# RANDOMIZED CONTROLS FOR LINEAR PLANTS AND CONFIDENCE REGIONS FOR PARAMETERS UNDER EXTERNAL ARBITRARY NOISE

Konstantin Amelin, Oleg Granichin

Saint-Petersburg State University

27 June 2012

# Introduction

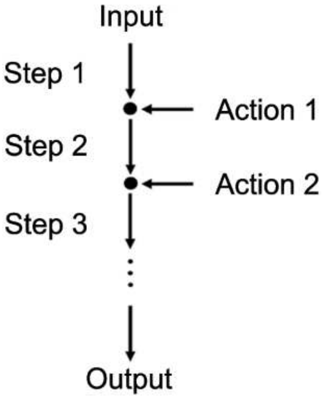
- Any model is never been the real system perfect description.
- It is important to determine the bounds of uncertainties for the model in which it can still be used.
- The key question in the system identification is a development of methods and procedures which are applicable for wide range uncertainties.

# Control via feedback

Usually in a real time environment we have:

- restrictions of resources;
- an insufficient amount of data with necessary varieties.

# Deterministic algorithm




Every new step is defined by a deterministic rule which uses results of previous steps (outputs, new information about the system).

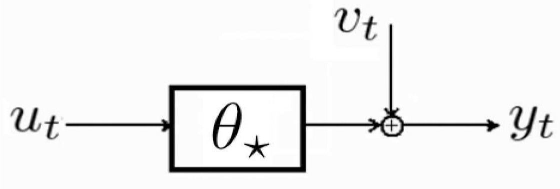
# Deterministic approaches often failed !

In theory and practice many difficulties arise when we try to make analytical investigation of “complex” systems.

In many practical applications traditionally efficient deterministic methods fail to yield a result when the system is complex.

In particular, this leads to the $NP$-hard problems.

# Example. The unknown parameter estimation




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

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

where we can

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

# An arbitrary external noise

There are no deterministic algorithms under arbitrary external noise!

$\theta_\star = 3$

| $t$ | 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 |

$$\widehat{\theta}_t = \frac{1}{7}\sum_{i=1}^{7} y_i$$

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

# Randomized algorithms

- A randomized algorithm is an algorithm where one or more steps are based on a random rule, that is, among many deterministic rules, one rule is selected according to a random scheme
- Randomization is a powerful tool for solving a number of problems deemed unsolvable with deterministic methods

# “Enriched” observation (“Flip a coin” approach)

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

$$u_t = \begin{cases} +1, \text{with probability} \frac{1}{2}, \\ -1, \text{with probability} \frac{1}{2}. \end{cases} \quad (2)$$

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

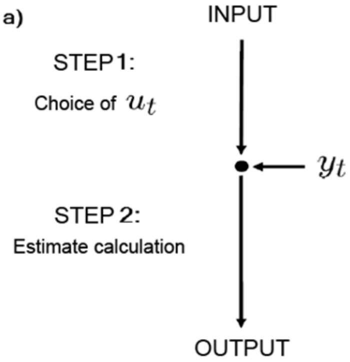
$$\tilde{y}_t = u_t \cdot y_t. \quad (3)$$

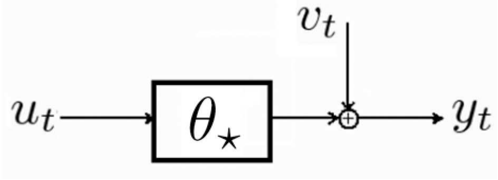
For the “new” observations we have a model

$$\tilde{y}_t = \theta_\star \cdot \tilde{u}_t + \tilde{v}_t, \quad (4)$$

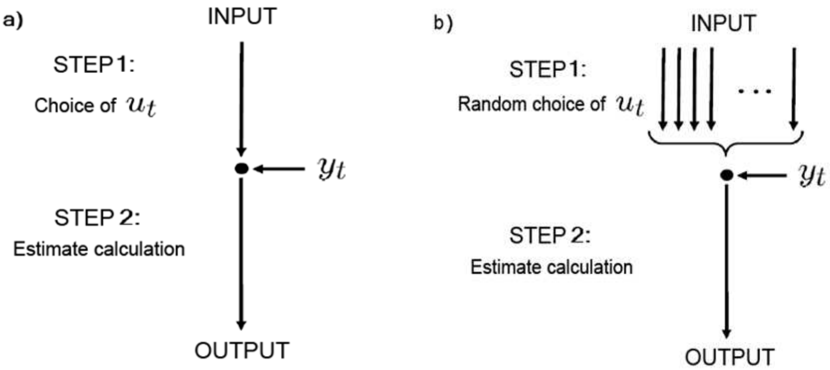
where $\tilde{u}_t = u_t^2 = 1$ and $\tilde{v}_t = u_t \cdot v_t$.

# Two types of the algorithms

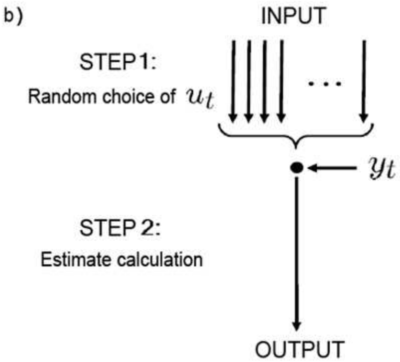


# Two types of the algorithms



# Two types of the algorithms



# Preliminary result

$\theta_\star = 3$

| $t$ | 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 |
| $\bar{u}_t$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\bar{y}_t$ | 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 |

$$\widehat{\theta}_t = \frac{1}{t}\sum_{i=1}^{t}\bar{y}_i = \frac{1}{t}\sum_{i=1}^{t}u_i y_i \to \theta_\star \text{ with probability 1}$$

(O. Granichin, TAC, 2004)

# Algorithm for the guaranteed set

1. Let be $M = 8$ and select randomly seven $(= M - 1)$ different groups of four indexes $T_1, \ldots, T_7$
2. Compute the partial sums $\bar{s}_i = \sum_{j \in T_i} \bar{y}_j,\ i = 1, \ldots, 7$
3. Build the confident interval

$$\widehat{\Theta} = [\min_{i \in 1:7} \bar{s}_i; \max_{i \in 1:7} \bar{s}_i],$$

which contains $\theta_\star$ with the probability $p = 75\%$ $(= 1 - 2 \cdot 1/M)$.

(M. Campi, EJC, 2010)

# The confidence interval

| $i$ | $T_i$ | $\bar{s}_i$ |
|---|---|---|
| 1 | {2, 3, 4, 5} | 3.375 |
| 2 | {1, 3, 4, 6} | 3.15 |
| 3 | {2, 3, 5, 6} | 3.4 |
| 4 | {1, 2, 6, 7} | 3.15 |
| 5 | {1, 4, 5, 7} | 3.075 |
| 6 | {2, 3, 5, 7} | 2.875 |
| 7 | {1, 4, 6, 7} | 3.275 |

The unknown parameter $\theta_\star$ belongs to the interval $\widehat{\Theta} = [2.875; 3.4]$ with the probability $p = 75\%$.

# Probability successful algorithm

Randomized algorithm is named
"probability successful" with the probability $p$
if the probability of it's true result is not less than $p$
(M. Campi, EJC, 2010)

# Randomization . . .

1930 . . .

- Fisher (remove bias)

1950 . . .

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

1980-90

- Polyak, Thzubakov, Luing, Guffi, Spall (fast algorithms)
- Granichin (arbitrary noise)
- Vadiyasagar (Randomized Learning Theory)

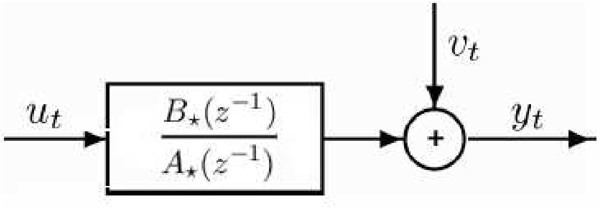
2000 . . .

- Tempo, Campi, Calafiore, Sherbakov *etc.*
- . . .

# Best features

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

# Adaptive control




Let's consider the dynamical system:

$$A_\star(z^{-1})y_t = B_\star(z^{-1})u_t + v_t,\ t = 1, 2, \ldots, N, \tag{5}$$

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)

$s \in \mathbb{N},\ N = s \cdot N_\Delta.$

$$u_{sn+i-l} = \begin{cases} \Delta_n + \bar{u}_{sn-l}, & i = 0, \\ \bar{u}_{sn+i-l}, & i = 1, 2, \dots, s-1, \end{cases} \quad n = 0, \dots, N_\Delta - 1,$$

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

$\Delta_n$ is a measurable random trial perturbation

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

# Unknown parameters estimation

$N < \infty, \ s = 1$

Leave-out Sign-dominant Correlation Regions (LSCR) method
(M. Campi and E. Weyer, TAC, 2010)

But

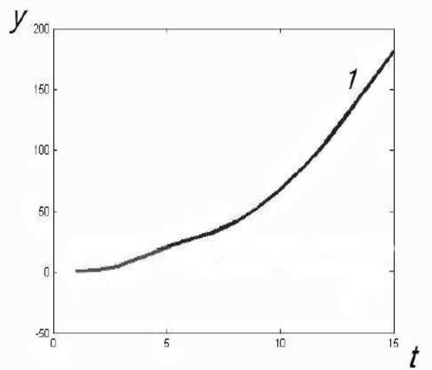
- Assumption: $\{\Delta_n\}_{n=0}^{N-1}$ and $\{v_t\}_{t=0}^{N}$ are independent. Hence, adaptive control schemes are not applicable. (We can not use the current estimates of parameters in a feedback).
- Randomization adds to the control channel at each step. It disturbs the system permanently.
- Algorithm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

# 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, $s = 1$

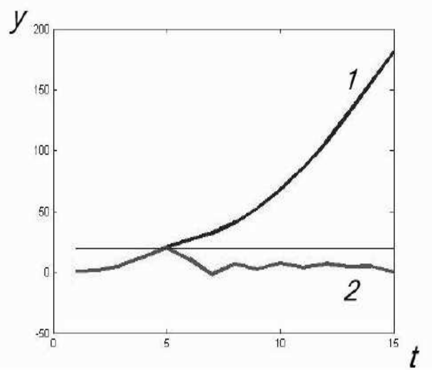


# Example

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

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

LSCR + adaptive stabilizing feedback



# Example

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

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

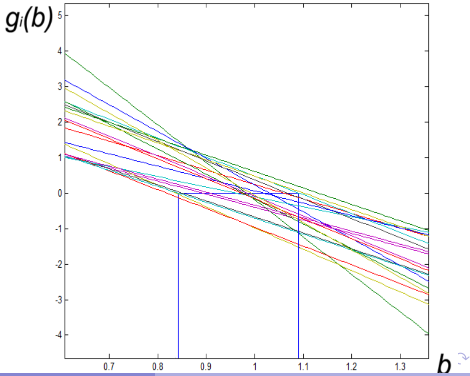
LSCR

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

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

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

The confidence interval is $[0.834; 1.090]$



# Reparametrization

We can rewrite the model (5) as a linear regression

$$y_{sn+k-1} = \Delta_n \theta_\star^{(k)} + \bar{v}_{sn+k-1} = \Delta_n \theta_\star^{(k)} + \sum_{i=0}^{k-1} \theta_\star^{(k-i)} \bar{u}_{sn-l+i} + \dots, \qquad (6)$$

with inputs $\Delta_n$ and regressors $\theta_\star^{(k)}$, $n = 1, \dots, N_\Delta$, $k = 1, \dots, s$.

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

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ a^1 & 1 & \dots & 0 & 0 \\ a^2 & a^1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & a^{n_a} \dots & a^1 & 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} b^l \\ \vdots \\ b^{n_b} \\ \vdots \\ 0 \end{pmatrix}.$$

(Granichin, Fomin, ARC, 1986)

Note, that $\Delta_n$ and $\bar{v}_{sn+k-1}$ are independent if the user can choose $\Delta_n$ and this choice does not affect to the external noise.

# How to choose $s$?

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

$s = n_a + n_b$ 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

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

# The algorithm

1. $\widehat{y}_{sn+k-1}(\theta) = \Delta_n \theta^k + \sum_{i=0}^{k-1} \theta^{k-i} \bar{u}_{sn+k-l-i}$.
2. $\epsilon_t(\theta) = y_t - \widehat{y}_t(\theta),\ t = 1, \dots, N$.
3. $f_{sn+k-1}(\theta) = \Delta_n \epsilon_{sn+k-1}(\theta),\ n = 0, \dots, N_\Delta - 1, k = 1, \dots, s$.
4. To choose $M > 2s$ and different random binary strings $(h_{i,1}, \dots, h_{i,N}),\ i = 0, \dots, M-1$.
   To calculate $g_i^k(\theta) = \sum_{n=0}^{N_\Delta - 1} h_{i,ns+k} \cdot f_{ns+k-1}(\theta),\ i = 0, \dots, M-1,$
5. To choose $q \in [1; M/2s]$ and for $k = 1, \dots, s$ to build the regions:

$$\widehat{\Theta}^k = \{\theta :\ \text{at least } q \text{ of } g_i^k(\theta) \text{ are } > 0 \text{ and at least } q \text{ are } < 0\}.$$

$$\widehat{\Theta} = \bigcap_{k=1}^{s} \widehat{\Theta}^k. \quad (9)$$

# Theorem

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

***Theorem***

*Let condition* **A1** *be satisfied. Consider* $k \in \{1, 2, \ldots, s\}$ *and assume that* $Prob(g_i^k(\theta_\star) = 0) = 0$. **Than**

$$Prob\{\theta_\star \in \widehat{\Theta}^k\} = 1 - 2q/M, \quad (10)$$

where $M$, $q$ and $\widehat{\Theta}^k$ is taken from 4 and 5 steps.

# 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 $\eta_i$ and $\eta_j$, $i > j$,:

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

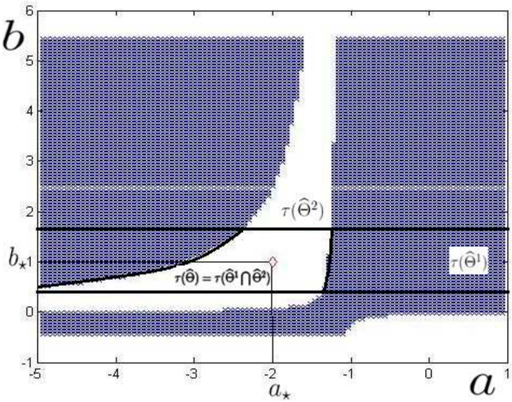
$E[\Delta_{i-1}] = 0$ ($\eta_1, \ldots, \eta_{N_\Delta}$ are mutually uncorrelated).
Take $g_{\bar{i}}^{(k)}(\theta_\star)$ in the $r$-th position.

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

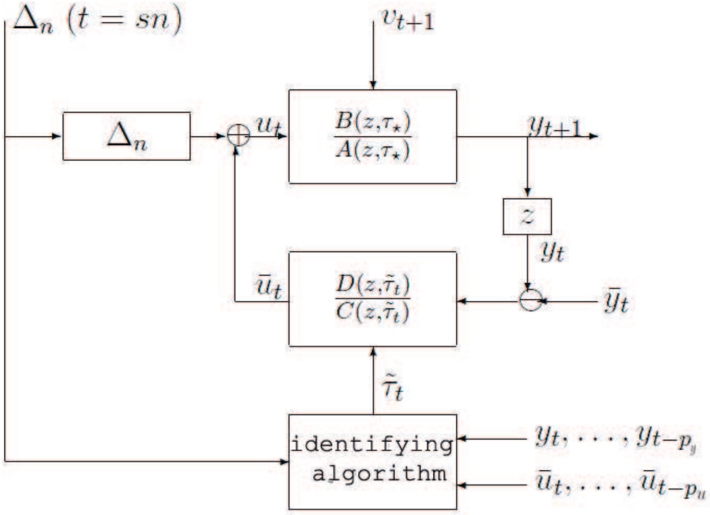
for $r-1$ selection of $i \in [0, M-1]$.
From Proposition 1: $Prob\{$"$r-1$ entries of $(H - H_{\bar{i}})\eta$ are negative"$\} = Prob\{$"$r-1$ entries of $H\eta$ are negative"$\}$, and it does not depend on $\bar{i}$.

# Example

$$y_t + a^1_\star y_{t-1} + y_{t-2} = b^1_\star u_{t-1} + 1,6u_{t-2} + v_t,\ t = 1, 2, \dots, 960,$$
$$Prob\{\theta_\star \in \widehat{\Theta}^k\} = 1 - 2q/M,$$
$$95\% = (1 - 2 \cdot 2 \cdot 6/480) \cdot 100\%.$$

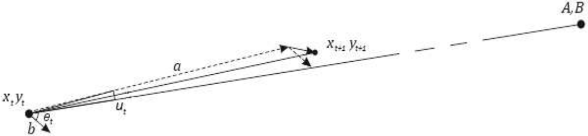


# Adaptive control block-scheme



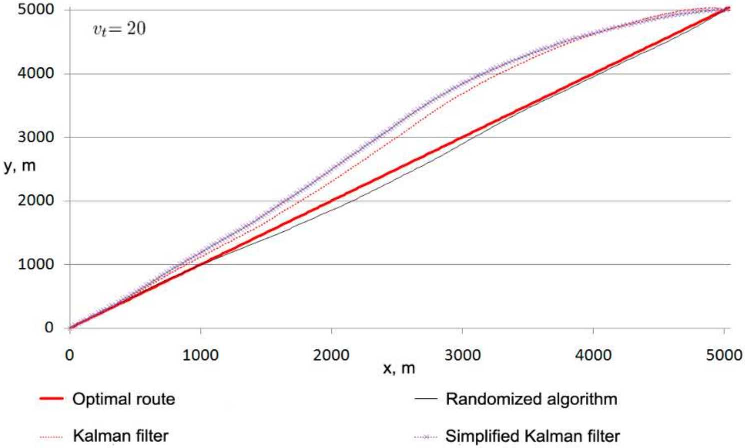
# Randomized algorithm for the small UAV flight optimization

Intel Lab. "Sprint"




$$\begin{cases} u_t = \bar{u}_{t-1} + \Delta_t, \\ \widehat{\theta}_{t+1} = \widehat{\theta}_t - \alpha \Delta_t \varepsilon_t, \\ \bar{u}_t = \frac{b}{a} \sin \widehat{\theta}_{t+1}. \end{cases} \tag{11}$$

# 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 |

Thank you for your attention!