduce a **novel estimator**: a **recursive** version of the *Frisch Scheme*.

Contribution:

- **Reduced size** of the solutions space w.r.t. standard methods;
- Improved computational efficiency

Originality:

Approximate the intersection of n -dimensional simplexes by means of bounding boxes.

Problem setting:

Extract **linear** relations from data affected by **additive noise**.

Approach:

We introduce a **novel estimator**: a **recursive** version of the *Frisch Scheme*.

Contribution:

- Reduced size of the solutions space w.r.t. standard methods;
- Improved computational efficiency

Originality:

Approximate the intersection of n -dimensional simplexes by means of bounding boxes.

Linear Relation:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

with

$$x_i = \hat{x}_i + \tilde{x}_i$$

Observation matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Parameters vector

$$\mathbf{A} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n$$

It holds:

$$\mathbf{X} = \hat{\mathbf{X}} + \tilde{\mathbf{X}}$$

$$\boxed{\hat{\mathbf{X}}\mathbf{A} = 0}$$

Linear Relation:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

with

Practical Examples:

- n -DOF robot: $\boldsymbol{\tau} = \mathbf{Y}(\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}})\boldsymbol{\theta}$
- Linear ODEs: $L\ddot{V} + R\dot{V} - \frac{1}{C}(V + u) = 0$
- Autoregressive models
- dynamical systems linear w.r.t. some parameters

 $\mathbf{X} =$

$$\begin{pmatrix} x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \quad \bigg| \quad \begin{pmatrix} \alpha_n \end{pmatrix}$$

It holds:

$$\mathbf{X} = \hat{\mathbf{X}} + \tilde{\mathbf{X}}$$

$$\hat{\mathbf{X}}\mathbf{A} = 0$$

We define the *sample covariance matrix* of the data:

$$\Sigma = \frac{1}{m} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$$

Assumption: Noise and noiseless samples are orthogonal:

$$\sum_{t=1}^m \hat{x}_{it} \tilde{x}_{jt} = 0 \quad \forall i, j$$

We have

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \quad , \quad \Sigma > 0, \quad \tilde{\Sigma} \geq 0$$

$$\hat{\Sigma} \mathbf{A} = 0$$

$$\hat{\Sigma} \geq 0 \quad \text{and} \quad \det(\hat{\Sigma}) = 0$$

We define the *sample covariance matrix* of the data:

$$\Sigma = \frac{1}{m} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$$

Assumption: Noise and noiseless samples are orthogonal:

$$\sum_{t=1}^m \hat{x}_{it} \tilde{x}_{jt} = 0 \quad \forall i, j$$

We have

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \quad , \quad \Sigma > 0, \quad \tilde{\Sigma} \geq 0$$

$$\hat{\Sigma} \mathbf{A} = 0$$

$$\hat{\Sigma} \geq 0 \quad \text{and} \quad \det(\hat{\Sigma}) = 0$$

We define the *sample covariance matrix* of the data:

$$\Sigma = \frac{1}{m} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$$

Assumption: Noise and noiseless samples are orthogonal:

$$\sum_{t=1}^m \hat{x}_{it} \tilde{x}_{jt} = 0 \quad \forall i, j$$

We have

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \quad , \quad \Sigma \succ 0, \quad \tilde{\Sigma} \succeq 0$$

$$\hat{\Sigma} \mathbf{A} = 0$$

$$\hat{\Sigma} \succeq 0 \quad \text{and} \quad \det(\hat{\Sigma}) = 0$$

Problem: Estimation Scheme [Kalmann, 1982]

Given a sample covariance matrix of noisy observations, Σ , determine positive definite or semidefinite noise covariance matrices $\tilde{\Sigma}$ such that

$$\hat{\Sigma} = \Sigma - \tilde{\Sigma} \succeq 0 \quad \text{and} \quad \det(\hat{\Sigma}) = 0$$

⇒ Any basis of $\text{null}(\hat{\Sigma})$ will describe a set of linear relations compatible with the data

$$\mathbf{A} = \text{null}(\hat{\Sigma})$$

Frisch Scheme's Assumption: The noise variables are mutually independent:

$$\sum_{t=1}^m \tilde{x}_{it} \tilde{x}_{jt} = 0 \quad \forall i \neq j \quad \Rightarrow \quad \tilde{\Sigma} \text{ is diagonal}$$

Frisch Scheme Solution:

Every positive definite or semidefinite diagonal matrix $\tilde{\Sigma}$

$$\tilde{\Sigma} = \text{diag}(\tilde{\sigma}_1^2 \quad \tilde{\sigma}_2^2 \quad \cdots \quad \tilde{\sigma}_n^2)$$

such that

$$\hat{\Sigma} = \Sigma - \tilde{\Sigma} \geq 0 \quad \text{and} \quad \det \hat{\Sigma} = 0$$

is a *solution* of the Frisch scheme.

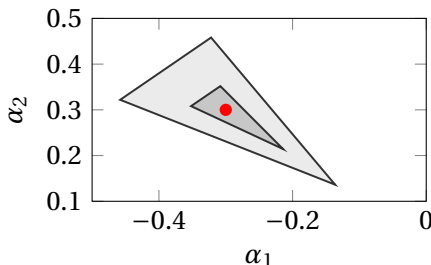
In general, the Frisch scheme has infinite solutions.

Define

$$\text{Maxcor}_F(\Sigma) = \max_{\tilde{\Sigma} \in \mathcal{D}} \{\dim [\text{null}(\Sigma - \tilde{\Sigma})]\}$$

Theorem [Kalman, 1982]

If $\text{Maxcor}_F(\Sigma) = 1$, the coefficients $\alpha_1, \dots, \alpha_n$ of all linear relations compatible with the Frisch scheme lie (by normalizing one of the coefficients to 1) inside the simplex whose vertices are defined by the *n least squares solutions*.



Advantages

- post-identification degree of freedom in the choice of a single solution
 - ⇒ solve physical feasibility problems
 - ⇒ seek “optimal” solution
- closed form computation of the *simplex of solutions*

Disadvantages

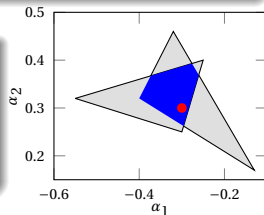
- To obtain a simplex of a useful size it is needed a *well-conditioned* experiment
 - ⇒ not suitable for diagnostics or online estimations
- If the solutions space is too wide, the post-identification process is meaningless

Research Aim:

- Reduce the size of the solutions space without introducing any further assumptions (*priors*)
- Develop a computationally efficient algorithm, suitable for online applications (control, diagnostics, etc.).

Recursive Frisch Scheme: Our Intuition

- If we compute the simplices corresponding to two different covariance matrices, **the true parameter must lie in both of them**, i.e. in their **intersection**.



Proposed Approach:

- Divide the data in several batches and compute the respective simplexes;
- Intersect those simplexes to obtain a smaller solution space.

Remark:

By **successively intersecting** simplexes computed from different data sets it is possible to significantly **reduce the size** of the solutions space.

↓size → ↑accuracy

Main Issues:

- ⇒ No computationally efficient algorithms compute the intersection of convex objects only from the knowledge of their vertices;
- ⇒ The number of vertices of the intersection may increase with the iterations.

Remark:

How can we approximate the intersection?

By **successively** adding more data sets it is possible to significantly **reduce the size** of the solutions space.

Main Issues:

- ⇒ No computationally efficient algorithms compute the intersection of convex objects only from the knowledge of their vertices;
- ⇒ The number of vertices of the intersection may increase with the iterations.

Remark:

By successively adding data sets it is possible to significantly reduce the size of the solutions space.

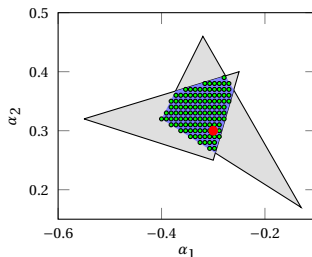
How can we approximate the intersection?

Current Method: Use particles!

Main Issues:

- ⇒ No computational objects only for
- ⇒ The number

of convex
iterations.



[Massaroli, et al., 2018]

Remark:

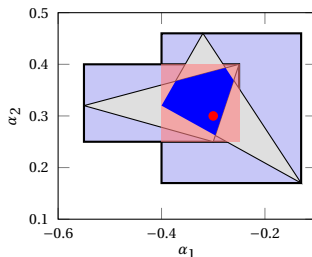
By successively adding more data sets it is possible to significantly reduce the size of the solutions space.

How can we approximate the intersection?

Proposed Method: Use bounding boxes

Main Issues:

- ⇒ No computation of convex objects only for
- ⇒ The number of iterations.



⇒ **Bounding-box recursive Frisch scheme**

Geometry of simplexes and bounding boxes

Simplex Matrix:

$$\mathbf{S}^p = (v_1, v_2, \dots, v_{p+1}) \in \mathbb{R}^{p \times p+1}$$

$$\text{Simplex: } \mathfrak{S}(\mathbf{S}^p) = \text{conv}(\mathbf{S}^p)$$

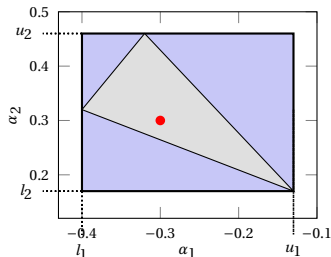
Simplex Bounds:

$$l(\mathbf{S}^p) = (l_1, \dots, l_p), \quad u(\mathbf{S}^p) = (u_1, \dots, u_p)$$

$$\text{with } l_i = \min_j (\mathbf{S}_{ij}^p) \text{ and } u_i = \max_j (\mathbf{S}_{ij}^p)$$

Simplex Bounding Box:

$$\mathfrak{B}(\mathbf{S}^p) = b_1 \times b_2 \times \dots \times b_p = \prod_{i=1}^p b_i, \quad b_i = [l_i, u_i]$$



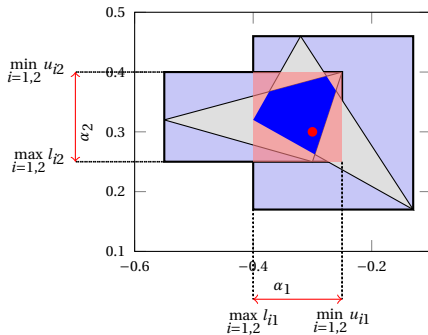
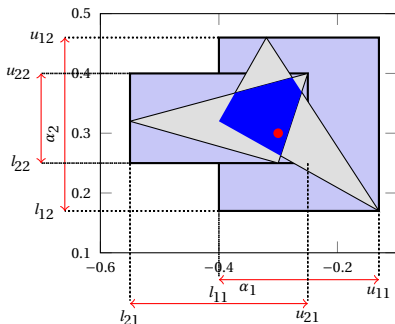
Proposition:

Given N simplex matrices $\mathbf{S}_1^n, \dots, \mathbf{S}_N^n$ it holds

$$\bigcap_{i=1}^N \mathfrak{S}(\mathbf{S}_i^n) \subseteq \bigcap_{i=1}^N \mathfrak{B}(\mathbf{S}_i^n)$$

Proposition:

$$\bigcap_{i=1}^N \mathfrak{B}(\mathbf{S}_i^n) = \prod_{j=1}^n [\max_i l_{ij}, \min_i u_{ij}]$$



BBRF Iteration

On-line estimation perspective: After the first m measurements, let's keep observing the system.

The state of the system at the time t_k is: $\mathbf{x}(t_k) = (x_1(t_k) \quad x_2(t_k) \quad \cdots \quad x_n(t_k))$

observation matrix	$\mathbf{X}(t_k) = (\mathbf{x}(t_{k-m})^\top \quad \mathbf{x}(t_{k-m+1})^\top \quad \cdots \quad \mathbf{x}(t_k)^\top)^\top \in \mathbb{R}^{m \times n}$
covariance matrix	$\Sigma(t_k) = \frac{\mathbf{X}(t_k)^\top \mathbf{X}(t_k)}{m} \in \mathbb{R}^{n \times n}$
simplex matrix	$\mathbf{S}^{n-1}(t_k) = (\mathbf{A}_1(t_k) \quad \mathbf{A}_2(t_k) \quad \cdots \quad \mathbf{A}_n(t_k)) \in \mathbb{R}^{n-1 \times n}$
simplex bounding box	$\mathfrak{B}(\mathbf{S}^{n-1}(t_k)) = \prod_{i=1}^{n-1} [l_i(t_k), u_i(t_k)]$

It holds:

$$l_i(t_k) \leq \alpha_i \leq u_i(t_k) \quad \forall i, k \in \mathbb{N}$$

Solution of Bounding Box Recursive Frisch Scheme:

Approximate the intersection of simplexes by intersecting their bounding boxes.

BBRF scheme:

Lower Solution Bounds

$$\gamma_i(t_k) = \max\{\gamma_i(t_{k-1}), l_i(t_k)\}$$

Upper Solution Bounds

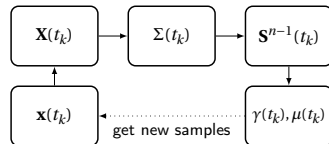
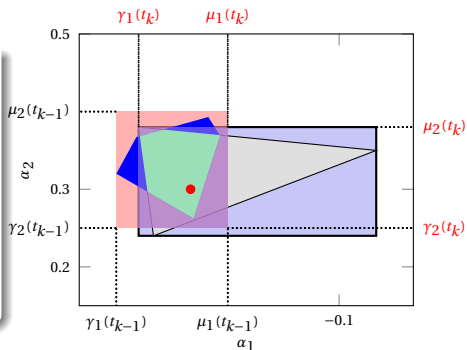
$$\mu_i(t_k) = \min\{\mu_i(t_{k-1}), u_i(t_k)\}$$

Solution Box

$$T(t_k) = \bigcap_{j=1}^k \mathfrak{B}(\mathbf{S}^{n-1}(t_j)) = \prod_{i=1}^{n-1} [\gamma_i(t_k), \mu_i(t_k)]$$

Proposition:

$$\lambda(T(t_k)) \leq \lambda(T(t_{k-1}))$$



Simulations

System:

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = 0 \Rightarrow \text{2D Simplex}$$

where

$$x_1, x_2 \sim \mathcal{N}(0, 1) \quad x_3 = -\alpha_1 x_1 - \alpha_2 x_2$$

and, thus

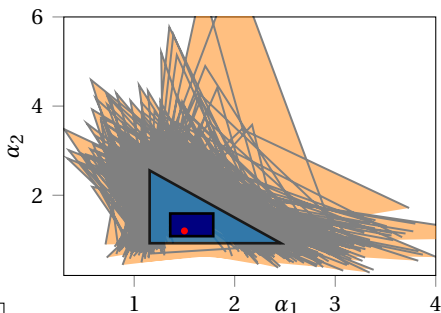
$$\mathbf{A} = \begin{pmatrix} 1.5 & 1.2 & 1 \end{pmatrix}^\top$$

- Independent noise has been added to x_1, x_2, x_3 ($\sigma_{\%} = 30$);
- Observation matrix size: $m = 20$;
- Monte Carlo simulation: 150 runs.

Results I.

“Smaller” solutions set w.r.t the simplex obtained with the whole dataset:

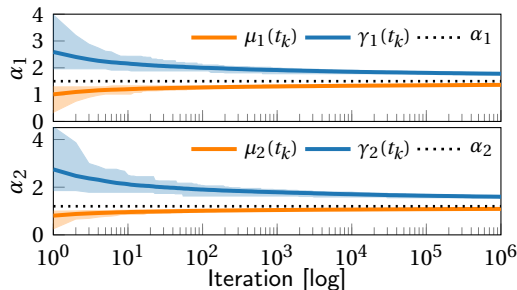
⇒ More efficient way to use data in the context of the Frisch scheme as alternative to perform a unique batch identification



Results II.

The solution bounds converge to constant values.

⇒ The uncertainty intervals at convergence are considerably smaller than the initial ones.



Results III.

Volume of solutions sets:

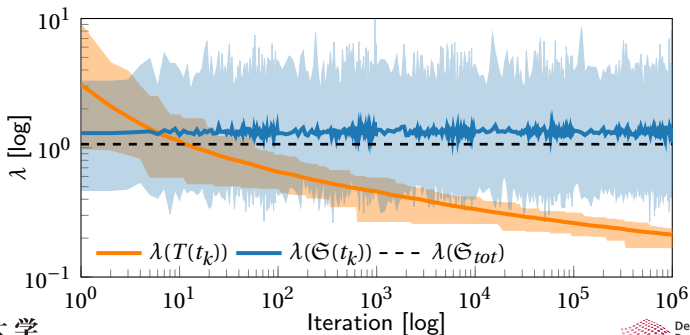
$$\lambda(T(t_k)) = \prod_{i=1}^{n-1} |\mu_i(t_k) - \gamma_i(t_k)|$$

$$\lambda(\mathfrak{S}(t_k)) = \frac{1}{(n-1)!} \left| \det \begin{pmatrix} \mathbf{S}^{n-1}(t_k) \\ 1 \dots 1 \end{pmatrix} \right|$$

$\lambda(T(t_k))$ decreases very fast in time \Rightarrow Quick convergence of the algorithm.

In average, for $k > 10$,

$$\lambda(T(t_k)) \leq \lambda(\mathfrak{S}(t_k)) \quad \wedge \quad \lambda(T(t_k)) \leq \lambda(\mathfrak{S}_{tot})$$



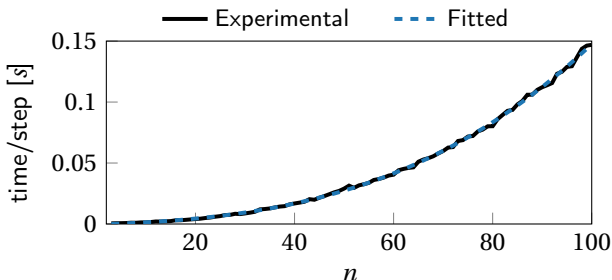
Computational Complexity Analysis

The computational complexity results to be $O(n^3)$ due to the extraction of null spaces bases, which is needed to compute the simplex vertices and performed in each iteration, i.e.

$$\mathbf{A}_i(t_k) = \text{null} [\Sigma(t_k) - \text{diag}(0, \dots, \tilde{\sigma}_i^2(t_k), \dots, 0)] \quad \forall i = 1, \dots, n$$

computed via SVD.

⇒ Suitable for online applications.



[Results obtained with Intel® Xeon E3-1240v5]

Conclusions and Future Works

Approach:

We proposed a **novel estimator**: the BBRF scheme.

Contribution:

- **Reduced size** of the solutions space w.r.t. standard methods;
 - Improved computational efficiency
-
- It is always more convenient to perform the BBRF scheme rather than use the whole data set for a single Frisch estimation;
 - The convergence speed and computational complexity of the proposed algorithm make it useful for online applications (diagnostics, control, etc);
 - This approach can be modified to treat *time-varying systems*.
⇒ Use **dynamic bounding boxes** instead of simple intersections.

Thank you for your attention

Any questions?