Stochastic Programming and Multiagent Technologies

Oleg Granichin

Saint Petersburg State University

3rd Algorithm Workshop
provided by Huawei Russian R&D Center and SPbUAI
09 June 2016

Outline

Outline

Introduction

In the recent years, control of multi-agent robotic systems has received a great deal of attention in science and technology. The miniaturization and increased computing power of sensors and actuators can further extend the range of applicability of the results of modern control, identification, and estimation theories. In particular, theoretical issues of adaptive control in changeable environment and under time-varying structure of the state space have been only marginally considered so far due to the limited ability of practical realization, and require more attention.

Control and Data Processing "on the Fly"

Usually, to "extract" information from the data in the real time environment we have:

- substantial restrictions of hardware resources
- an insufficient ammount of input data with the necessary diversity

Outline

Stochastic Programming

Stochastic programming is a framework for modeling of the optimization problems that involve uncertainty. Whereas deterministic optimization problems are formulated with known parameters, real world problems almost always include some unknown parameters. One of the approaches for solving such problems, when the parameters are known only within the certain bounds, is called the robust optimization. Here, the goal is to find a solution, which is feasible for all such data and is optimal in some sense. Stochastic programming models are similar in style, but take the advantage of the fact that probability distributions governing the data are known or can be estimated. The goal here is to find some policy that is feasible for all (or almost all) the possible data instances and minimizes the expectation of some decision functions and the random variables. More generally, such models are formulated, solved analytically or numerically, and analyzed in order to provide useful information to a decision-maker.

Let us assume that we can choose points of measurements $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathbb{R}^d$ and on each iteration we can measure:

$$y_t = f(\mathbf{x}_t, \mathbf{w}_t) + v_t, \tag{1}$$

where $f : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$, $\mathbf{w}_t$ is a random vector which is defined on the basic probability space $\{\Omega, \mathscr{F}, P\}$. It represents non-controlled random uncertainty, and v_n is an external arbitrary observation noise.
Consider the minimization problem:

$$F(x) = \int f(\mathbf{x}, \mathbf{w}) \mathrm{P}(d\mathbf{w}) \to \min_{\mathbf{x}}. \tag{2}$$

Finite Difference Approach

The number of observations per one iteration is $N = 2d$
$\widehat{\theta}_0 \in \mathbb{R}^d$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{2\beta_n}(\mathbf{Y}_n^+ - \mathbf{Y}_n^-),$$

$$x_n^{(i,\pm)} = \widehat{\theta}_{n-1} \pm \beta_n e_i$$

$$\mathbf{Y}_n^\pm = \begin{pmatrix} f\big(x_n^{(1,\pm)}, w_n^{(1,\pm)}\big) + v_n^{(1,\pm)} \\ f\big(x_n^{(2,\pm)}, w_n^{(2,\pm)}\big) + v_n^{(2,\pm)} \\ \vdots \\ f\big(x_n^{(d,\pm)}, w_n^{(d,\pm)}\big) + v_n^{(d,\pm)} \end{pmatrix}$$

Randomized Stochastic Approximation (SPSA)

We reduce the number of observations up to 1 or 2 instead of $2d$ (!)
One measurement form

$$\mathbf{x}_n = \widehat{\theta}_{n-1} + \beta_n \Delta_n, \;\; \Delta_n = \begin{pmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{pmatrix}$$

$$y_n = f(\mathbf{x}_n, \mathbf{w}_n) + v_n$$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \mathscr{K}_n(\Delta_n) y_n$$

Two measurements form

$$\mathbf{x}_n^{\pm} = \widehat{\theta}_{n-1} \pm \beta_n^{\pm} \Delta_n$$

$$y_n^{\pm} = f(\mathbf{x}_n^{\pm}, \mathbf{w}_n^{\pm}) + v_n^{\pm}$$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n^{+} + \beta_n^{-}} \mathscr{K}_n(\Delta_n)(y_n^{+} - y_n^{-})$$

Advantages of Algorithms

- Asymptotic-optimal rate of convergence (Polyak and Tsybakov, 1990)
- Min number of observations per iteration (Spall, 1992, 1997)
- Consistency under arbitrary external noise (Granichin, 1989, 2002)
- Allows tracking (Granichin et. al. 2009, 2015)
- Easy to implement on quantum computer (Granichin et. al. 2006, 2015)
- Finite (small) amount of data (Granichin et. al. 2012, 2016 – LSCR)

Tracking

Let Ξ be a set, $\{f_\xi(\mathbf{x},\mathbf{w})\}_{\xi\in\Xi}$ be a family of differentiable functions

$$y_t = f_{\xi_t}(\mathbf{x}_t,\mathbf{w}_t) + v_t$$

$$F_t(\mathbf{x}) = \int f_{\xi_t}(\mathbf{x},\mathbf{w}_t)\mathrm{P}(d\mathbf{w}_t)\mathrm{P}(d\xi_t) \to \min_{\mathbf{x}} \tag{3}$$

Granichin et. al. (2015, 2009) SPSA with constant step-size $\alpha_n = \alpha > 0$
Example:

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \zeta,\ \mathbf{x}_n \in \mathbb{R}^d$$

Randomized Algorithm and Quantum Computing

The representation of SPSA algorithm is associated with something well known to those familiar with the fundamentals of quantum computing. Virtually all known effective quantum algorithms implement a similar scheme:

- preparing input "superposition"
- processing
- measuring a result

$$\mathbf{u} = \frac{1}{2^{\frac{d}{2}}} \sum_{\mathbf{\Delta}_i \in \{-1,+1\}^d} |\widehat{\mathbf{x}} + \beta \mathbf{\Delta}_i\rangle = \mathbf{H}_\beta |\widehat{\mathbf{x}}\rangle$$

$$\mathbf{U}_f |\mathbf{u}\rangle |0\rangle = \frac{1}{2^d} \sum_{\mathbf{\Delta}_i \in \{-1,+1\}^d} |\widehat{\mathbf{x}} + \beta \mathbf{\Delta}_i\rangle |f(\widehat{\mathbf{x}} + \beta \mathbf{\Delta}_i)\rangle$$

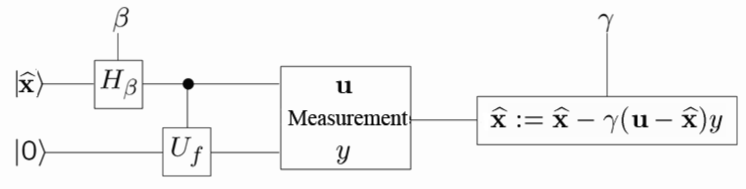


Figure: The quantum circuit for "on the fly" computing of the gradient.

Machine Learning and Network Optimization Applications

- Machine Learning
- UAV Soaring
- Server Adaptive Optimization
- Network Load Balancing

Machine Learning

Vapnik and Chervonenkis (1968)
A fundamental connection between learning problems and "Uniform convergence of the frequencies of occurrence of events to their probabilities":

- if convergence takes place uniformly, minimizing empirical values y from (1) with respect to $\mathbf{x}$ returns a nearly minimizer of the corresponding expected error $E_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})$

UAV Soaring

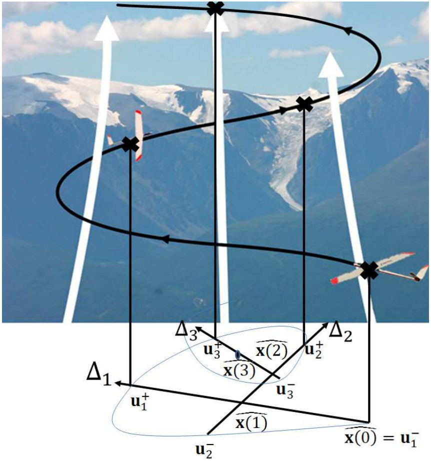


Figure: The sequence of estimates and waypoints.

Server Adaptive Optimization

$$F(x) = q(x) + L(x) \equiv q(x) + \lim_{T} \frac{1}{T} \sum_{t=1}^{T} Ey_t(x) \to \min_{x}$$

$$y_n = \frac{t_n - t_{n-1}}{x} d_{load} + \frac{1}{N} \sum_{t_i^{out} \in [t_{n-1}, t_n]} (t_i^{out} - t_i^{in} - d_i) \qquad (4)$$

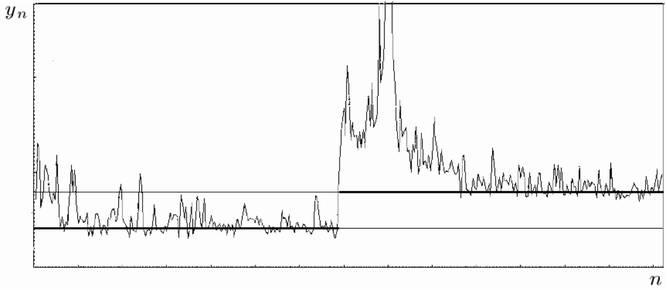


Figure: The behavior of y_n for tuning abrupt changes in the parameters of the input stream.

Network Load Balancing

Let $N = \{1, \dots, n\}$ be a set of agents (nodes) with states $x_t^i = q_t^i / p_t^i$:

- q_t^i is a task queue length of agent i
- p_t^i is an agent i productivity

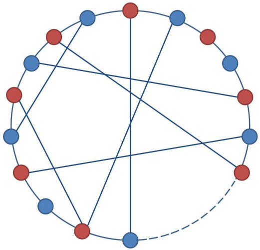
Dynamic model:

$$x_{t+1}^i = x_t^i + f_t^i + u_t^i, \ i \in N = \{1, \dots, n\},$$

or

$$q_{t+1}^i = q_t^i - p_t^i + z_t^i + p_t^i u_t^i; \ \ i \in N, \ t = 0, 1, \dots$$

Network topology example:



Optimal Rules

IEEE TAC 2015 (Granichin, Amelina)

$$z_t = \sum_{i=1}^{n} x_t^i,\ x_t^i = ?$$

$$T(\mathbf{x}) = \max_{i \in N}\ time_i(x_t^i) \to \min_{\mathbf{x}}$$

Consensus:

$$p_t^1 x_t^1 = p_t^2 x_t^2 = \cdots = p_t^n x_t^n$$

Mean risk:

$$F_t(\mathbf{x}) = \sum_{i,j=1}^{n} a_t^{i,j}(time_t^i - time_t^j)^2 \ \to\ \min_{\mathbf{x}}$$

IEEE TAC 2015 (Granichin, Amelina)

Outline

Consensus Control

Agents' dynamics:

$$x_{t+1}^i = f(x_t^i, u_t^i),\ i \in N = \{1, \dots, n\}$$

Consistent behavior (consensus):

$$x_t^i \approx x_t^j,\ i, j \in N$$

Observations:

$$y_t^{i,j} = x_{t-d_t^{i,j}}^j + w_t^{i,j},\ j \in N_t^i$$

Local Voting Protocol:

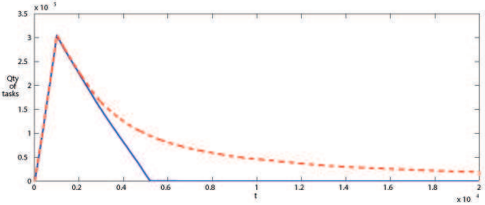
$$u_t^i = \alpha \sum_{j \in \bar{N}_t^i} a_t^{i,j}(y_t^{i,j} - y_t^{i,i})$$

IEEE TIT 2015 (Amelina, Fradkov *et. al.*)

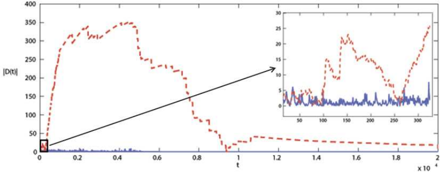
Simulation

- $n = 1024$
- Number of tasks is 10^6

Queue lengths:



The average deviation from consensus:



Usual Model of Data Processing

The Increasing Complexity of Interactions

The Increasing Complexity of Interactions

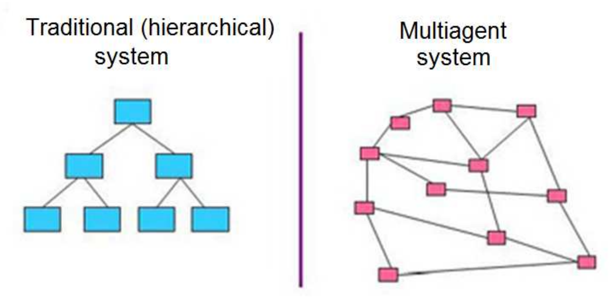
The Increasing Complexity of Network Interactions

The Increasing Complexity of Network Interactions

The Increasing Complexity of Network Interactions

Time to Collect and Prepare the Data Becomes Critical

Muliagent Systems (From 1990s)



The achievements in the following areas were taken as a basis:

- Artificial Intelligence
- Parallel Computing
- Distributed Problem Solving

Multiagent Control

The application of the traditional mathematical models of system motion with a large number of transducers / sensors and actuators often leads to extremely complex problems, involving extremely high-dimensional state spaces. The multi-agent technology can effectively solve many of the problems arising in this context by replacing the general model of interaction with a complex system, containing the multiple local models and their aggregation (clustering).

- Network Load Balancing
- UAV Group Planning
- Swarm Control

New Directions in Distributed Systems Research

Networked/cooperative control systems and distributed parameter systems are two instances of dynamical systems, distributed in discrete and continuum space.

Among new directions in distributed systems research were outlined fascinating connections between distributed systems theory on the one hand, and canonical problems in turbulence and statistical mechanics on the other. In one class of problems, spatio-temporal dynamical analysis clarifies old and vexing questions in the theory of shear flow turbulence. In another class of problems, structured, distributed control design exhibits dimensionality-dependence and phase transition phenomena similar to those in statistical mechanics.

What is an actual structure of complex systems?

Complex systems:

$X(t)$ is a state vector, $W = \begin{pmatrix} u \\ w \end{pmatrix}$ is an external disturbances

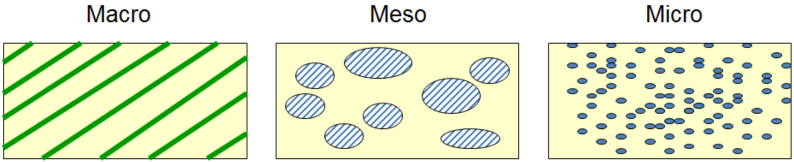
Dynamic equations:

$$\dot{x}_i = g_i(X, W),\ i = 1, 2, \ldots, n,\ X \in \mathbb{R}^n \tag{5}$$

or

$$\dot{x}_\gamma = g_\gamma(X, W),\ X = \{x_\gamma\},\ \gamma \in [0, 1] \tag{6}$$

Levels of Model Description, Self-organization



In open thermodynamic systems, synergistic processes often form new dynamic structures at the mesoscopic level. These processes are associated with an internal control feedback, which, together with an external control, leads to the discretization of state space and time of non-equilibrium systems.

Cluster Flow

Let W be structured, and structure s_k changes in time instances $T_0, T_1, T_2, \ldots$

Clustering of state space:

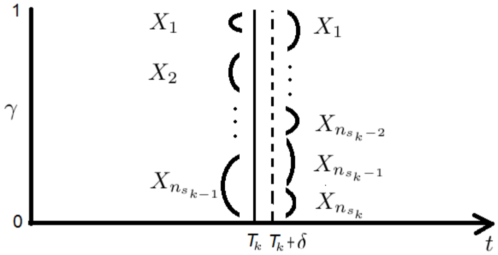
$$\mathscr{X}_{s_k} = \{X_1, X_2, \ldots, X_{n_{s_k}}\} : \; X = \cup_{i=1,2,\ldots,n_{s_k}} X_i, \; X_i \subset X \tag{7}$$

Dynamic Equations:

$$\dot{\bar{x}}_i = \bar{g}_i(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{n_{s_k}}, u, w, \theta_{s_k}), \; i = 1, 2, \ldots, n_{s_k}, \tag{8}$$

where $\bar{x}_i$ is a set of averaged x_γ on cluster X_i,
θ_{s_k} is a finite set of "current" parameters

State Space Structure Changing



$$\delta << \zeta = \min_k |T_{k+1} - T_k|$$

System Identification Issues

Measurements:

$$Y(t) = \int_{t-\Delta}^{t} \int_{\mathbb{M}} f(X, W) dX dt' = \int_{t-\Delta}^{t} \bar{f}_k(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{n_{s_k}}, u, w, \theta_{s_k}) dt' \quad (9)$$

After disretization and simplifications, we have:

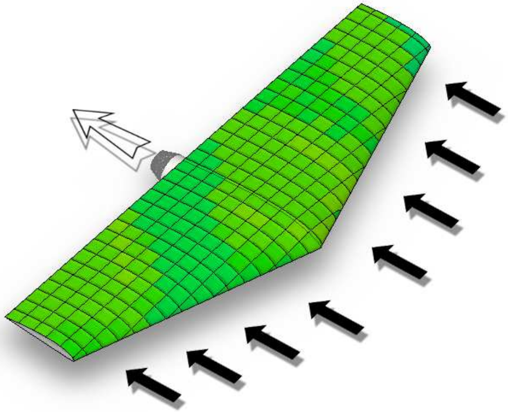
$$y_i = \tilde{f}_k(x_i, u_i, w_i, \theta(s_k)) + \xi_i \quad (10)$$

where $y_i = Y(t_i)$, $t_i \in [T_k, T_{k+1}]$, $\tilde{f}_k(\cdot)$ are functions of

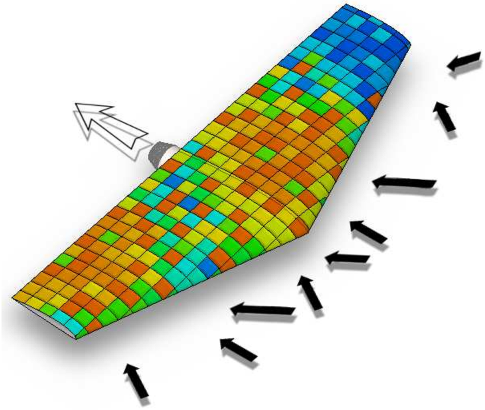
$$x_i = col(\bar{x}_1(t_i), \bar{x}_2(t_i), \ldots, \bar{x}_{n_{s_k}}(t_i)),\ u_i = u(t_i),\ w_i = w(t_i),\ \theta_{s_k},$$

$\xi_i = \xi_i' + \xi_i(s_k)''$ is a standard error (noise),
ξ_i' is a random noise, $\xi_i(s_k)''$ is a systematic errors

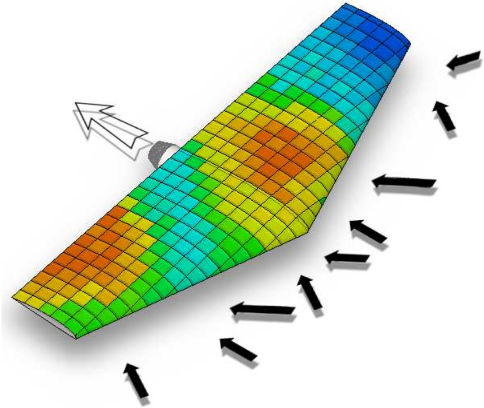
Airplane with “Feathers” in Laminar Wind Flow



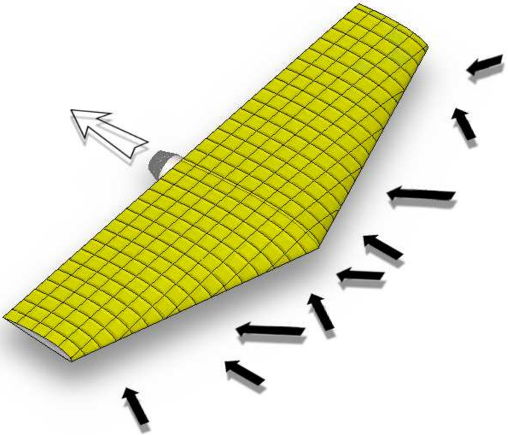
Turbulent Flow



Clustering of “Feathers”

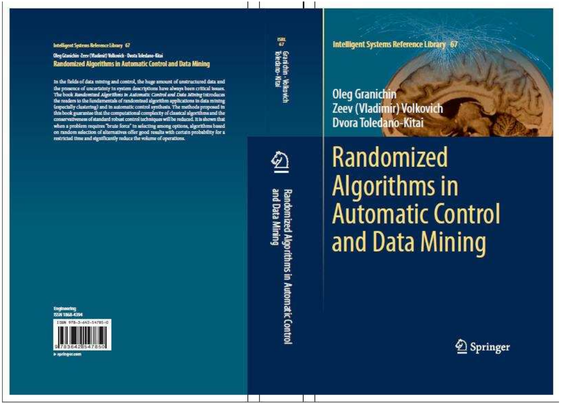


“Alignment” of Pressures in Turbulent Wind Flow



Outline

About Myself



New book

Recent Publications

1) *Randomized control strategies under arbitrary external noise* // IEEE Transactions on Automatic Control (WoS impact-factor 3.713), 2016, vol. 61, May, issue 5, pp. 1328–1333. Amelin K., Granichin O.

2) *Modeling and visualization of media in Arabic* // Journal of Informetrics (WoS impact-factor 2.932), 2016, vol. 10, issue 2, pp. 439–453. Volkovich Z., Granichin O., Redkin O., Bernikova O.

3) *Literary writing style recognition via a minimal spanning tree-based approach* // Expert Systems With Applications (WoS impact-factor 2.571), 2016, vol. 61, Nov., pp.145–153. Shalymov D., Granichin O., Klebanov L., Volkovich Z.

4) *Estimating the position of a moving object based on test disturbance of camera position* // Automation and Remote Control, 2016, vol. 77, No. 2, pp. 297–312. Krivokon' D.S., Vakhitov A.T., Granichin O.N.

5) *Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances* // IEEE Transactions on Automatic Control (WoS impact-factor 3.713), 2015, vol. 60, June, issue 6, pp. 1653–1658. Granichin O., Amelina N.

Thank you for your attention!

Randomization for Reducing a Computational Complexity

The general idea is:
Suppose that a deterministic algorithm requires a huge amount of computing resources to process all available information.
Then we can intentionally give up part of the information and proceed to the solution of the simplified problem with partial information.
In this case, however, a deterministic solvability may be impossible to achieve, but as above we can consider a randomized approach to defining solutions with a high probability of success. The end result is a compromise between the full guarantee of success and computational feasibility (the opportunity to get a real answer for a limited time).

Randomized and Bayesian Approaches

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

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

In Bayesian approach 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 Q in a Bayesian approach.

Randomization ...

1928-30 ...

- von Neumann (minimax theorem), 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

SPSA Ground

$$E\{\widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n}\Delta_n y_n | \mathscr{F}_{n-1}\} =$$

$$= \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n}(E\{\Delta_n f(x_n) | \mathscr{F}_{n-1}\} + E\{\Delta_n\}E\{v_n | \mathscr{F}_{n-1}\}) =$$

$$= \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n}(E\{\Delta_n f(\widehat{\theta}_{n-1} + \beta_n \Delta_n) | \mathscr{F}_{n-1}\} \approx$$

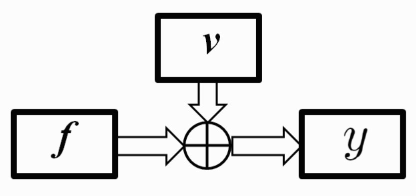
$$\approx \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n}(E\{\Delta_n f(\widehat{\theta}_{n-1}) + \frac{\beta_n \Delta_n \Delta_n \nabla f(\widehat{\theta}_{n-1})}{2} | \mathscr{F}_{n-1}\} =$$

$$= \widehat{\theta}_{n-1} - \frac{\alpha_n}{2}\nabla f(\widehat{\theta}_{n-1})$$

Arbitrary External Noise

If some signal f goes into the recorder with a noise v then the "instantaneous" observation y_t can be written as

$$y_t = \mathrm{A} f_t + v_t \tag{11}$$



- $v = 0$.
- $v \approx 0$.
- $v_t \to 0$ as $t \to \infty$.
- v_t, $t = 1, \ldots, T$, is i.i.d. with $\sigma_v < \infty$
- Arbitrary external noise

Data Averaging

If our registering apparatus averages signals f_t coming at $t = 1, \ldots, T$ then at the output we receive

$$y = \frac{1}{N}\sum_{t=1}^{T} f_t + \frac{1}{N}\sum_{t=1}^{T} v_t \tag{12}$$

If v_t, $t = 1, \ldots, T$, is i.i.d. with mean value M_v and variance $\sigma_v < \infty$ then

$$\mathrm{Prob}\left\{ |\frac{1}{N}\sum_{t=1}^{T} v_t - M_v| > \varepsilon \right\} \to 0 \ \ as \ \ t \to \infty.$$

Hence, we can use estimates

$$\hat{f} = \frac{1}{N}\sum_{t=1}^{T} y_t - M_v.$$

Estimation under Arbitrary Noise

Is it possible to get smart estimates?

Modernize the problem by including into the observations model the controllable input u. Following the paradigm inseparability of an information and control, we assume that the measured signal f at time t is directly determined by the current input u_t and some unknown parameter $\theta_\star$ (an unknown coefficient of gain/attenuation inputs).

$$f_t = u_t \theta_\star. \tag{13}$$

The problem is *to find or estimate* the unknown parameter $\theta_\star \in \mathbb{R}$ by the sequence of inputs and outputs $\{u_t, y_t\}$ without any restrictions for the sequence $\{v_t\}$ of external noises.

Problem Description

The model of observations (11) can be rewritten as:

$$y_t = u_t \theta_\star + v_t. \tag{14}$$

And we can

- chose the inputs (controls) u_t, $t = 1, 2, ..., T$,
- measure the outputs y_t (see Fig. 4).

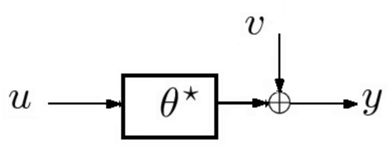
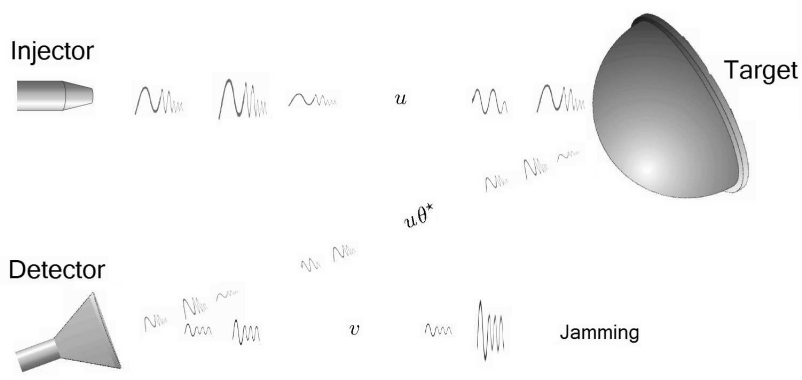


Figure: The model of observations

If we use $u_t \equiv 1$, we obtain the traditional problem of estimating of unknown parameter $\theta_\star$ observed with noise

The Source, Target and Detector of the Reflected Signal



An Algorithm of Estimation of $\theta_\star$

1. Control u_t selection and feeding it to the system input.
2. Receive the response from the system y_t.
3. Estimate 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.

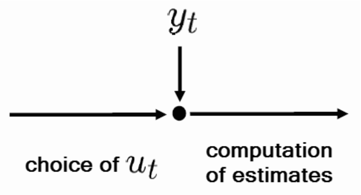


Figure: A model of an estimation algorithm.

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.

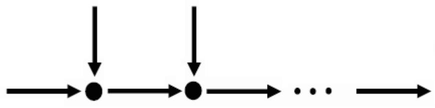


Figure: A model of a deterministic algorithm.

Deterministic Approaches are Often Fail !

- 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 notion of *NP*-hard problems.

There are no Deterministic Algorithm Under Arbitrary External Noise!

Let be $\theta_\star = 3$

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

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

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

$$u_t = \begin{cases} +1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}. \end{cases} \tag{15}$$

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

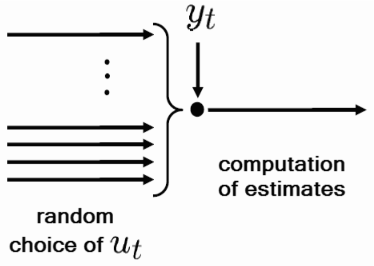
$$\tilde{y}_t = u_t \cdot y_t.$$

For the “new” sequence of observations we have a similar to (14) model

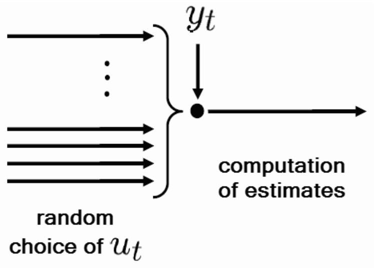
$$\tilde{y}_t = \tilde{u}_t \cdot \theta_\star + \tilde{v}_t,$$

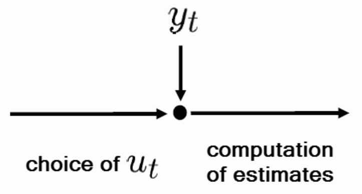
where $\tilde{u}_t = u_t^2$ and $\tilde{v}_t = u_t \cdot v_t$.

Diagram of Randomized Algorithm



Two Kinds of Algorithms





Results of Simulation

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

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 ($N = 7$) a new rigorous mathematical result of a guaranteed set of possible values of the unknown parameter $\theta_\star$ can be obtained for arbitrary external noise v_t following by the method described by M. Campi [EJC, 2010]:

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 $\bar{s}_i = \frac{1}{4} \sum_{j \in T_i} \bar{y}_j$, $i = 1, \ldots, 7$.
3. Build the confidence 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 $p = 75\%$ ($= 1 - 2 \cdot 1/M$).

Confidence Interval

For the previous data $\{(u_t, y_t)\}$ we obtain by the described method:

Table:

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

Hence,

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

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