Article

Not peer-reviewed version

Dynamic System Identification via Randomized Stochastic Optimization Under Unknown-but-Bounded Noise

Oleg Granichin , Yury Ivanskiy * , Kseniia Denisovna Kopylova

Posted Date: 7 December 2023

doi: 10.20944/preprints202312.0504.v1

Keywords: mesoscopic observations; randomized stochastic optimization; unknown-but-bounded noise

Article

Dynamic System Identification via Randomized Stochastic Optimization Under Unknown-but-Bounded Noise

Oleg Granichin [1,2], Yury Ivanskiy [1,2,*] and Kseniia Kopylova [1]

1 Saint Petersburg State University, 199034 St. Petersburg, Russia; o.granichin@spbu.ru
2 Institute for Problems in Mechanical Engineering Russian Academy of Sciences, 199178 St. Petersburg, Russia
* Correspondence: y.ivansky@spbu.ru

Abstract: Discretization in time and in the state space of the system leads to the necessity to solve the parameter identification problems for dynamic systems in a limited time (at a finite time interval) using observations obtained under the influence of unknown-but-bounded noise. Finding the solution in this case is more difficult compared to traditional identification problem setting which considers random independent zero-mean noise. For system parameter identification problem under unknown-but-bounded noise, a randomized stochastic optimization algorithm is given in the paper, estimates for the mean square values of the residuals for a finite observation interval are obtained. An example of application of the given method to the problem of tuning the parameters of a multi-mirror telescope is considered.

Keywords: mesoscopic observations; randomized stochastic optimization; unknown-but-bounded noise

1. Introduction

An exact solution to any problem is possible with an accurate formulation of the problem, but connections and relationships in the real world are so complex and diverse that it is almost impossible to describe many phenomena strictly in mathematical language. A typical approach in theory is to select a mathematical model close to real processes and include various noises related, on the one hand, to the roughness of the mathematical model and, on the other hand, characterizing uncontrolled external disturbances affecting the considered object or system. For all mathematical models, the result of the experiment is a mathematical object: a number, a set of numbers, a curve, etc. From a mathematical point of view, a significant range of applied problems aims to restore characteristics from experimental data (parameters) of the object. At the same time, real systems are rarely described thoroughly by limited mathematical models. When choosing a model to solve a real problem, it is common to consider so-called systematic error (model error), which can be quantified by the distance from the real operator to the selected model. Another type of error that an experimenter may encounter is associated with measurement errors. Such errors are called statistical errors (random errors). The process of selecting characteristics (parameters) of a model from a given class of models to describe the results in a best way is one of the general definitions of estimation. In practice the estimation process can often be related to some quantitative characteristic of the quality of estimation and, it is natural while choosing the estimates, to try to minimize the negative impact of errors, both statistical and, if possible, systematic [2]. Examples include:

- formation of multiscale vortex structures in turbulent fluid flows and plastic flow of solids under pulsed load;
- clustering in a stream of concentrated dispersive mixtures;
- propagation of the shock wave front inside the substance;
- transition layers near interphase boundaries;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

- processes of protein formation in cells;
- hierarchy of structures in living systems;
- processes of fission of heavy elements nuclei;
- thermonuclear fusion;
- as well as the behavior of groups of people.

2. Discretization in System State Space

Among new directions in distributed systems research connections between distributed systems theory on the one hand, and canonical problems in turbulence and statistical mechanics on the other could be suggested. In one class of problems, spatio-temporal dynamical analysis clarifies old and complex 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.

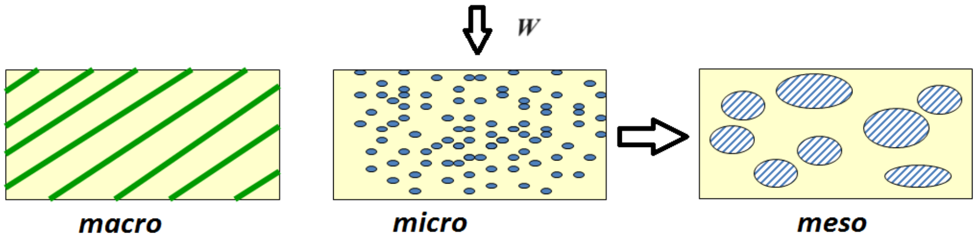

Figure 1. Mesoscale system structure can occur as a result of external influence at the system with microscale structure. It exhibits the properties of both micro- and macroscale structure. On the one hand, the clusters at meso-scale can be considered as a set of different structures, as single particles at microscale but of larger size. On the other hand each cluster could be viewed as a rigid structure, as a macroscale object but of smaller size.

Assume that (unidirectional) time t is introduced, and consider non-isolated systems consisting of elements. Evolution of each system over time is determined by the current states of both the system itself and other elements of the system. The evolution is also affected by external disturbing influences W (the absence of influence can be interpreted as zero impact). External influences W can be formally included in the general set of system states. The inclusion of W into of the system state can significantly complicate the descriptive model. External influences W naturally fall into two groups:

$$W = \begin{pmatrix} u \\ w \end{pmatrix}, \tag{1}$$

controlled ones u (or simply, control) and uncontrolled ones w.

It is usually assumed that at time t the system state $X(t)$ is finite-dimensional, and to describe the dynamics of the system a system of differential equations is used

$$\dot{x}_i = g_i(X, W), \ X = \{x_i\}, \ i \in \mathbb{M} = \{1, 2, \ldots, m\} \tag{2}$$

with some functions $g_i(\cdot)$ and external disturbances W. In models of complex systems of this type, consisting of a huge number of components, it is customary to assume a large dimension n of state space. But, on the one hand, the choice of the threshold for n significantly limits the "upper bound" of complexity, and, on the other hand, does not allow to take possible "flexibility" of the system into account during its changing process. We will assume that the system consists of a continuum of elements $X = \{x_\gamma\}$, parameterized by $\gamma \in [0, 1]$, and the evolution of each of the elements is described by equation

$$\dot{x}_\gamma = g_\gamma(X, W), \ X = \{x_\gamma\}, \ \gamma \in \mathbb{M} = [0, 1]. \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

Such setups can occur, for instance, in stochastic games with many players [3] and mean-field games with almost infinitely many players [4]. Additionally, we assume that the external influence W for all its arbitrariness at each moment of time k has a structure s_k of finite order. Finite structure of external influence after some transition process causes discretization of spatial elements (clustering) in the considered complex system

$$\mathcal{X}_{s_k} = \{X_1, X_2, \ldots, X_{m_{s_k}}\} : \ X = \cup_{i=1,2,\ldots,m_{s_k}} X_i, \ X_i \subset X. \tag{4}$$

Discretization occurs due to the self-organization of groups of elements and their synchronization. For clusters $i = 1, \ldots m_{s_k}$, a set of $m_{s_k} \in N$ variables $\bar{x}_i$ averaged over cluster X_i is naturally introduced. Set of $\bar{x}_i$, $i = 1, \ldots, m_{s_k}$ could be generalized as a set of some integrals over the clusters. Such approach is usually applied to simplify physical models (dimension reduction). The general integral characteristics of clusters of elements with similar properties are introduced and dynamic models in reduced state spaces are considered. Experiments show that such simplifications are justified and often give good results.

Typically, the process of clustering (self-organization) in a system is not "one-shot", but is constantly reproduced due to changes in external influences and critical changes in internal states. But at the same time we will assume that a change in the structure of external influences does not occur permanently, but in some time instants $T_0, T_1, T_2, \ldots$. This leads to the necessity to consider the dynamic processes under condition of state space structure change over time.

So, when the structure of external influence changes, the discretization of spatial elements may change. Let us assume that this transient process takes a duration of time no longer than some $\delta \geq 0$ (see Figure 2). We will assume that δ is many times smaller than the intervals between successive changes in the structure of external influences:

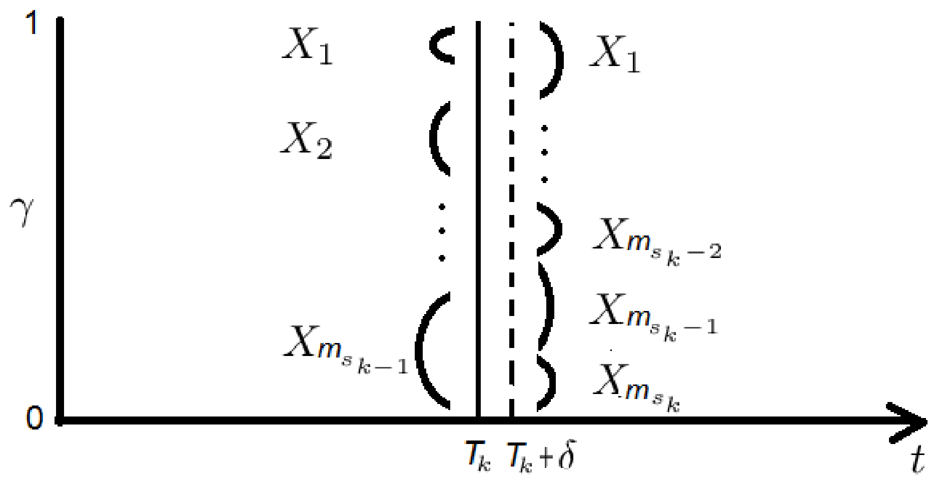

Figure 2. State space structure change at time interval $[T_k, T_k + \delta]$. Up to time instant T_k the system has the structure s_{k-1}. After T_k the system structure changes as a result of some disturbance. During time period δ the system transformation to a new state happens and from time instant $T_k + \delta$ the system has new structure s_k.

$$\delta << \zeta = \min_k |T_{k+1} - T_k| \tag{5}$$

In addition to discretizing spatial variables we obtain time sampling, neglecting duration of transient process intervals. After such discretization in many practical applications the system of differential equations describes the dynamics of changes in the original complex system over the time interval from $T_k + \delta$ to T_{k+1}:

$$\dot{\bar{x}}_i = \bar{g}_i(\bar{X}, u, w, \theta_{s_k}), \ i = 1, 2, \ldots, m_{s_k}, \tag{6}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

where $\bar{x}_i$ is aggregated state $\{x_i\}$ or $\{x_\gamma\}$ from cluster X_i, $\bar{X} = col(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{m_{s_k}})$, θ_{s_k} is a finite set of current parameters at time interval $[T_k + \delta, T_{k+1})$.

3. Control Problem and Discretization on Time

Consider a control problem of choosing the strategy of control minimizing the cost function comprised of local cost functions computed at different parts of the system:

$$L(\{u\}) = \sum_{i \in \mathbb{M}} l(x_i, u) \to \min_u$$

or, in case the system consists of continuum elements (3):

$$L(\{u\}) = \int_{\mathbb{M}} l(x_\gamma, u) d\gamma \to \min_u .$$

Previously it was assumed perturbation W has a "finite structure" at each moment, and the structure of s_k changes at times $T_0, T_1, T_2, \ldots$, causing clustering of the state space:

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

where $\bar{x}_i$ is aggregated state $\{x_i\}$ or $\{x_\gamma\}$ of cluster X_i, $\bar{X} = col(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{m_{s_k}})$, θ_{s_k} is finite set of current parameters. We assume that for any k the dimension of θ_{s_k} is bounded by d. For $t \in [T_k + \delta, T_{k+1})$ due to integral (sum) additive property the loss function could be changed to:

$$L = \int_{\mathbb{M}} l(x_\gamma, u) dx_\gamma \approx \sum_i^{m_{s_k}} \bar{l}_k(\bar{x}_k, u, \theta_{s_k}) = \bar{L}_k(\bar{X}, u, \theta_{s_k}).$$

Assume that control strategy u are piece-wise constant and changing in the end of each time interval of length h,

$$u(t) = u_n, \ t \in [(n-1)h, nh).$$

The feedback u will be computed on the base of noised observations of the loss function $\tilde{L}$. After sampling in time and space, we obtain an observation model for the loss function:

$$y_n = \tilde{L}_k(\bar{X}_n, u_n, \theta_{s_k}) + \xi_n, \tag{8}$$

where $t_n \in [T_K + \delta, T_{k+1})$, $\tilde{L}_k(\cdot)$ are functions from $\bar{X}_n = \bar{X}(t_n), u_n = u(t_n), \theta_{s_k}, \xi_n = \xi'_n + \xi_n(s_k)$" is discrepancy (error) composed of some random noise ξ'_n independent of current system structure s_k, and systematic error $\xi_n(s_k)$" which is, in general, is some function of the current system state.

4. Parameter Identification Problem

Estimation of system parameters values can be formulated as an optimization problem. The discrepancy between estimated and real parameter values could be expressed in terms of some loss function value and thus to solve the system identification problem one has to minimize given loss function. To identify the system structure it is required to choose such control formation strategy $\{u\}$, which minimizes some loss function.

Suppose that for a given parameter vector $\theta_{s_k} \in \mathbb{R}^d$ the optimal control strategy

$$\{u_n\} = \mathcal{U}(\theta_{s_k})$$

is known. After substituting it in (8), we get the problem of minimizing the function

$$f_{s_k}(\theta) = \tilde{L}_k(\bar{X}_n, \mathcal{U}(\theta), \theta_{s_k})$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

when observing its values against under the influence of noise ξ_n. Under the assumptions made, the minimum of the function $f_{s_k}(\theta)$ is reached when $\theta = \theta_{s_k}$. To solve the formulated problem, the method from [5] can be used.

The distributed optimization task can be formulated in terms of finding the method of constructing cluster (meso-) control, in which the same control action is applied to all elements of the cluster. In this case, the discretized loss function can be represented as a distributed functional

$$f_{s_k}(\theta) = \sum_{i=1}^{m_{s_k}} \tilde{l}_k(\bar{X}_n^i, \mathcal{U}^i(\theta), \theta_{s_k}).$$

This problem could be solved by appliying the stochastic optmimization type method from [15].

The considered problem setting is a particular case of more general problem of minimizing a non-stationary differentiable function with respect to θ. Let $\mathcal{F}_{n-1}$ be the σ-algebra of all probabilistic events which happened during time interval $n-1$ before start of time interval n. Hereinafter $\mathbb{E}_{\mathcal{F}_{n-1}}$ is a symbol of the conditional mathematical expectation with respect to the σ-algebra $\mathcal{F}_{n-1}$, $\mathbb{E}$ is a symbol of the mathematical expectation. The minimum point θ_{s_k} of function

$$F_n(\theta) = \mathbb{E}_{\mathcal{F}_{n-1}} f_{s_k}(\theta) \to \min_{\theta}$$

needs to be estimated.

More precisely, using the observations $y_1, y_2, \ldots, y_n$ and inputs $\theta_1, \theta_2, \ldots, \theta_n$, construct an estimate $\hat{\theta}_n$ of an unknown vector θ_{s_k} minimizing the time-varying mean-risk functional.

4.1. Assumptions

Let us formulate Assumptions about disturbances and functions $f_{s_k}(\theta), F_n(\theta)$.

1. For $n = 1, 2, \ldots$, the successive differences $\bar{\xi}_n = \xi_n^+ - \xi_n^-$ of observation noise are bounded: $|\bar{\xi}_n| \leq c_\xi < \infty$, or $\mathbb{E}\bar{\xi}_n^2 \leq c_\xi^2$ if a sequence $\{\xi_n\}$ is random, where ξ_n^+, ξ_n^- are observation noises occurred during the same time interval n but at different time instants.
2. Functions $F_n(\cdot)$ have unique minimum points θ_{s_k} and $\forall \theta \ \langle \theta - \theta_{s_k}, \mathbb{E}_{\mathcal{F}_{n-1}} \nabla f_{s_k}(\theta) \rangle \geq \mu \|\theta - \theta_{s_k}\|^2$ with a constant $\mu > 0$. Here and further $\langle \cdot, \cdot \rangle$ is a scalar product of two vectors.
3. The gradient ∇f_{s_k} is uniformly bounded in the mean-squared sense at the minimum points $\theta_t : E\|\nabla f_{s_k}(\theta_t)\|^2 \leq g^2$
4. $\forall s_k \in S$ the gradient $\nabla f_{s_k}(\theta)$ satisfies the Lipschitz condition: $\forall \theta', \theta''$
$$\|\nabla f_{s_k}(\theta') - \nabla f_{s_k}(\theta'')\| \leq M\|\theta' - \theta''\|$$
with a constant $M \geq \mu$.
5. $\forall n \geq 1$ random vector Δ_n does not depend on $\bar{w}_n$, random vectors $\bar{w}_n, \Delta_n$ do not depend on $\bar{w}_1, \ldots, \bar{w}_{n-1}$; if $\{\bar{v}_n\}$ are random variables, then $\bar{w}_n, \Delta_n$ also do not depend on $\bar{v}_1, \ldots, \bar{v}_n$.

Using available observations, it is necessary to construct a sequence of estimates $\hat{\theta}_n$ of the unknown vector θ_{s_k} that minimizes the function $f_{s_k}(\theta)$. To solve the problem, we will use an iterative algorithm with two measurements.

Let the trial simultaneous disturbance Δ_n, $n = 1, 2, \ldots$ be an observable (set or user-controlled) sequence of independent random vectors with known distribution functions $\mathrm{P}_n(\cdot)$ — and specified vector functions $\mathcal{K}_n(\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^1$, satisfy the conditions

$$\int \mathcal{K}_n(x)\mathrm{P}_n(dx) = 0; \int \mathcal{K}_n(x)x^{\mathrm{T}}\mathrm{P}_n(dx) = \mathrm{I}, \tag{9}$$

$$\sup_n \int \|\mathcal{K}_n(x)\|^2 \mathrm{P}_n(dx) < \infty, n = 1, 2, \ldots. \tag{10}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

4.2. Algorithm

Let's choose an arbitrary initial estimate vector $\hat{\theta}_0 \in \mathbb{R}^d$ and scalar parameters α, β for an iterative algorithm

$$\begin{cases} \theta_n^+ = \hat{\theta}_{n-1} + \beta\Delta_n, \; \theta_n^- = \hat{\theta}_{n-1} - \beta\Delta_n \\ y_n^+ = f_n(\theta_n^+) + \xi_n^+, \; y_n^- = f_n(\theta_n^-) + \xi_n^- \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \frac{\alpha}{2\beta}\mathcal{K}_n(\Delta_n)(y_n^+ - y_n^-). \end{cases} \tag{11}$$

4.3. Main Result

Let us introduce the following notation:

$$\nu = \alpha\left(0.5 - \mu - \frac{L\alpha C_\tau}{2}\right), \quad \phi = \alpha\gamma + \frac{L}{2}\alpha^2 C_\sigma, \quad \gamma = 0.5\,(C_1\beta)^2, \quad \psi = \frac{\phi}{\nu}$$

Here C_τ, C_σ could be chosen to satisfy inequation $\mathbb{E}\|\frac{1}{2\beta}(f_n(\theta+\beta\Delta_n) + \xi_n^+ - f_n(\theta-\beta\Delta_n) - \xi_n^-)\|^2 \leq C_\sigma + C_\tau\|\theta - \theta_{s_k}\|^2$ and $C_1 \geq \int \|\mathcal{K}_n(x)\|M\|x\|^2 \mathrm{P}_n(dx)$ and α chosen to satisfy following conditions:

$$\begin{aligned} & 0 \leq \alpha \leq \frac{4\mu - 2}{LC_\tau}, \quad \alpha \leq \frac{2\mu - 1 - \sqrt{(2\mu-1)^2 - 2LC_\tau}}{LC_\tau} \\ & \alpha \geq \frac{2\mu - 1 + \sqrt{(2\mu-1)^2 - 2LC_\tau}}{LC_\tau}. \end{aligned} \tag{12}$$

Theorem 1. *Let Assumptions 1–5 and conditions for kernels $\mathcal{K}$ (9)-(10) and α (12) be satisfied. Set $\hat{\theta}_0$, choose interval size parameter k*

$$\mathbb{E}\{\|\hat{\theta}_n - \theta_{s_k}\|^2\} \leq \mathbb{E}\{\|\hat{\theta}_0 - \theta_{s_k}\|^2\}(1-\nu_i)^n + \psi(1 - (1-\nu_i)^n). \tag{13}$$

Proof. To analyze the convergence of the algorithm (11) estimates, a method from [10] is used. Choose

$$V(\hat{\theta}_n) = \frac{1}{2}\|\hat{\theta}_n - \theta_{s_k}\|^2 \tag{14}$$

as Lyapunov function. To prove the theorem (13) is true it is sufficient to show the following six propositions are satisfied.

Proposition 1. *The iterative process Y_n, which defines the direction of the estimate change, is a Markov process, i.e. the distribution of the random vector Y_n depends only on $\hat{\theta}_n$ and n:*

$$Y_n = \frac{1}{2\beta}\mathcal{K}(\Delta_n)(y_n^+ - y_n^-) \tag{15}$$

For algorithm (11) we have

$$\mathbb{E}\{Y_n\} = \mathbb{E}\{\mathcal{K}_n(\Delta_n)\frac{y_n^+ - y_n^-}{2\beta}|\mathcal{F}_{n-1}\},$$

where the right hand side depends only on $\hat{\theta}_n$ and n in the sense that Δ_n does not depend on any other random variables. Proposition is true.

Proposition 2. $V(\hat{\theta}_n) \geq 0$, $\inf V(\hat{\theta}_n) = 0$, $V(\hat{\theta}_n)$ *has first-order derivative, and its gradient satisfies Lipshitz condition:*

$$\|\nabla V(x) - \nabla V(\theta)\| \leq L\|x - \theta\| \quad \forall x, \theta \in \mathbb{R}^d.$$

doi:10.20944/preprints202312.0504.v1

The proposition is valid due to the choice of Lyapunov function (14).

Proposition 3. *Pseudo-gradient condition:*

$$\langle \nabla V(\hat{\theta}_n), \mathbb{E}\{Y_n\}\rangle \geq \delta_n V(\hat{\theta}_n) - \gamma_n,\ \delta_n > 0,\ \gamma_n \geq 0. \tag{16}$$

At first consider $\mathbb{E}\{Y_n\}$. Due to (11) pseudo-gradient (15) after using Assumption 5 becomes

$$\mathbb{E}\{Y_n\} = \mathbb{E}\{\mathcal{K}_n(\Delta_n)\frac{1}{2\beta}\left(f_{s_k}(\theta_n^+) - f_{s_k}(\theta_n^-)\right)|\mathcal{F}_{n-1}\} \tag{17}$$

Consider the expression under mathematical expectation. Using Taylor series representation it could be written as:

$$\frac{1}{2}\nabla F_n(\hat{\theta}_n) + \frac{1}{2\beta}\int \mathcal{K}_n(x)x^T\int_0^1\left(\nabla_x f_{s_k}(\hat{\theta}_n + t\beta x) - \nabla f_{s_k}(\hat{\theta}_n)\right)dtP_n(dx) +$$
$$+\frac{1}{2}\nabla F_n(\hat{\theta}_n) - \frac{1}{2\beta}\int \mathcal{K}_n(x)x^T\int_0^1(\nabla f_{s_k}(\hat{\theta}_n - t\beta x) - \nabla f_{s_k}(\hat{\theta}_n))dtP_n(dx)$$

Estimate absolute value of the sum of integral elements in the obtained expression. After using (9), Assumption 4 we get

$$|\int(\cdot)P_n(dx)| + |\int(\cdot)P_n(dx)| \leq \frac{2\beta^2}{2\beta}\int \|\mathcal{K}_n(x)\|\|x\|M\|x\|P_n(dx) \leq C_1\beta. \tag{18}$$

Substitute the estimate, elements containing gradients of F_n and (14) into (16), regard the relation $\int(\cdot)P_n(dx) \geq -|\int(\cdot)P_n(dx)|$ and Cauchy–Bunyakovsky–Schwarz inequality:

$$\langle \hat{\theta}_n - \theta_{s_k}, \mathbb{E}\{Y_n\}\rangle \geq \langle \hat{\theta}_n - \theta_{s_k}, \nabla F_n(\hat{\theta}_n)\rangle - \|\hat{\theta}_n - \theta_{s_k}\|C_1\beta.$$

Apply Assumption 2 and estimate $\|\hat{\theta}_n - \theta_{s_k}\|C_1\beta \leq \frac{1}{2}(\|\hat{\theta}_n - \theta_{s_k}\|^2 + (C_1\beta)^2)$:

$$\langle \hat{\theta}_n - \theta_{s_k}, \mathbb{E}\{Y_n\}\rangle \geq (2\mu - 1)\frac{1}{2}\|\hat{\theta}_n - \theta_{s_k}\|^2 - \frac{1}{2}C_1^2\beta^2$$

Proposition is true for $\mu > 1/2$.

Proposition 4.

$$\mathbb{E}\{\|Y_n\|^2\} \leq \sigma_n^2 + \tau V(x),\ \sigma_n \geq 0,\ \tau_n \geq 0.$$

Using (10), Cauchy–Bunyakovsky–Schwarz inequality, Assumptions 1 and 3 it could be shown that

$$\mathbb{E}\{\|Y_n\|^2\} \leq \frac{1}{2}\sup_x \mathcal{K}_n(x)^2\mathbb{E}\{(\xi_n^+)^2 + (\xi_n^-)^2|\mathcal{F}_{n-1}\} +$$
$$+\int\left(f_{s_k}(\hat{\theta}_n^+) - f_{s_k}(\hat{\theta}_n^-)\right)^2\|\mathcal{K}_n(x)\|^2\mathrm{P}_n(dx) \leq C_2\beta^2(\|\hat{\theta}_{n-1} - \theta_{s_k}\|^2) + C_3\beta^4 + C_4\xi_n^2$$

Proposition holds true with $\tau = 2C_2\beta^2$, $\sigma_n^2 = C_3\beta^4 + C_4\xi_n^2$.

Proposition 5. $\mathbb{E}V(\hat{\theta}_0) < \infty$.

Proposition is valid due to arbitrariness of initial approximation $\hat{\theta}_0$ choice and an assumption regarding final order s_k of the external disturbance W affecting the system and thus the final order of the system state vector.

Proposition 6. $0 \leq \nu \leq 1; \sum_n \nu = \infty,\ n \to \infty$

The first inequality could be met by choice of α and the second one is true since ν is constant. Proposition is true.

The fulfillment of the given propositions allow to prove the theorem using the result in [10]. □

Remark 1. *After the algorithm converges the parameter estimates $\hat{\theta}_n$ continue to fluctuate around the true parameter value θ_{s_k} until the new change of system structure and value of θ_{s_k}.*

Remark 2. *The most common problem setup for unknown-but-bound disturbances in existing works is formulated as follows. It is required to minimize the value of the objective function $f(\theta)$ with some adversarial deterministic noise $\xi(\theta)$ such that $|\xi(\theta)| \leq \xi$ and $\xi > 0$):*

$$\tilde{f}(\theta) = f(\theta) + \xi(\theta).$$

The considered setup implies that external noise depends on system state s_k which is not a direct function of θ but rather is affected by values of θ_t during some time period $[t-d, t)$

$$\tilde{f}(\theta) = f(\theta) + \xi(s_k).$$

Value s_k is formed on basis of previous values of θ.

5. Application for Orientation Improvement of Radio Telescope's Elements

For astronomers, an urgent task was to obtain images of objects that can't be recorded using optical telescopes, which are situated on earth or in space. This problem was largely solved using radio telescopes [11]. The main tasks of the telescope are: to collect radiation that falls on the mirror system with minimal losses, and also obtain the most accurate image of the object [6]. There are various ways to solve this problem, consider for example, [8]. One of them is improving the quality of the device [12], which collects radiation for obtaining images. Another one is combining radio telescopes into systems [9]. If the radiation is collected with significant errors, then the image will have disturbances [27].

An important and time-consuming part of image acquisition is the precise tuning of the radio telescope antenna (or systems consisting of such antennas) [13]. The quality of the image obtained on a radio telescope directly depends on the quality of the construction of a reflecting system of mirrors that focuses the radiation coming from outside. To improve the image quality, it is necessary to focus the radiation of the device in such a way that it works as accurately as possible, especially if it is located in space [14]. Traditional antenna tuning algorithms are sufficient for the task. However, they lose their effectiveness under uncontrolled unpredictable external influences. We consider the case when these are deformations of the radio telescope shields that arise due to environmental influences such as temperature changes, wind and other influences.

In practice, radiation is subject to various distorting influences, and as a result the quality of the observed image decreases, despite the presence of various stabilizers and filters [11]. One of the ways to solve such a problem is using randomized stochastic optimization algorithms [15]. A method for improving image quality by improving the tuning characteristics of the radio telescope mirror system is considered in [25].

In an ideal antenna system, the signal is reflected from different mirror plates and assembled at one point. Deformations of radio telescope structures, external temperature, wind, mechanical influences lead to deviation of the optical path lengths of the rays from the required ones. As a result, the focus point on the plane of the receiver shifts [12]. To improve reflection accuracy on the surface of radio telescope mirrors, the following methods are used: autocollimation [22], telescope calibration by spectral density radiation flux, synchronous calibration method [21], laser geodetic measurements [6],

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

improvement of the kinematics of antenna elements [19], radio holography physical method [24]. The goal is to develop a stochastic optimization type algorithm to improve the quality of the settings of the radio telescope mirror system model which could be used in a system similar to Radioastron [20,26]. The main criterion for efficiency is the recording power of the desired signal and the time required for adjusting the parameters.

The antenna segments can be set to the optimal position to improve the quality of image recording. Consider an irradiator (radiation generator), a receiver and a mirror system of a radio telescope, consisting of identical plates that reflect the incoming signal ($i = 1, \dots, N, N = 895$ in the RATAN-600 installation). Radiation is created in the irradiator, which, falling on the plates and focused in the receiver. Let's divide the number axis into time intervals of duration δ, starting from some moment t_0, where T_k is the k-th time interval k is the index of the time interval. Assume that we know:
1) the position (orientation) of each i-th plate, which is specified by the vector of parameters $(a_i, b_i, c_i)^T$, where a_i is the rotation angle of the i-th reflective element horizontally; b_i is the vertical rotation angle, c_i is the forward horizontal displacement of the i-th reflecting element. Let θ_k be a vector that contains all parameters of the mirror system in a given time interval k, $\theta_k = (a_1, a_2, ..., a_N, b_1, b_2, ..., b_N, c_1, c_2, ..., c_N)^T$,
2) radiation coming from each mirror (z_i),
3) the common signal coming from all mirrors to the receiver $Z(\theta_k) = \sum_{i=1,\dots,N}(z_i)$,
4) characteristics of the signal in the generator.
The front of a signal is the sum of harmonics that comes to us with different phases from a certain direction. The perfect placement of the plates brings all radiation from the objects into focus. We obtain the signal as a sum of sines with phases. Different z_i arrive at different times with different phases. Let us evaluate the difference between the signals from an ideal antenna and from a real one (with deformations). Signals reflected from ideal mirror segments will have the same phase ϕ_i ($\phi_1 = \phi_2 = \dots = \phi_N$). The signals from segments with deformations will look like this:

$$z_1 = \sin(\omega t + \phi_1), z_2 = \sin(\omega t + \phi_2), \dots z_i = \sin(\omega t + \phi_i) \tag{19}$$

The objective function of the problem ($F(\theta)$ is the signal power) is defined as follows:

$$F(\theta) = \overline{\lim_k} \int_{t \in T_k} |Z(\theta, t)|^2 dt. \tag{20}$$

It is required to maximize the objective function:

$$F(\theta) \to \max_\theta \tag{21}$$

We consider the problem of optimizing the position of mirrors in the limit over time (not over a specific time interval).

The reflective elements of the antenna are made exactly the same, so they provide equivalent observations in all directions.

At the same time, if one moves along an ideal reflective surface, its local characteristics change. Therefore, a real reflective surface composed of identical elements will repeat deviations from the ideal surface from element to element. These deviations will be greater, if the shape of the surface of the element differs from the shape of that portion of the ideal surface which this element should represent. If the size of the reflective elements increases, then the deviations naturally increase. These deviations are an error distributed over the reflector of a variable profile antenna, which, at large values, creates unacceptable distortions reducing the efficiency of the antenna. We will call v_k the signal power measurement errors arising due to unknown and uncontrolled deformations in the reflective elements of the antenna caused by weather, wind, and temperature changes.

After conventional procedures for adjusting the inclination angles and positions of the radio telescope mirrors, we obtain an initial approximation $\hat{\theta}_0$ to the optimal tuning values. Then there

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 December 2023 doi:10.20944/preprints202312.0504.v1

is still the possibility of "tuning" in a certain neighborhood T containing $\hat{\theta}_0$ and the optimal value $\theta^\star$ corresponding to the maximum power of the received signal. We use a stochastic optimization algorithm with two measurements per iteration, which allows us to reduce the negative impact of various disturbances on the power of the recorded signal [7]:
Select θ_0.
$n-1 \to n$
We sequentially generate vectors Δ_n with components from +/-1, which are chosen with equal probability.
We measure the power values for two positions of the antenna system:
$(\theta_n + \beta\Delta_n)$ and $(\theta_n - \beta\Delta_n)$.
Measurements are obtained with noise v_{2n} and v_{2n-1} :

$$y_{2n} = P_{2n}(\theta_{2n} + \beta\Delta_n) + \xi_{2n}; \tag{22}$$

$$y_{2n-1} = P_{2n-1}(\theta_{2n-1} - \beta\Delta_n) + \xi_{2n-1}; \tag{23}$$

Next, we form the following estimate θ_{2n} according to the rule:

$$\theta_{2n} = \mathcal{P}_{\mathcal{T}}(\theta_{2n-1} + \frac{\alpha K_n(\Delta_n)}{2\beta}(y_{2n} - y_{2n-1})), \tag{24}$$

where α, β are the parameters of the algorithm, $\mathcal{P}_{\mathcal{T}}$ is the projection onto the set $\mathcal{T}$.

To construct the kernels $K_0(\cdot)$ and $K_1(\cdot)$ on the interval $[-1/2, 1/2]$ orthogonal Legendre polynomials could be used. In this case, for initial values $\ell = 1, 2 \quad$ (i.e. $2 \leq \gamma \leq 3$) the type of kernels is as follows:
$K_0(q) = 12q, \quad K_1(q) = 1, \quad |\, q \,| \leq 1/2,$
for $\ell = 3, 4 \quad$ (i.e. $3 < \gamma \leq 5$):
$K_0(q) = 5q(15 - 84q^2),$
$K_1(q) = 9/4 - 15q^2, \quad |\, q \,| \leq 1/2,$
and for $|q| > 1/2$ both functions are equal to zero.

The test disturbance is formed in such a way that $\forall n \geq 1$ random vector Δ_n does not depend on $\bar{v}_1, \ldots, \bar{v}_k$ and $\mathrm{E}\{(v_{2n} - v_{2n-1})^2/2\} \leq \sigma_2^2$, $(\mathrm{E}\{v_n^2\} \leq \sigma_1^2)$;

The main element that reflects the incoming signal is the segment of the mirror system. It is important to configure these shields so that the signal comes into focus. The system allows to customize the required dimensions of the reflective shield and its curvature. Stochastic optimisation algorithm (11) can be applied for tuning a system which consists of a large number of telescopes in space under conditions of interference and also when individual elements of the system are deformed. About 3000 parameters of the mirror surfaces are adjusted, during focusing of the mirror system.The faster setting up of the mirrors is very important for the accuracy of observations. For acceleration, it is proposed to use the Nesterov acceleration method in distributed form [16–18]. In Figure 3 on page 11, the dependence of the signal power on the number of iterations of the algorithm is given.

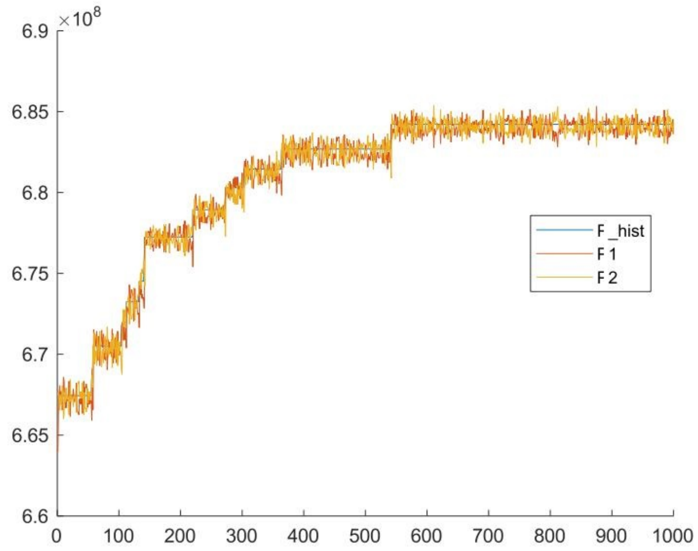

Figure 3. Convergence of the algorithm.

Author Contributions: Conceptualization, O.G.; methodology, O.G.; software, K.D.; formal analysis, Y.I.; writing—original draft preparation, K.D. and Y.I.; writing—review and editing, O.G, and Y.I.; visualization, K.D.; All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the IPME RAS by Russian Science Foundation (project no. 21-19-00516).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Granichin, O.; Uzhva, D.; Volkovich, Z. Cluster Flows and Multiagent Technology. Mathematics 2021, 9, 22. https://doi.org/10.3390/math9010022
2. Granichin, O. What is an Actual Structure of Complex Information-Control Systems? Stochastic optimization in informatics. 2016. Is. 12, No. 1
3. Parilina, E.; Petrosyan, L. On a Simplified Method of Defining Characteristic Function in Stochastic Games. Mathematics 2020, 8, 1135. https://doi.org/10.3390/math8071135
4. Achdou, Y., Cardaliaguet, P., Delarue, F., Porretta, A. and Santambrogio, F., 2021. Mean Field Games: Cetraro, Italy 2019 (Vol. 2281). Springer Nature.
5. Granichin, O. and Amelina, N., 2014. Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances. IEEE Transactions on Automatic Control, 60(6), pp.1653-1658.
6. Ermakov A. N., Kovalev Yu. A. Project "RadioAstron". Calibration of a space telescope in flight - automation of measurement processing of the ASC FIAN, Moscow, Russia, Proceedings of the Institute of Applied Astronomy of the Russian Academy of Sciences, vol. 54, 2020
7. Granichin O.N., Polyak B.T. Randomized estimation and optimization algorithms under almost arbitrary noise. M.: Nauka, 2003. 291 p.
8. Droszcz, A., Jedrzejewski, K., Kłos, J., Kulpa, K., Pozoga M. Beamforming of LOFAR Radio-Telescope for Passive Radiolocation Purposes. Remote Sens.2021, 13, 810. https://doi.org/ 10.3390/rs13040810
9. Leonid I. Gurvits. Advances in Space Research Volume 65, Issue 2, 15 January 2020, Pages 868-876 Space VLBI: from first ideas to operational missions Author links open overlay panel
10. Polyak, B.T. Convergence and rate of convergence in iterative stochastic processes. I. The general case. 1976. Avtomatika i telemekhanika, No. 12, pp.83-94.
11. Kardashev N.S. and others. "RADIOASTRON": results of the implementation of the scientific research program for 5 years of flight // Bulletin of NPO im. S.A. Lavochkina. 2016. No. 3. P. 4-24
12. Dubarenko V.V., Kuchmin A.Yu., Artemenko Yu.N., Shishlakov V.F. Millimeter-wave radio telescopes with adjustable mirror surfaces: monograph. – St. Petersburg: GUAP, 2019.

13. Monakhova U.V., Ivanov D.S. Formation of a swarm of nanosatellites using decentralized aerodynamic control taking into account communication restrictions // Preprints of the Institute for Problems of Materials Science. M.V.Keldysh. 2018. No. 151. 32 p. doi:10.20948/prepr-2018-151
14. Bentum M.J., Verma M.K., Rajan R.T., Boonstra A.J., Verhoeven C.J.M., Gill E.K.A., van der Veen A.J., Falcke H., Klein Wolt M., Monna B., Engelen S., Rotteveel J., Gurvits L.I. A Roadmap towards a Space-based Radio Telescope for Ultra-Low Frequency Radio Astronomy. /doi.org/10.48550/arXiv.1909.08951
15. Granichin O.N., Erofeeva V.A., Ivanskiy Y.V., Jiang Y. Simultaneous Perturbation Stochastic Approximation-Based Consensus for Tracking Under Unknown-But-Bounded Disturbances. IEEE Transactions on Automatic Control 2021, Vol. 66, Is. 8 PP. 3710–3717
16. Rogozin A., Yarmoshik D., Kopylova K. Gasnikov A. Decentralized Strongly-Convex Optimization with Affine Constraints: Primal and Dual Approaches 2022
17. Nesterov Y. A method of solving a convex programming problem with convergence rate o(1/k 2). Soviet Mathematics Doklady, 27(2):372–376, 1983.
18. Nesterov Y. Introductory Lectures on Convex Optimization: a basic course. Kluwer Academic Publishers, Massachusetts, 2004.
19. Zharov V.I., Sotnikova Yu.V. Methodology for determining the kinematic characteristics of the elements of the main mirror of the RATAN-600 radio telescope using modern laser measuring systems. Astrophysical Bulletin, 2017. Vol. 72, No. 4, pp. 520–526.
20. Moisheev A.A. Creation of space segments of astrophysical observatories. Bulletin 2.
21. Sotnikova Yu.V., Kovalev Yu.A., Erkenov A.K., Method of synchronous calibration of RATAN-600 using 2 of its sectors. Astrophysical Bulletin, 2019. Vol. 74, No. 4, pp. 535–543.
22. Khaikin V.B., Bursov N.N. Autocollimation automatic adjustment and control of the efficiency of elements of the RATAN-600 radio telescope. Journal of Radioelectronics, 2016, pp. 1684-1719.
23. Mingaliev M.G. RATAN-600 - current state and prospects. Abstracts of the report. Conf. RT-2002, in Pushchino, 2002. P. 80.
24. Khaikin V.B., Lebedev M.K., Ripak A.M. A method for radio-holographic monitoring of the surface of the main mirror of the RATAN-600 radio telescope with radial movement of the support element. Journal of Radioelectronics, 2016, pp. 1684-1719.
25. Kopylova K.D., Granichin O.N. Minimizing the systematic error of a radio astronomy telescope using a randomized stochastic optimization algorithm. Proceedings of the 13th multi-conference on management issues, 2020.
26. Kovalev Yu.A., Sotnikova Yu.V., Erkenov A.K., Popkov A.V., Volvach L. N., Vasilkov V. I., Lisakov M. M., Semenova T. A., Tsybulev P. G. Features of calibration of the space radio telescope "RadioAstron" and the radio telescope RATAN-600. // Proceedings of the IPA RAS. – St. Petersburg: IPA RAS, 2018. Issue. 47.- pp. 38-42.
27. Minchenko B.S. Synthesis of radio images on the RATAN-600 radio telescope Izv. universities Radiophysics. 1983. T. 26, No. 11. P. 1463–1471.