Robust Exact Differentiation via Sliding Mode Technique*

ARIE LEVANT†

Key Words—Differentiators; measurement noise; nonlinear control; sliding mode; robustness.

Abstract—The main problem in differentiator design is to combine differentiation exactness with robustness in respect to possible measurement errors and input noise. The proposed differentiator provides for proportionality of the maximal differentiation error to the square root of the maximal deviation of the measured input signal from the base signal. Such an order of the differentiation error is shown to be the best possible one when the only information known on the base signal is an upper bound for Lipschitz's constant of the derivative. © 1998 Elsevier Science Ltd. All rights reserved.

Introduction

Differentiation of signals given in real time is an old and well-known problem. Construction of a special differentiator may often be avoided. For example, if the signal satisfies a certain differential equation or is an output of some known dynamic system, the derivative of the given signal may be calculated as a derivative with respect to some known dynamic system. Thus, the problem is reduced to the well-known observation and filtration problems. In other cases, construction of a differentiator is inevitable. However, the ideal differentiator could not be realized. Indeed, together with the basic signal it would also have to differentiate any small high-frequency noise which always exists and may have a large derivative.

The main approach to linear differentiator construction is to approximate the transfer function of the ideal differentiator on a definite signal frequency band (Pei and Shyu, 1989; Kumar and Roy, 1988; Rabineer and Steiniglitz, 1970). The frequency band of the noise being known, low-pass filters are used to damp noises. Stochastic features of the signal and the noise may also be considered (Carlsson et al., 1991). In the latter case, the stochastic models of both the signal and the noise are presumed to be known. Linear observers (Luenberger, 1971) may be used. In any case, a linear differentiator with constant coefficients may provide for asymptotically exact differentiation for a rather than class of inputs and does not calculate exact derivatives of other noise-free signals.

If nothing is known on the structure of the signal except some differential inequalities, then sliding modes (U'kin, 1992) are used. In the absence of noise the exact information on the signal derivative may be obtained by averaging high-frequency switching signals. Also, sliding observers (Slotine et al., 1987) or observers with large gains (Nicosia et al., 1991) are successfully employed. However, in all these cases the exact differentiation is provided only when some differentiator parameters tend to inadmissible values (like infinity). Thus, here too the resulting differentiator cannot calculate exact derivatives of noise-free signals.

* Received 24 June 1996; revised 29 April 1997; received in final form 5 November 1997. This paper was recommended for publication in final form by Associate Editor Fassan Khalil under the direction of Editor Tamer Basar. Corresponding author: Dr. A. Levant. Tel. 972 7 6231313; Fax 972 7 6231211; E-mail: levant@bgu mail.bgu.ac.il.
† Institute for Industrial Mathematics, 22 Ha-Histadrut St., Beer-Sheva 84213, Israel.

The performance of the known differentiators follows the following principle: only approximate differentiation is provided in the absence of noise, at the same time the differentiator is insensitive to any high-frequency signal components considered to be noises. Thus, differentiation is robust but not exact, the error does not tend to zero in the presence of vanishing noise at any fixed time, and no asymptotic error analysis is sensible for any fixed differentiator parameters and time. Another principle employed here combines exact differentiation (with finite transient time) for a large class of inputs with robustness in respect to any small noises. A known approach (Golembio et al., 1976) is chosen: high-quality tracking of $f(t)$ by $x(t)$, $x = u$, having been provided, control $u(t)$ may be used for evaluation of $f(t)$. The new result is attained here due to application of a 2-sliding algorithm (Levantovski, 1985; Emelyanov et al., 1986; Levantovski, 1993; Fridman and Levant, 1996) which forms continuous control $u(t)$ providing for keeping the equalities $\phi = x - f(t) = 0$, $\phi = u - f(t)$ = 0 after a finite-time transient process. The purpose of this paper is

- to clear some inherent restrictions on exact robust differentiation and its error asymptotics;
- to propose a robust first-order differentiator that is exact on signals with a given upper bound for Lipschitz's constant of the derivative;
- to ensure the best-possible error asymptotic order when the input noise is a measurable (Lebesgue) bounded function of time.

Robust exact differentiation limitations

Let input signals belong to the space $M[a, b]$ of measurable functions bounded on a segment $[a, b]$ and let $|f| = sup |f(t)|$. Define abstract differentiator as a map associating an output signal with any input signal. A differentiator is called exact on some input if the output coincides with its derivative. The differentiator order is the order of the derivative which it produces. Differentiator $D$ is called robust on some input $f(t)$ if the output tends uniformly to $Df(t)$ while the input signal tends uniformly to $f(t)$. A differentiator is called correct on some input if it is exact and robust on it. The ideal differentiator cannot be considered as an abstract one, for it does not operate for nondifferentiable inputs.

Being exact on two inputs, any differentiator will actually differentiate a difference between these inputs which may be considered as a noise. Thus, differentiator design is a simple trade-off: the denser the exactness class in $M[a, b]$, the more sensitive will the differentiator be to small noises. For example, being correct on a thin class of constant inputs, the differentiator producing identical zero is totally insensitive to noises. It is easily seen that the correctness set of any abstract differentiator cannot be locally dense in the set of continuous functions, otherwise small noises with large derivatives would be exactly differentiated, which contradicts robustness. In particular, no differentiator is correct on all smooth functions or on all polynomials.

Let $W_0, \ldots, n$ be the set of all input signals whose $n - 1^{th}$ derivatives have Lipschitz's constant $C > 0$. The statements below are valid for any sufficiently small $\varepsilon > 0$ and noises not exceeding $\varepsilon$ in absolute value.
Proposition 1. No exact on \( W(C, n), n > 0 \), differentiator of the order \( i \leq n \) may provide for accuracy better than \( C^3 e^{-\alpha n} \).

Proof. Consider noise \( v(t) = \varepsilon \sin(C|\varepsilon|^{1/2} \tau) \). It is easy to check that \( \sup_j |v(t)| = C^{1/2} e^{-n^{1/2} \tau} \) for \( i = 0, 1, \ldots, n \) and small \( \varepsilon \). The differentiator has to differentiate properly both \( 0 \) and \( 0 + v(t) \).

Proposition 2. There are such constants \( \gamma_i(n) \geq 1, n = 1, 2, \ldots, i = 0, 1, \ldots, n \), \( \gamma_0(n) = 1, \gamma_1(n) = 2 \), that for any \( i \) there exists an \( i \)-th order differentiator \( D \), correct on any \( f \in W(C, n) \), which provides for the accuracy \( \|Df + f\| \leq \gamma_i(n) C^3 e^{-\alpha n} \) for any noise \( v \) with \( \|v\| < \xi \).

Proof. Let \( \Lambda \) be a map associating with any \( g \in M[a, b] \) some function \( \cphi \) closest to \( g \) in \( W(C, n) \). Such a function exists, for any bounded subset of \( W(C, n) \) is precompact, but it is certainly not unique. Define \( \cD \) as \( \langle \Delta \xi \rangle \). The proposition is now a simple consequence of Lemma 1.

Lemma 1. There are such constants \( \beta_i(n) \geq 1, i = 0, 1, \ldots, n \), \( \beta_0(n) = 1 \), that the following inequalities hold:
\[
C^3 e^{-\alpha n} < \sup_{g \in W(C,n)} \sup_{f \in W(C,n)} \|f(t_0)\| \leq \beta_0(n) C^3 e^{-\alpha n}.
\]

Here and below all proofs are in the appendices. These are some evaluations of \( \beta_0(n) \): \( \beta_0(2) = \sqrt{\frac{2}{\pi}}, \beta_0(3) = 7.07, \beta_0(3) = 6.24 \).

Remarks.
- With \( C = 0 \) \( W(C, n) \) is the set of polynomials of degree \( n - 1 \). In that case Proposition 1 is trivial, as are Proposition 2 and Lemma 1. The restrictions discussed here do not preclude possibilities to receive a better differentiation accuracy, provided another exactness set \( s \) is considered. For instance, it may be shown that the differentiator suggested in the proof of Proposition 2 provides for the accuracy linear on \( \varepsilon \) in the case \( C = 0 \). The corresponding gain tends to infinity when \( \varepsilon \) or the segment length tends to zero. \( W(0, n) \) is too thin for reasonable, such differentiators are suitable for local usage only.
- All the above reasons are also true in the case when the differentiators are allowed to have transient time uniformly bounded by a constant less than the segment length.
- The above results also hold for infinit time intervals, the inputs being measurable locally bounded functions and \( e \) being any positive number. Only a minor change in the proof of Proposition 2 is needed.

Practical first-order robust exact differentiator

The above abstract differentiators were not intended for realization. Consider now a practical real-time differentiation problem. Let input signal \( f(t) \) be a measurable locally bounded function defined on \([0, \infty)\) and let it consist of a base signal having a derivative with Lipschitz's constant \( C > 0 \) and a noise. In order to differentiate the unknown base signal, consider the auxiliary equation
\[
\dot{x} = u.
\]

Applying a modified 2-sliding algorithm (Levant, 1993) to keep \( x - f(t) = 0 \), obtain
\[
\begin{align*}
\dot{u} & = u_1 - \lambda |x - f(t)|^{1/2} \text{sign}(x - f(t)), \\
\dot{u} & = -x \text{sign}(x - f(t)),
\end{align*}
\]

where \( u, \lambda > 0 \). Here \( u(t) \) is the output of the differentiator. Solutions of system 3.4 are unbounded in the Filippov sense (Filippov, 1985).

Define a function \( \Psi(\tau, \xi, C) = |\Psi(\tau, \xi)| \), where \( \xi(t) \), \( \Psi(t) \) is the solution of
\[
\begin{align*}
\dot{\xi} & = -|\xi|^{1/2} + \Psi, \\
\dot{\Psi} & = \frac{1}{2} (\xi - C) + |\xi|^{1/2} + \Psi > 0, \\
2\dot{C} & = 0 \quad \Psi(0) = 1,
\end{align*}
\]

where \( \tau > C, \lambda > 0 \), and \( \tau_0 = \inf\{t > 0 : \xi(t) = 0, \Psi(t) < 0\} \). It is easy to check that \( \tau_0 < \infty \). In practice, \( \Psi(\lambda, C) \) is to be calculated by computer simulation.

Theorem 1. (Convergence criterion). Let \( \lambda > 0, \lambda > 0, \Phi(\lambda, C) < 1 \). Then, provided \( f(t) \) has a derivative with Lipschitz's constant \( \xi(f \in W(C, 2)) \), the equality \( u(t) = f(t) \) is fulfilled identically after a finite-time transient process. There is no convergence of \( u(t) \) to \( f(t) \) for some \( f \in W(C, 2) \) if \( \Phi(\lambda, C) > 1 \).

The less \( \Phi(\lambda, C) \), the faster the convergence. \( \Phi \) is obviously the same for all \( (\lambda, C) \), with \( \lambda = \mu_C, \lambda = \mu_C \), where \( \mu_C > 1 \), \( \mu_C > 0 \) are some constants. Also, any increase of \( \lambda \) decreases \( \Phi \).

Following are the sufficient conditions for the convergence of \( u(t) \) to \( f(t) \):
\[
\begin{align*}
\lambda > C, \\
\tau^2 > 4C^2 + C.
\end{align*}
\]

Condition (6) results from a very crude estimation. Calculation of \( \Phi \) shows that many other values, e.g. \( \lambda > C = 12, \lambda = 1.1C \left( \Phi \approx 0.986 \right) \), or \( \lambda > 0.5C^{1/2}, \tau > 4C^2 + C \left( \Phi \approx 0.736 \right) \), may also be taken.

The conditions \( \lambda > C, \lambda > 0, \Phi < 1 \) are assumed to be satisfied in the following theorems.

Theorem 2. Let input signal be presented in the form \( f(t) = f_0(t) + v(t) \), where \( f_0(t) \) is a differentiable base signal, \( f_0(t) \) has a derivative with Lipschitz's constant \( C > 0 \), and \( v(t) \) is a noise, \( |v(t)| \leq \xi \). Then there exists such a constant \( B > 0 \) dependent on \( \lambda = C^{1/2} \) and \( \lambda = C \) such that after a finite time the inequality \( |u(t) - f(t)| < ab \|\xi\|^{1/2} \) holds.

The inequality \( \Psi(t) = f(t) - f(t) < ab \|\xi\|^{1/2} \) is to hold after a finite-time transient process. Here \( a > 0 \) is some constant dependent on \( \lambda = C^{1/2} \) and \( \lambda = C \).

Let \( x, u, t \) be measured at discrete times with time interval \( \tau \), and let \( t_{n+1}, t_n \) be successively measured times and the current time, \( t = [t_n, t_{n+1}] \). As a result, achieve the following modified algorithm:
\[
\begin{align*}
\dot{\xi} & = u, \\
u & = u(t) - \lambda |x(t) - f(t)|^{1/2} \text{sign}(x(t) - f(t)), \\
\dot{u} & = -x \text{sign}(x(t) - f(t)).
\end{align*}
\]

Theorem 3. Provided \( f(t) \) has a derivative with Lipschitz's constant \( C > 0 \), algorithm (7) (8) enables the inequality \( \Phi(t) = f(t) < a b \|\xi\|^{1/2} \) to hold after a finite-time transient process. Here \( a > 0 \) is some constant dependent on \( \lambda = C^{1/2} \) and \( \lambda = C \).

If the discrete measurements are carried out with some small measurement noise of magnitude \( \xi \) and \( \tau < C^{1/2} \), then an infinitesimally of a higher order has to be added to the right-hand side of the inequality in the statement of Theorem 2 (or Theorem 3). It may be shown that in the general case at least the accuracy of the order \( \varepsilon + \varepsilon^2 \) is provided (very crude estimation).

The transient process time is uniformly bounded if the initial deviations \( |x(t) - f(t)| \) and \( |w(t) - f(t)| \) are bounded. This may be attained by any inexact preliminary evaluation of \( f(t) \) and \( f(t) \). In such a case the transient time may be arbitrarily shortened.

Computer simulation

It took that \( \tau_0 = 0 \), initial values of the internal variable \( x(t) \) and the measured input signal \( f(t) \) coincided, initial value of the output signal \( u(t) \) is zero. The simulation was carried out by the Euler method with measurement and integration steps equaling \( 10^{-4} \).

Compare the proposed differentiator (3), (4) with a simple linear differentiator described by the transfer function \( (p_1 t + 1)^2 \). Such a differentiator is actually a combination of the ideal differentiator and a low-pass filter. Let \( \lambda = 8, \lambda = 6 \). Checking the convergence criterion, achieve \( \Phi = 0.064 \).
for \( C = 2 \) and \( \Phi = 0.20 \) for \( C = 7 \). The output signals for inputs \( f(t) = \sin t + 5t \), \( f(t) = \sin t + 5t + 0.0 \cos 10t \), and \( f(t) = \sin t + 5t + 0.001 \cos 30t \) and ideal derivatives \( \dot{f}(t) \) are shown in Fig. 1. The linear differentiator is seen not to differentiate exactly. At the same time it is highly insensitive to any signals with frequency above 30. The proposed differentiator handles properly any input signal \( f \) with \( \dot{f} \leq 7 \) regardless of the signal spectrum.

Let \( x = 2.2 \), \( \lambda = 2 \). Checking the convergence criterion achieves \( \Phi = 0.596 \) for \( C = 2 \). The output for the input base signal \( f(t) = \sin t + 5t + 1 \) in the presence of a measurement high-frequency noise with magnitude 0.04 is shown in Fig. 2a. Simulation shows the differentiation accuracy to take on successively the values 0.456, 0.213, 0.106, 0.053, while the noise magnitude takes on the values 0.04, 0.01, 0.0025, 6.25 \times 10^{-4}, respectively. Thus, the differentiation error is proportional to the square root of the noise magnitude with the gain 2.13 (the same gain is 5.9 for \( x = 8 \), \( \lambda = 6 \)). With precise measurements the accuracy \( 1.2 \times 10^{-3} \) was achieved. Remind that the gain is not less than 1 and in the absence of noise the accuracy is proportional to the measurement time step (Proposition 1, Theorem 3).

The output signal of the differentiator consists, essentially, of the accurate derivative and some small high-frequency noise. Therefore, implementation of a smoothing element may be practically useful if significant noise is assumed. Output \( u_{\text{sm}} \) of the simple smoothing element \( p/(0.05p + 1) \) for the base signal \( f(t) = \sin t + 5t + 1 \) and noise magnitude 0.04 is given in Fig. 2b (the accuracy achieved is 0.054). However, the ideal differentiation ability in the absence of noises is inevitably lost in this case (the accuracy sup\( |f - u_{\text{sm}}| = 0.05 \) is achieved). Such a differentiator is insensitive to high-frequency components of the input signal.

**Conclusions**

Inherent restrictions on exact robust differentiation and its error asymptotics were found. The existence of an arbitrary-order

![Fig. 1. Comparison of the proposed differentiator with a linear one.](image_url)
robust differentiator with the optimal order of error asymptotics
was established.
A first-order robust exact differentiator was proposed providing
for maximal derivative error to be proportional to the square
root of the input noise magnitude after a finite-time transient
process. This asymptotics order was shown to be the best attain-
able in the case when the only restriction on the input signal is
that Lipschitz's constant of its derivative is bounded by a given
constant and the noise is a measurable bounded function of
time.
Discrete measurements with a small time step were shown to
not be destructive to the differentiator features. In the absence of
noises the differentiation error is proportional to the measure-
ment step.
The differentiator considered features a simple form and an
eyasy design. It may be employed both in real-time control
systems and for numeric differentiation. Its use is preferable
in high precision systems with small noises. In the presence of
considerable measurement noises a simple smoothing element
may be implemented. However, this leads to loss of ideal differ-
entiation in the absence of noises.
The differentiator allows, obviously, successive implementa-
tion for higher-order exact robust differentiation. However, the
optimal error asymptotics will not be attained in that case.

References
tiation based on stochastic signal models. IEEE Trans. Signal
Higher order sliding modes in the binary control systems. Sov.

Appendix: proofs
Proof of Lemma 1. The first inequality follows from the proof of
Proposition 1. The Lagrange Theorem implies that there are
constants $L_T$ satisfying the property that for any $k$-smooth
function $f(t), |f''(t)| \leq \epsilon$, and any $T > 0$ there is a point $t_*$ on
any segment of the length $L_T$ where $|f''(t_*)| \leq 2\epsilon / T^2$. Obviously,
$L_0 = 0, L_1 = 2L_{k - 1} + 1$.

Denote $S_k(c, C, n) = \sup_{x \in [c, C]} \sup_{\{0 \leq l \leq k\}} \sup_{[0,1]} |f^{(l)}(t)|$. Using
identities $S_k(c, C, n) = C, S_k(c, C, n) = n, \epsilon = \text{the inequality } S_k(c, C, n) \leq
S_{k-1}(c, C, n), n - 1, \text{ and minimizing with respect to } T$, the
right-hand side of the inequality $|f''(t)| \leq 2\epsilon / T^2$ is
$T^{k-1} + CL_{k-1} - T$, obtain a recursive-on-$n$ definition of $\beta(n)$

$\beta(0) = 1, \beta(n) = \beta(n-1)\beta_{n-1}(\epsilon)^{n-1}, i < n,$

$\beta_{n-1} = 2^{n-1}\epsilon/2^{n-1} \left( \frac{1}{n-1} \right)^{n-1} + (n-1)^{-1} \epsilon^n \right]. \quad \Box$

The plan of the theorems proofs is as follows. Using the inclusion $f \in [-C, C]$, true almost everywhere, system (3), (4) is
replaced by an autonomous differential inclusion on plane $\sigma, t
\sigma = -f(\t).$ Intersections with axis $\delta$ having been studied, the
paths are shown to make an infinite number of rotations around
the origin converging to it in finite time. With a small measure-
ment step and in the presence of small noises a small attraction
set appears around the origin. After changing coordinates a
linear transformation preserving the paths is found, which allows to evaluate the attraction set asymptotics

Remind that according to the definition by Filippov (1988) any differential equation \( \sigma = f(\sigma) \), where \( \sigma \in \mathbb{R}^q \) and \( f \) is a locally bounded measurable vector function, is replaced by an equivalent differential inclusion \( \sigma \in F(\sigma) \). In the simplest case, when \( f \) is continuous almost everywhere, \( F(\sigma) \) is the convex closure of the set of all possible limits of \( f(\sigma) \) as \( \sigma \to z \), while \( \{\sigma \} \) are continuity points of \( f \). Any solution of the equation is defined as an absolutely continuous function \( \tau(\sigma) \), satisfying \( \sigma(t) \) at differential inclusion almost everywhere. Also, any differential inclusion is to be similarly replaced by a special one.

**Proof of Theorem 1.** Let \( \sigma = x - f(t) \). By calculating achieve

\[
\sigma = -f' - \frac{1}{2} \delta \sigma |\sigma|^{-1/2} - 2 \text{sign} \sigma.
\]

(9)

It does not matter that \( f' \) exists almost everywhere, but not at any \( t \), for solutions are understood in the Filippov sense. Strictly speaking, equation (9) is valid only for \( \sigma \neq 0 \). Nevertheless, as follows from (Filippov, 1988), solutions of system (3), (4) and, therefore, also \( \sigma(t) = \sigma_0 = f(t) \) and \( \delta(t) = \delta_0(\sigma(t)) \) are well defined for any initial conditions \( (\sigma_0, \delta_0) \) and any \( t > 0 \).

Remind that \( |f| \leq C \). Denote \( R = |x| + |\text{sign} \sigma| \). Then \( R \in [x - C, x + C] \) and

\[
\sigma = -\frac{1}{2} \delta \sigma |\sigma|^{-1/2} - R \text{sign} \sigma.
\]

(10)

Any solution of equation (9) has to satisfy the following differential inclusion understood in the Filippov sense:

\[
\sigma \in -\frac{1}{2} \delta \sigma |\sigma|^{-1/2} - [x - C, x + C] \text{sign} \sigma.
\]

(11)

The operations on sets are naturally understood here as sets of operation results for all possible combinations of the operand set elements. Consider the trajectory of inclusion (11) on plane \( \sigma \). Let \( \sigma = 0, \delta = \delta_0 > 0 \) at the initial moment (Fig. 3). Any trajectory \( \Gamma \) of (11) lies between trajectories \( \Gamma \) of (10) given, respectively, by

\[
R = \begin{cases} x - C, \quad \sigma > 0, \\ x + C, \quad \sigma < 0, \\ \sigma = 0, \quad \sigma = 0, \end{cases} \quad R = \begin{cases} x + C, \quad \sigma > 0, \\ x - C, \quad \sigma < 0, \end{cases}
\]

Extending trajectory \( \Gamma \), achieve sequence \( \delta_0, \delta_1, \ldots \) of successive intersections \( \Gamma \) with axis \( \sigma \). It follows from Lemma 2 below that \( |\delta_1 - \delta_0| |d_1| / |d_0| < \text{const} < 1 \). Hence, \( \delta_i \) converge to \( \sigma \) as a geometric progression. Also, calculating trajectory \( \Gamma \), achieve that (6) implies \( \delta(\sigma, \tau) \leq \text{const} < 1 \). The last point to be checked here is the finite time convergence of the algorithm.

The state coordinates of system (3), (4) are \( x \) and \( u_1 \). Consider new coordinates \( \sigma = x - f(t) \) and \( \xi = u_1 - f(t) \). In these coordinates system (3), (4) takes on the form

\[
\sigma = -\frac{1}{2} \delta \sigma |\sigma|^{-1/2} \text{sign} \sigma + \xi,
\]

(12)

and the corresponding differential inclusion is

\[
\xi = [x - C, x + C] \text{sign} \sigma.
\]

(13)

As follows from equation (12), \( \xi = \sigma \) when \( \sigma = 0 \). Hence, as follows from (13), the convergence time is estimated by the inequality

\[
T \leq \frac{2}{C} \sum_{i=0}^{1} |d_i|.
\]

**Lemma 2.** The ratio of successive intersections of trajectory \( \Gamma \) with axis \( \sigma \) is constant and coincides with the value of function \( \Phi(\sigma, x, C) \).

**Proof.** Let \( G_{\sigma, \eta} > 0 \) be an operator constituted by a combination of the linear coordinate transformation \( g_\xi : (\sigma, \xi) \rightarrow (\eta^1, \eta^2) \) and the time transformation \( t \rightarrow t \). It transfers any vector \( \epsilon \) from the tangential space at the point \( \zeta = (\sigma_0, \xi_0) \) into the vector \( \eta = [d_{\sigma_0} / d_{\xi_0}] \epsilon \) at the point \( \sigma_0, \xi_0 \). It is easily checked that inclusion (12), (13) is invariant with respect to this transformation.

Introducing new coordinates \( \Sigma = \sigma / \lambda^2 \) and \( \Psi = \xi / \lambda^2 \) and taking sign \( \sigma = 1 \) for positive \( \sigma \), achieve that system (5) describes trajectory \( \Gamma \) with some special initial conditions. Let \( \zeta_0 = \sigma_0, \quad \xi_0 = \xi_0 \), be the intersections of \( \Gamma \) with the axis \( \sigma = 0 \). It follows from the invariance of the system with respect to transformation \( G_{\sigma, \eta} \) that the value of \( |d_{\xi} / d_{\sigma}| \) does not depend on the initial value \( \sigma_0 \). Hence, \( \Phi = \Phi(t) \) and \( \Phi > |d_{\xi} / d_{\sigma}| \).

**Proofs of Theorems 2 and 3.** Consider the case when both measurement noises and discrete measurements are present. Let the input signal consists of a base signal \( f_0(t) \) and a measurement noise not exceeding \( \varepsilon \) in absolute value, \( \sigma = x - f_0(t) \) and \( \xi = u_1 - f_0(t) \). Let also the measurements be carried out at discrete times with time interval \( \tau \) and \( t \in \{t, t+\tau, \ldots \} \) where \( t_i, t_i+\tau \) are successive measurement times and \( t \) is the current time.

Then the following differential inclusion holds:

\[
\sigma \in -\lambda [\sigma, [-\varepsilon, \varepsilon]]^2 \text{sign} \sigma, \quad [-\varepsilon, \varepsilon], \quad t \leq t_i + \tau - t_i = \tau.
\]

(14)

\[
\xi \in [-\varepsilon, \varepsilon] \text{sign} \sigma, \quad \xi \in [-\varepsilon, \varepsilon], \quad t \leq t_i + \tau - t_i = \tau.
\]

(15)

If the continuous measurement case is considered, any appearance of indices and \( t \) is to be omitted in inclusion (14), (15).

With \( \varepsilon, \tau \) being zero, inclusion (14), (15) coincides with (12), (13), whose trajectories converge in finite time to the origin. It is easily seen that this implies the existence of a bounded invariant set attracting all the trajectories in finite time, when \( \varepsilon, \tau \) are small. All that is needed now is to show that its size has the asymptotics defined by the statements of the theorems when \( \varepsilon, \tau \rightarrow 0 \).
Applying operator $G_\lambda$ (see the Lemma 2 proof) to inclusion (14), (15) achieve in the coordinates $\Sigma = \sigma/\lambda^2$ and $\Psi = \zeta/\lambda^2$

$$\Sigma \in -\left[\Sigma_0 + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right] \text{sign}\left(\Sigma + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right) + \Psi,$$

$$\Psi \in -\frac{1}{\lambda^2} [x - C, x + C] \text{sign}\left(\Sigma + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right).$$

(16)

$$t_{i+1} - t_i = \eta \tau.$$ 

(17)

Consider the case when $\epsilon = 0$ (Theorems 3). Fix some values of the expressions $(x - C)/\lambda^2$ and $(x + C)/\lambda^2$ and measurement step $\tau_0$. Let the attraction set of inclusion (14), (15) be given by the inequalities $|\Sigma| \leq k_1, |\Psi| \leq k_2$. Let $\epsilon$ be another measurement step. After transformation $G_{\lambda / \eta}$ with $\eta = \epsilon / \tau_0$ achieve that $|\sigma| \leq \lambda^2 k_1, |\eta| \leq x^2 (k_2/\tau_0)^2, |\zeta| \leq \lambda^2 (k_2/\tau_0)^2$. Here $k_1, k_2, k_3, \lambda_0$ are some constants. Theorem 3 follows now from the equality $\delta = -\lambda |\sigma|^{1/2} \text{sign} \sigma + \zeta$.

To prove Theorem 2 achieve from system (16), (17) the following inclusion (continuous measurements are considered):

$$\Sigma \in -\left[\Sigma + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right] \text{sign}\left(\Sigma + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right) + \Psi,$$

$$\Psi \in -\frac{1}{\lambda^2} [x - C, x + C] \text{sign}\left(\Sigma + \frac{1}{\lambda^2} \left[ -\eta^2 \sigma, \eta^2 \sigma \right]\right).$$

(18)

(19)

Fix some values $\epsilon_0, \lambda_0, \tau_0, C_0$ and let the attraction set of (18), (19) with $\eta = 1$ be given by the inequalities $|\Sigma| \leq k_1, |\Psi| \leq k_2$. Consider other values $\epsilon, \lambda, \alpha, C$, keeping the ends of the segment $(x^2 / \lambda^2)(x - C, x + C)$ fixed. After transformation $G_{\lambda / \alpha}$ with $\eta = (\epsilon / \epsilon_0)^{1/2} \lambda_0 / \lambda$ achieve that $|\sigma| \leq x^2 \lambda_0 \eta^2 = x^2 (k_1, k_2, k_3, \lambda_0, \lambda, \alpha)$, $|\zeta| \leq x^2 \lambda_0 \eta^2 = \lambda (k_1, k_2, k_3, \lambda_0, \lambda, \alpha)$, $|\zeta| \leq x^2 \lambda_0 \eta^2 = \lambda (k_1, k_2, k_3, \lambda_0, \lambda, \alpha)$. Here $k_1, k_2, k_3, \lambda_0$ are some constants. Taking into account that

$$\delta = -\lambda |\sigma + [\sigma, \zeta]|^{1/2} \text{sign} |\sigma + [\sigma, \zeta]| + \zeta$$

achieve the statements of Theorem 2. \hfill \Box