Combined Procedure with Randomized Controls for the Parameters' Confidence Region of Linear Plant under External Arbitrary Noise

Konstantin Amelin, Natalia Amelina, Oleg Granichin,
and Olga Granichina

Saint-Petersburg State University,
St. Petersburg, Russia

51st IEEE Conference on Decision and Control
Maui, USA
11 December 2012

Introduction

- Any model is never been the real system perfect description.
- It is important to determine the bounds of uncertainties for the model in which it can still be used.
- The key question in the system identification is a development of methods and procedures which are applicable for wide range uncertainties.

Contradiction of Processes of Control and Gain Knowledge

A goal of control is to achieve some kind of generally stable state (if possible do not change over time).
In this state, invariability provides a "very little information" and, therefore, it is impossible to identify or establish new links, values, and so on.

Feldbaum's concept of *dual control*: *control must be not only directing but also learning.*

Control Via Feedback

Usually, we have

- substantial restrictions of resources,
- an insufficient number of data with the necessary diversity

to "extract" an information from the data in the real time environment.

Example

Let's consider the simple problem of an unknown parameter $\theta_\star$ estimating from the observations:

$$y_t = \theta_\star \cdot u_t + v_t,\ t = 1, 2, ..., N. \tag{1}$$

We can

- to choose the inputs (control actions) u_t;
- to measure the outputs y_t.

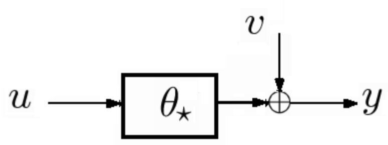
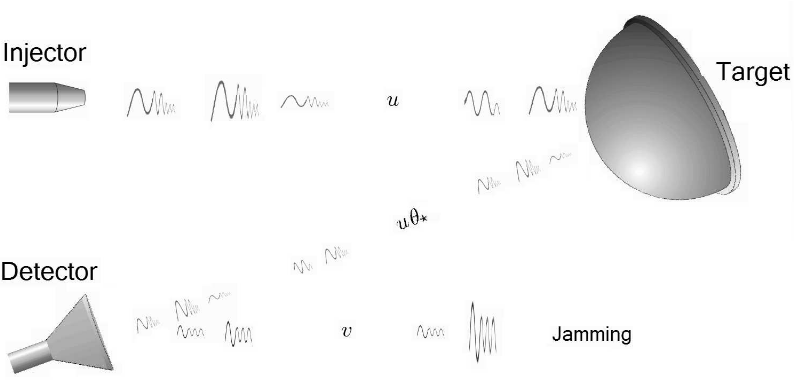


Figure: The model of observations.

The Source, Target and Detector of the Reflected Signal



Algorithm of $\theta_\star$ Estimation

1. Control u_t selection and feeding it to the system input.
2. Receiving the response from the system y_t.
3. Estimation of the parameter $\theta_\star$ based on the data obtained u_t, y_t (for example, calculation of an estimate $\widehat{\theta}_t$ or set $\widehat{\Theta}_t$ containing $\theta_\star$).
4. Repeat steps 1–3.

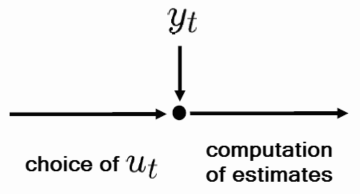


Figure: A model of an estimation algorithm.

Deterministic Algorithm

Definition

An algorithm is called a *deterministic algorithm* if each of its steps defined by the user is given by deterministic rules using the results of the previous steps, and obtained new data (output) is returned for using in subsequent steps of the algorithm.

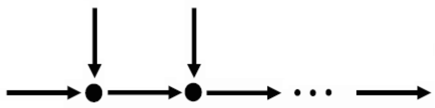


Figure: A model of a deterministic algorithm.

Deterministic Approaches Often Failed!

- In theory and practice many difficulties arise when we try to make analytical investigation of “complex” systems.
- For many practical applications traditionally efficient deterministic methods fail to yield a result when the system dimension is high.
- In particular, this leads to the notion of NP-hard problems.

There are no Deterministic Algorithms Under Arbitrary External Noise!

Let be

$$\theta_\star = 3, \quad \widehat{\theta}_t = \frac{1}{t}\sum_{i=1}^{t} y_i$$

Table:

t	1	2	3	4	5	6	7
u_t	1	1	1	1	1	1	1
$v_t = rand() - 0.5$							
y_t	2.9	2.8	3.2	3.3	2.6	3.4	2.7
$\widehat{\theta}_t$	2.9	2.85	2.97	3.05	2.96	3.03	2.99
$v_t = rand() - 0.5 + m,\ m = 1$							
y_t	3.9	3.8	4.2	4.3	3.6	3.9	4.2
$\widehat{\theta}_t$	3.9	3.85	3.97	4.05	3.96	4.03	3.99

Note, the bias m is an unknown for the user!

Randomized Algorithms

Randomization is a powerful tool for solving a number of problems deemed unsolvable with deterministic methods

Definition

An algorithm is called a *randomized algorithm* when the execution of one or more steps, which are defined by the user, is based on a random rule (that is, among many deterministic rules one is chosen randomly according to a probability P).

"Enriched" Observations

Consider the following rule of a random input selection for the first step

$$u_t = \begin{cases} +1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}. \end{cases} \tag{2}$$

At the second step from the known values (u_t, y_t) we form a value

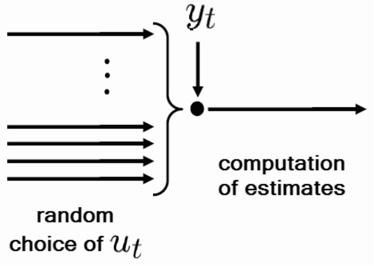
$$\tilde{y}_t = u_t \cdot y_t.$$

For the "new" sequence of observations we have a similar to (1) model

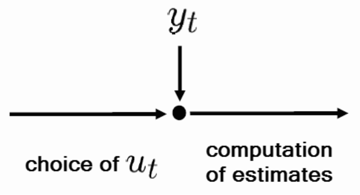
$$\tilde{y}_t = \tilde{u}_t \cdot \theta_\star + \tilde{v}_t,$$

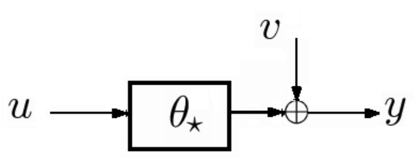
where $\tilde{u}_t = u_t^2$ and $\tilde{v}_t = u_t \cdot v_t$.

Diagram of Randomized Algorithm

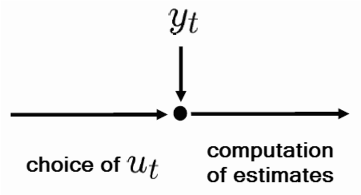


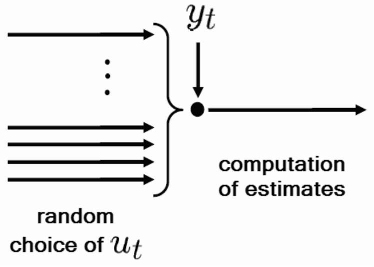
Two Kinds of Agorithms



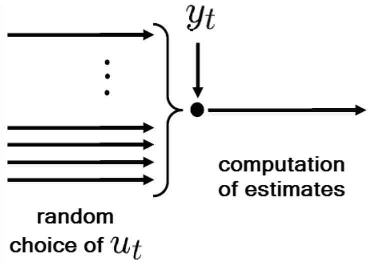


Two Kinds of Agorithms





Two Kinds of Agorithms



Preliminary result

$$\theta_\star = 3, \qquad \widehat{\theta}_t = \frac{1}{t}\sum_{i=1}^{t} \tilde{y}_i = \frac{1}{t}\sum_{i=1}^{t} u_i y_i$$

Table:

t	1	2	3	4	5	6	7
u_t	-1	1	-1	1	1	1	-1
$v_t = rand() - 0.5 + m,\ m = 1$							
y_t	-2.1	3.8	-1.8	4.3	3.6	4.4	-2.3
$\tilde{u}_t$	1	1	1	1	1	1	1
$\tilde{y}_t$	2.1	3.8	1.8	4.3	3.6	4.4	2.3
$\widehat{\theta}_t$	2.1	2.95	2.57	3.00	3.12	3.33	3.19

$$\forall t,\ \forall \varepsilon > 0\ \ Prob\{|\widehat{\theta}_t - \theta_\star| \geq \varepsilon\} \leq \frac{1}{t}\frac{\mathrm{E}\{v_t^2\}}{\varepsilon^2} + o\left(\frac{1}{t}\right).$$

[Granichin, TAC, 2004]

Non-Asymptotic Result

For the finite number of observations ($N = 9$) a new rigorous mathematical result of a guaranteed set of possible values of the unknown parameter $\theta_\star$ can be obtained for an arbitrary external noise v_t following by the method described by M. Campi [EJC, 2010]:

1. Let be $M = 10$ and select randomly nine ($= M - 1$) different groups of four indexes $T_1, \ldots, T_9$.
2. Compute the partial sums $\bar{s}_i = \frac{1}{4} \sum_{j \in T_i} \bar{y}_j,\ i = 1, \ldots, 9$.
3. Build the confidence interval

$$\widehat{\Theta} = [\min_{i \in 1..9} \bar{s}_i; \max_{i \in 1..9} \bar{s}_i],$$

which contains $\theta_\star$ with the probability $p = 80\%$ ($= 1 - 2 \cdot 1/M$).

The Confidence Interval

$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ s_9 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \cdot (-2.1) \\ 1 \cdot 3.8 \\ -1 \cdot (-1.8) \\ 1 \cdot 4.3 \\ 1 \cdot 3.6 \\ 1 \cdot 4.4 \\ -1 \cdot (-2.3) \end{pmatrix} = \begin{pmatrix} 3.375 \\ 3.15 \\ 3.4 \\ 3.15 \\ 3.075 \\ 2.875 \\ 3.275 \\ 3.025 \\ 3.275 \end{pmatrix}$$

- The unknown parameter $\theta_\star$ belongs to the interval $\widehat{\Theta} = [2.875, 3.4]$ with probability $p = 80\%$.

The randomization in the process of the input data selection can get quite reasonable results.

Randomized and Bayesian Approaches

An alternative probabilistic approach is a Bayesian estimation when the noise v_t probability is attributed a priori to a nature Q.

But Bayesian and randomized approaches are quite different from the practical point of view.

In a Bayesian approach the probability Q describes a probability of a value of v_t in a comparison with other, i. e. the choice of Q is a part of the problem model.

In contrast, the probability P in a randomized approach is selected artificially. P exists only in our algorithm, and therefore, there is no a traditional problem of "a bad model" as can happen with the Q in a Bayesian approach.

Randomization …

1930 …

- Fisher (remove bias)

1950 … 1975

- Metropolis, Ulam (method Monte-Carlo)
- Rastrigin, Kirkpatrick, Holland (random search, simulation annealing, genetic algorithm)

1980 … 1999

- Granichin, Fomin, Chen, Guo (randomized control strategies)
- Polyak, Thzubakov, Luing, Guffi, Spall (fast algorithms)
- Granichin (arbitrary noise)
- Vadiyasagar (randomized learning theory)

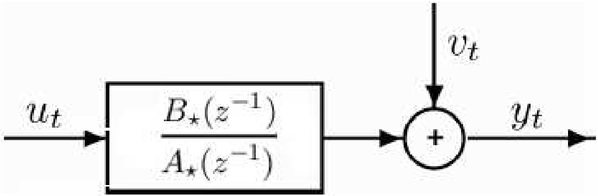
2000 …

- Tempo, Campi, Calafiore, Dabbene, Polyak, Sherbakov *etc.* (probabilistic methods in a control syntheses, scenario approach)
- Candes, Donoho, Romberg, Tao (compressive sensing)

Best Features

- Significantly decreasing the number of operations
- Annihilating the systematic errors (the bias effect or an arbitrary noise)
- Accuracy usually not depend on the dimension of data

Adaptive Control



Let's consider the dynamical system:

$$A_\star(z^{-1})y_t = B_\star(z^{-1})u_t + v_t,\ t = 1, 2, \ldots, N, \tag{3}$$

where $A_\star(\lambda) = 1 + a_\star^{(1)}\lambda + \cdots + a_\star^{(n_a)}\lambda^{n_a},\ B_\star(\lambda) =$
$b_\star^{(l)}\lambda^l + b_\star^{(l+1)}\lambda^{l+1} + \cdots + b_\star^{(n_b)}\lambda^{n_b},$
$\tau_\star = col(a_\star^{(1)}, a_\star^{(2)}, \ldots, a_\star^{(n_a)}, b_\star^{(l)}, b_\star^{(l+1)}, \ldots, b_\star^{(n_b)})$ is the vector of parameters some of which are unknown.

Control Strategy Randomization

$$\text{Goal:} \quad \overline{\lim_{t\to\infty}} |y_t| \to \min, \ \sup_t |y_t| + |u_t| < \infty$$

(Granichin, Fomin, ARC, 1986)

Let be $s \leq n_a + n_b - l + 1$ and $N = s \cdot K \cdot N_\Delta$.

$$u_{sn+i-l} = \begin{cases} \beta_{n \div N_\Delta} \Delta_n \ + \ \bar{u}_{sn-l}, \ i = 0, \\ \bar{u}_{sn+i-l}, \ i \in [1..s-1] \text{ or } i \in [-s+1..-1], \end{cases}$$

$n \in [1..KN_\Delta]$ and "own" controls $\{\bar{u}_t\}$ are determined by the adjustable feedback law

$$\bar{u}_t = \mathcal{U}_t(y_t, y_{t-1}, \ldots, \bar{u}_{t-1}, \ldots), \, \bar{u}_{-t} = 0, \, t \geq 0.$$

"Strip"-algorithm

The type and characteristics of a feedback depend on practical problems specifics:
1) $\bar{u}_t = 0,\ t \in [1..N - l]$
2)stabilized regulator

$$C(z^{-1}, \tilde{\tau}_t)\bar{u}_t = D(z^{-1}, \tilde{\tau}_t)y_t \tag{4}$$

with parameters $\tilde{\tau}_t = \widehat{\tau}_{t-s}$ which are tuning by the "Strip"-algorithm

$$\widehat{\tau}_t = \widehat{\tau}_{t-1} - \frac{(\varphi_t^{\mathrm{T}}\widehat{\tau}_{t-1} - y_t)\mathbf{1}_{\{|\varphi_t^{\mathrm{T}}\widehat{\tau}_{t-1} - y_t| - 2C_v - \delta\|\varphi_t\| > 0\}}}{\|\varphi_t\|^2}\varphi_t, \tag{5}$$

where $\varphi_t = (-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-l}, \ldots, u_{t-n_b})^{\mathrm{T}}$, and
$A(\lambda, \tau)C(\lambda, \tau) - B(\lambda, \tau)D(\lambda, \tau)$ is a stable polynomial.

Random Perturbations

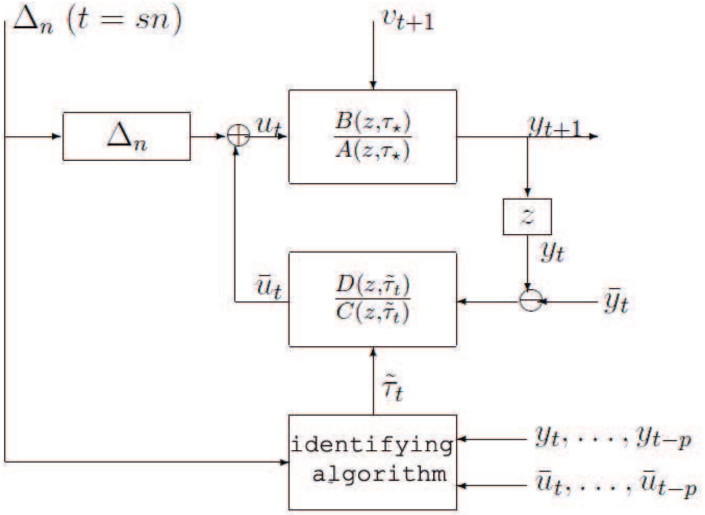
$\Delta_0, \Delta_1, \ldots, \Delta_{KN_\Delta - 1}$ are measurable:

$$E\{\Delta_n\} = E\{\Delta_n^3\} = 0,\ E\{\Delta_n^2\} = \sigma_\Delta^2,\ E\{\Delta_n^4\} \leq M_4$$

$$\text{For example,}\quad \Delta_n = \begin{cases} +1, \text{with probability} \frac{1}{2}, \\ -1, \text{with probability} \frac{1}{2}. \end{cases}$$

A1. The user can choose Δ_n and this choice does not affect to the external noise $v_{sn}, \ldots, v_{s(n+1)-1}$.

Diagram of Adaptive Control



Reparametrization

We can rewrite the model (3) as a linear regression

$$y_{sn} = \beta_{n \div N_\Delta} \Delta_n \theta_\star^{(1)} + \theta_\star^{(1)} \bar{u}_{sn-l} + \bar{v}_{sn}, \quad \theta_\star^{(1)} = b_\star^{(l)}$$

and

$$y_{sn+i-1} = \beta_{n \div N_\Delta} \Delta_n \theta_\star^{(i)} + \sum_{j=0}^{i-1} \theta_\star^{(i-j)} \bar{u}_{sn-l+j} + \bar{v}_{sn+i-1}. \tag{6}$$

$$\theta_\star = \theta(\tau_\star),\ \theta(\tau) = \mathbb{A}^{-1}(\tau)\mathbb{B}(\tau), \tag{7}$$

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ a_\star^{(1)} & 1 & \dots & 0 & 0 \\ a_\star^{(2)} & a_\star^{(1)} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & a_\star^{(n_a)} \dots & a_\star^{(1)} & 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} b_\star^{(l)} \\ \vdots \\ b_\star^{(n_b)} \\ \vdots \\ 0 \end{pmatrix}.$$

(Granichin, Fomin, ARC, 1986)

How to choose s?

$$\mathbf{A2}. \qquad s : \exists \;\; \tau(\theta) = \theta^{-1}(\tau)$$

$s = n_a + n_b$ if polynomials $A_\star(\lambda)$ and $B_\star(\lambda)$ are mutually prime

Example

Consider the second-order plant

$$y_t + a_\star^{(1)} y_{t-1} + y_{t-2} = b_\star^{(1)} u_{t-1} + 1.6 u_{t-2} + v_t, \tag{8}$$

$t = 1, 2, \ldots, N$, with unknown coefficients $a_\star^{(1)}$ and $b_\star^{(1)} \neq 0$. Denote

$$\tau_\star = col(a_\star^{(1)}, b_\star^{(1)}).$$

Let $s = 2$ and vector $\theta_\star$ of the "new" parameters be

$$\theta_\star = \begin{pmatrix} b_\star^{(1)} \\ 1.6 - a_\star^{(1)} b_\star^{(1)} \end{pmatrix} \in \mathbb{R}^2.$$

In this case, the inverse function $\tau(\theta)$ is

$$\tau(\theta) = \begin{pmatrix} \frac{1.6-\theta^{(2)}}{\theta^{(1)}} \\ \theta^{(1)} \end{pmatrix}.$$

Unknown Parameters Estimation

We combine two different algorithms:

- Stochastic approximation algorithm with perturbation in the input (Granichin, Fomin, ARC, 1986)
- Leave-out Sign-dominant Correlation Regions (LSCR) method (M. Campi and E. Weyer, TAC, 2010)

Stochastic Approximation Algorithm

$$\begin{cases} \bar{\tau}_t = \tau(\widehat{\theta}_{n-1}),\ t \in [s(n-1)+1..sn],\, n \in [1..KN_\Delta], \\ \widehat{\theta}_n^{(i)} \ = \ \widehat{\theta}_{n-1}^{(i)} - \frac{1}{n}\Delta_n(u_{sn-l}\widehat{\theta}_{n-1}^{(i)} - y_{sn+i-1}),\ i \in [1..s]. \end{cases} \quad (9)$$

$$\tilde{\tau}_t = \begin{cases} \bar{\tau}_t, & \text{if } |y_t| + |u_{t-1}| < \bar{R}, \\ \widehat{\tau}_{t-s} & \text{— otherwise.} \end{cases} \quad (10)$$

Theorem

If **A1–A2** *and* $2\sigma_\Delta^2 > 1,\ E\{v_t^2\} \le \sigma_v^2,\ \sum_{n=1}^{\infty} \frac{\beta^{-1}_{n\div N_\Delta}}{n} = \infty,$ $\sum_{n=1}^{\infty} \frac{\beta^{-2}_{n\div N_\Delta}}{n^2} < \infty$ **then** $\forall \rho > 0$

$$E\{|\widehat{\theta}_n^{(i)} - \theta_\star^{(i)}|^2\} \le \frac{\beta^{-1}_{n\div N_\Delta}}{n} \frac{\rho C_{\bar{u}}^2 + \sigma_{i,\bar{v}}^2}{2\sigma_\Delta^2 - 1} + o\left(\frac{\beta^{-1}_{n\div N_\Delta}}{n}\right),\ i \in [1..s], \quad (11)$$

$C_{\bar{u}} = \sup_t \bar{u}_t,\ \sigma_{i,\bar{v}}^2 = \sigma_v^2 + (C_{\bar{u}}^2 + \sigma_\Delta^2)\sum_{j=1}^{i-1} |\theta_\star^{(j)}|^2.$

LSCR method

1 $\forall k \in [1..K]$ consider $[k's..k's + sN_\Delta - 1]$ where $k' = (k-1)N_\Delta$.

2 $\hat{y}_{k's+sn+i-1}(\theta) = \beta_k \Delta_{k'+n} \theta^{(i)} + \sum_{j=0}^{i-1} \theta^{(i-j)} \bar{u}_{k's+sn+i-l-j}$.

3 $\epsilon_t(\theta) = y_t - \hat{y}_t(\theta),\ t \in [k's..k's + sN_\Delta - 1]$.

4 $f_{k's+sn+i-1}(\theta) = \Delta_{k'+n} \epsilon_{k's+sn+i-1}(\theta)$.

5 Choose $M > 2s$ and construct M different binary stochastic strings $(h_{j,1}, \ldots, h_{j,sN_\Delta})$. We calculate

$$g_{k,j}^{(i)}(\theta) = \sum_{n=0}^{N_\Delta - 1} h_{j,ns+i} \cdot f_{k's+ns+i-1}(\theta),\ i \in [1..s].$$

6 Choose $q \in [1, M/2s]$ and construct $\widehat{\Theta}_k^{(i)}$: at least q of the $g_{k,j}^{(i)}(\theta)$ functions are strictly higher than 0 and at least q functions are strictly lower than 0.

Confidence Set

We define the confidence set by the formula

$$\widehat{\Theta}_k = \bigcap_{i=1}^{s} \widehat{\Theta}_k^{(i)}. \tag{12}$$

Theorem

Let condition **A1** *be satisfied. Consider* $i \in [1..s]$ *and assume that* $Prob(g_{k,j}^{(i)}(\theta_\star) = 0) = 0$. **Then**

$$Prob\{\theta_\star \in \widehat{\Theta}_k^{(i)}\} = 1 - 2q/M \tag{13}$$

$$Prob\{\theta_\star \in \widehat{\Theta}_k\} \geq 1 - 2sq/M \tag{14}$$

where M, q *and* $\widehat{\Theta}_k^{(i)}$ *are from steps 5 and 6 of the above-described procedure.*

Combined Algorithm

1. For $k = 1, 2, \dots, K$.
2. To generate the sequence $\{\widehat{\theta}_n\}$ by the algorithm (9).
3. To choose $\varepsilon > 0$ and to build the parallelepiped

$$\bar{\Theta}_k = \prod_{i=1}^{s} \{\theta^{(i)} : |\widehat{\theta}^{(i)}_{kN_\Delta} - \theta^{(i)}| \leq \varepsilon\}.$$

4. To choose q and M and to compute the region $\widehat{\Theta}_k$ by the algorithm (12).
5. We define the confidence set as the intersection $\widetilde{\Theta}_k = \widehat{\Theta}_k \cap \bar{\Theta}_k$.

Theorem

$$Prob\{\theta_\star \in \tilde{\Theta}_k\} \geq (1 - 2sq/M) \left(1 - \frac{\beta_k}{kN_\Delta} \frac{\rho C^2_{\bar{u}} + \sigma^2_{s,\bar{v}}}{\varepsilon^2 (2\sigma^2_\Delta - 1)}\right)^s + o\left(\frac{\beta_k}{kN_\Delta}\right).$$

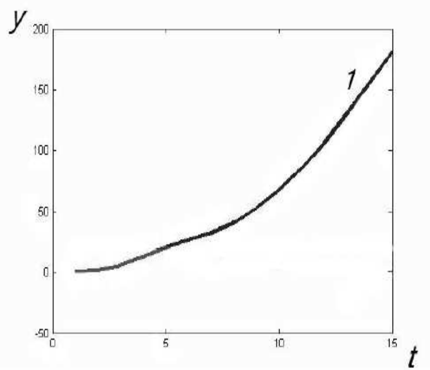
Thank you for your attention!

Example

$$y_t - 2y_{t-1} + y_{t-2} = b_\star u_{t-1} + 1.6u_{t-2} + v_t,\ t = 1, \ldots, 15,$$

$y_0 = y_{-1} = u_{-1} = 0$, $b_\star$ is an unknown coefficient, v_t is an unknown external arbitrary noise.

LSCR, $s = 1$

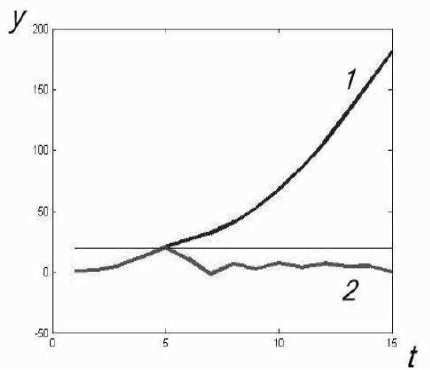


Example

$$y_t - 2y_{t-1} + y_{t-2} = b_\star u_{t-1} + 1.6u_{t-2} + v_t,\ t = 1, \dots, 15,$$

$y_0 = y_{-1} = u_{-1} = 0$, $b_\star$ is an unknown coefficient, v_t is an unknown external arbitrary noise.

LSCR + adaptive stabilizing feedback



Example

$$y_t - 2y_{t-1} + y_{t-2} = b_\star u_{t-1} + 1.6u_{t-2} + v_t,\ t = 1, \dots, 15,$$

$y_0 = y_{-1} = u_{-1} = 0$, $b_\star$ is an unknown coefficient, v_t is an unknown external arbitrary noise.

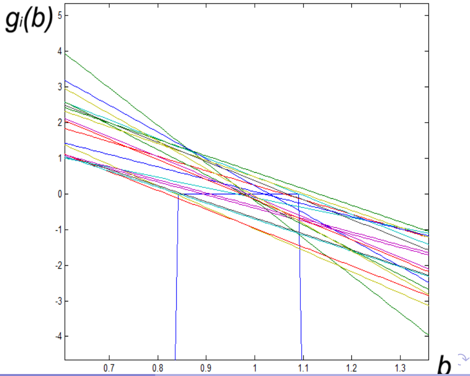
LSCR

$b_\star = 1,\ Ev_t = 0.5,\ \sigma_v = 0.1$

$$g_i(b) = \sum_{t=1}^{15} h_{i,t} \cdot \Delta_{t-1}\epsilon_t(b),$$

$$h_{i,t} \in \{0, 1\},\ i = 1, \dots, 19.$$

The confidence interval is $[0.834; 1.090]$

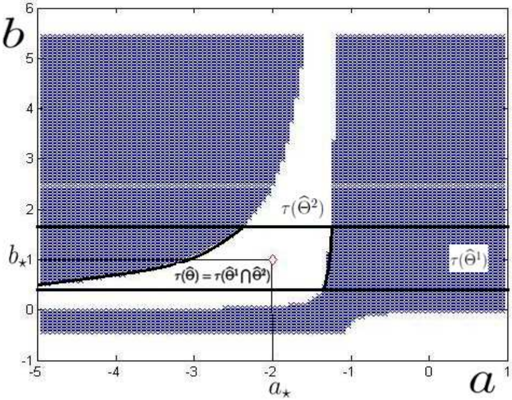


Example

$$y_t + a_\star^1 y_{t-1} + y_{t-2} = b_\star^1 u_{t-1} + 1.6 u_{t-2} + v_t,\ t = 1, 2, \ldots, 960,$$

$$Prob\{\theta_\star \in \widehat{\Theta}^k\} = 1 - 2q/M,$$

$$95\% = (1 - 2 \cdot 2 \cdot 6/480) \cdot 100\%.$$

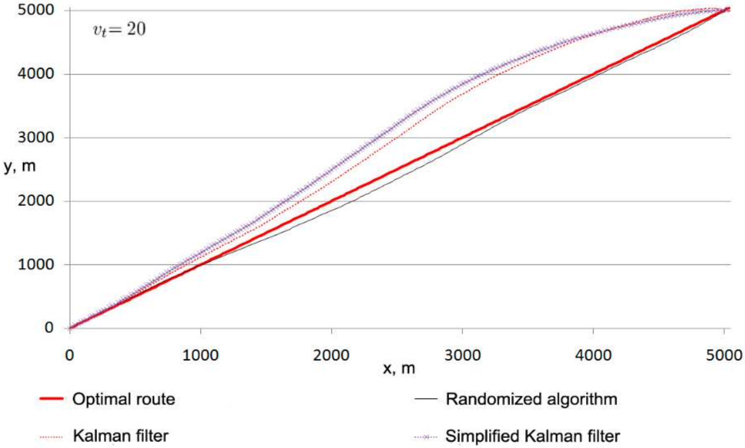


Randomized algorithm for the small UAV flight optimization

Intel Lab. "Sprint"

$$\begin{cases} u_t = \bar{u}_{t-1} + \Delta_t, \\ \widehat{\theta}_{t+1} = \widehat{\theta}_t - \alpha \Delta_t \varepsilon_t, \\ \bar{u}_t = \frac{b}{a} \sin \widehat{\theta}_{t+1}. \end{cases} \tag{15}$$

Simulation results



External noise	RA	KF	SKF
$v_t = 10 \cdot (\text{rand}\,() \cdot 4 - 2)$	41.36	38.15	42.65
$v_t = 0.1 \cdot \sin(t) + 19 \cdot \text{sign}(50 - \text{t mod}100)$	53.4	197.64	212,45
$v_t = 20$	45.15	276.35	169.48

- Assumption: $\{\Delta_n\}_{n=0}^{N-1}$ and $\{v_t\}_{t=0}^{N}$ are independent. Hence, adaptive control schemes are not applicable. (We can not use the current estimates of parameters in a feedback).
- Randomization adds to the control channel at each step. It disturbs the system permanently.
- Algorithm dimension is so high (even for the simplest cases).

Sketch of the proof

Proposition 1: Fix $k \in [1, \ldots, s]$. Let H be a stochastic $M \times N_\Delta$ matrix with elements $h_{i,ns+k}$, $i = 0, 1, \ldots, M-1$, $n = 0, \ldots, N_\Delta - 1$, from step 4 of the algorithms in Section VI, and let $\eta = col(\eta_1, \ldots, \eta_{N_\Delta})$ be a vector independent of H, consisting of mutually uncorrelated random variables symmetrically distributed around zero. Given an $i \in [0, M-1]$, let H_i be the $M \times N$ matrix, whose rows are equal to the i-th row of H. **Then**, $H\eta$ and $(H - H_i)\eta$ have the same M-dimensional distribution provided that the i-th element of $(H - H_i)\eta$ (which is 0) is repositioned as the first element of the vector.

(M. Campi and E. Weyer, TAC, 2010)

Sketch of the proof

Denote $\eta_n := \Delta_{n-1}\epsilon_{(n-1)s+k-1}(\theta_\star)$.
For the correlation between η_i and η_j, $i > j$,:

$$E[\eta_i\eta_j] = E[\Delta_{i-1}]E[\epsilon_{(i-1)s+k-1}(\theta_\star)\Delta_{j-1}\epsilon_{(j-1)s+k-1}(\theta_\star)] = 0$$

$E[\Delta_{i-1}] = 0$ ($\eta_1, \dots, \eta_{N_\Delta}$ are mutually uncorrelated).
Take $g_{\bar{i}}^{(k)}(\theta_\star)$ in the r-th position.

$$g_i^{(k)}(\theta_\star) - g_{\bar{i}}^{(k)}(\theta_\star) = \sum_{n=0}^{N_\Delta}(h_{i,ns+k} - h_{\bar{i},ns+k})\eta_n < 0$$

for $r-1$ selection of $i \in [0, M-1]$.
From Proposition 1: $Prob\{$"$r-1$ entries of $(H - H_{\bar{i}})\eta$ are negative"$\} = Prob\{$"$r-1$ entries of $H\eta$ are negative"$\}$, and it does not depend on $\bar{i}$.