State estimation of complex oscillatory system with uniform quantization under data rate constraints

Boris Andrievsky and Alexander L. Fradkov

Abstract—A scheme for state estimation for multiple-output nonlinear system under communication constraints is extended for the case of exogenous excitation and implemented for Mechatronic Multipendulum Setup. Experimental results are presented, showing efficiency of the proposed method.

Keywords: nonlinear system, state estimation, communication constraints, mechatronic setup

I. INTRODUCTION

During the last decade substantial interest has been shown in networked control systems (NCS). The idea is to use serial communication networks to exchange system information and control signals between various physical components of the systems that may be physically distributed. NCS are real-time systems where sensor and actuator data are transmitted through shared or switched communication networks, see e.g. [1]–[4]. Transmitting sensor measurement and control commands over wireless links allows rapid deployment, flexible installation, fully mobile operation and prevents the cable wear and tear problem in an industrial environment. The possibility of NCS motivates development of a new chapter of control theory in which control and communication issues are integrated, and all the limitations of the communication channels are taken into account. The introduction of a communication network into a NCS can degrade overall control system performance through quantisation errors, transmission time delays and dropped measurements. The limitations of estimation and control under constraints imposed by a finite capacity information channel have been investigated in detail in the control theoretic literature, see the surveys [5], [6], the monograph [4] and the references therein. It has been shown that stabilization of linear systems under information constraints is possible if and only if the capacity of the information channel exceeds the entropy production of the system at the equilibrium (Data Rate Theorem), [7]–[9].

Continuous-time nonlinear systems were considered in [10]–[12], where several sufficient conditions for different estimation and stabilisation problems were obtained. In [11], uniformly observable systems were considered and an “embedded-observer” decoder and a controller were designed, which semi-globally stabilizes this class of systems under data-rate constraints. In most of the above mentioned papers the coding-decoding procedure is rather complicated: the size of the required memory exceeds or equals to the dimension of the system state space. Such a draw-back was overcome in [13], where a first order coder scheme was proposed for SISO nonlinear autonomous systems, represented in the Lurie form (linear part plus nonlinearity, depending only on measurable outputs) and the limit possibilities of synchronization and state estimation under information constraints are established. Complexity of the scheme of [13] does not grow with the dimension of the system state.

The results of [13] are extended in [14], [15] to the MIMO case. Under the assumptions that the measurements on the transmitter’s side are perfect and the channel distortions and errors may be neglected, in [13]–[17] it is shown that the upper bound of the limit estimation error is proportional to the upper bound of the transmission error. As a consequence, it is proportional to the maximum rate of the coupling signal and inversely proportional to the information transmission rate (the channel capacity). In the present paper the data transmission scheme of [13]–[17] is modified for non-autonomous systems, and implemented for the multipendulum mechatronic setup. The state estimation scheme is studied experimentally for the real-world system, subjected to measurement errors and data losses.

The paper is organized as follows. The state estimation scheme of [13]–[17] is extended to the case of additional exogenous input in Section II. Results of experiments with the multipendulum setup are described in Section III. Concluding remarks are given in Section IV. Coding procedure and analytical evaluation of the transmission error are given in Appendix A.

II. STATE ESTIMATION SCHEME

Consider a system model in the Lurie form. In additional to [13]–[17], let us assume presence of an exogenous input signal and consider the following non-autonomous system:

\[ \dot{x}(t) = Ax(t) + \phi(y(t)) + Bu(t) , \quad y(t) = Cx(t) , \quad (1) \]

where \( x(t) \in \mathbb{R}^n \) is the state variables vector; \( y(t) \in \mathbb{R}^l \) is the system output; \( u(t) \in \mathbb{R}^m \) denotes an exogenous input; \( A \) is \((n \times n)\)-matrix; \( B \) is \((n \times m)\)-matrix; \( C \) is \((l \times n)\)-matrix; \( \phi(y) \) is a continuous nonlinear vector-function, \( \phi : \mathbb{R}^l \to \mathbb{R}^n \). We assume that the system is dissipative: all the trajectories of the system (1) belong to a bounded set \( \Omega \) (e.g. attractor of a...
chaotic system). Such an assumption is typical for oscillatory and chaotic systems. Let the signals \( u(t) \), \( y(t) \) be measured at the side of the source system (1) and be transmitted to the remote observer over the digital communication channel.

The observer has the following form:

\[
\dot{\hat{x}}(t) = A\hat{x}(t) + \varphi(\hat{y}(t)) + L(\hat{y}(t) - y(t)) + B\bar{u}(t), \quad \hat{y} = C\hat{x},
\]

(2)

where \( \hat{x}(t) \in \mathbb{R}^n \) is the vector of the state estimates, produced by the observer; \( \bar{u}(t), \hat{y}(t) \) are, respectively, plant (1) input and output signals, transmitted over the channel and restored by the decoder; \( L \) is the vector of the observer parameters (gain). Apparently, in absence of the transmission errors (i.e. if \( \bar{u}(t) \equiv u(t), \hat{y}(t) \equiv y(t) \)), the dynamics of the state error vector \( e(t) = x(t) - \hat{x}(t) \) are described by a linear equation

\[
\dot{e} = A_L e, \quad y = Cx,
\]

(3)

where \( A_L = A - LC \).

![Block-diagram of state estimation over a discrete communication channel](image)

In the case if the pair \((A, C)\) is observable, there exists \( L \) providing the matrix \( A_L \) with any given real eigenvalues. In particular all eigenvalues of \( A_L \) can have negative real parts, i.e. the system (3) can be made asymptotically stable and \( e(t) \to 0 \) as \( t \to \infty \). Therefore, in the absence of measurement and transmission errors the estimation error decays to zero.

Now let us take into account transmission errors. Let the observed signals \( u_i(t) \in \mathbb{R}^l, y_j(t) \in \mathbb{R}^l \) (\( i = 1, \ldots, m, j = 1, \ldots, k \)) be coded with symbols from a finite alphabet at discrete sampling time instants \( T_{s,i} = k_i T_{s,j} \), where \( T_{s,i}, T_{s,j} \) are the sampling periods. Disregarding difference in representation of digital numbers and symbols of the coding alphabet, let us denote the output codewords as \( \bar{u}_i[k] = \bar{u}(t_{k,i}), \bar{y}_j[k] = \bar{y}(t_{k,j}) \) (respectively). The block-diagram, illustrating the considered remote estimation scheme is shown in Fig. 1 (cf. [18], [19]).

At the present stage of the study we assume that the observations are not corrupted by the observation noise, and that transmission delay and distortion may be neglected.

Assume that zero-order extrapolation is used to convert the digital sequences \( \bar{u}_i[k], \bar{y}_j[k] \) (\( i = 1, \ldots, m, j = 1, \ldots, k \)) to the continuous-time input signals \( \bar{u}_i(t), \bar{y}_j(t) \) of the observer (2): \( \bar{u}_i(t) = \bar{u}_i[k], \bar{y}_j(t) = \bar{y}_j[k] \) as \( k_i T_{s,i} \leq t < (k_i + 1)T_{s,i} \), \( k_j T_{s,j} \leq t < (k_j + 1)T_{s,j} \). Then the transmission error vectors are as follows:

\[
\delta_{u_i}(t) = u(t) - \bar{u}(t) \in \mathbb{R}^m, \quad \delta_{y_j}(t) = y(t) - \bar{y}(t) \in \mathbb{R}^l.
\]

(4)

In presence of the transmission errors, equation (3) reads as

\[
\dot{e} = A_L e + \varphi(y) - \varphi(y + \delta_y) - B\delta_{u} - L\delta_{y}.
\]

(5)

We are interested in limitations imposed on the estimation precision by limited transmission rate. To this end introduce an upper bound on the limit output estimation errors \( Q_i = \sup_{t \to \infty} \| e_i(t) \| \), where \( e_i(t) = y_i(t) - \hat{y}_i(t), i = 1, \ldots, l \), and the supremum is taken over all admissible transmission errors. In [13], [14] it has been proved analytically and illustrated by numerical examples, that for the considered estimation scheme, applied to an autonomous system, the total estimation error is proportional to the upper bound of the norm of the transmission error and, in turn, is inversely proportional to the transmission rate. It is easy to show that this relationship is also valid for non-autonomous system, if an exogenous signal is transmitted to the observer at the receiver’s end by means of the coding-decoding procedure (8), (10), (11). In the next section, the experiments on the mechatronic laboratory setup for evaluation of the estimation error are described. It is shown that the experimental results agree with the general statements of [13], [14].

III. EXPERIMENTS ON STATE ESTIMATION OVER THE COMMUNICATION CHANNEL

A. Experimental setup

The Multipendulum Mechatronic Setup of the IPME RAS (MMS IPME) consists of the set of interconnected pendulum sections, electrical equipment (with the computer interface facilities), the electric computer-controlled motor, the personal computer for data processing and real-time representation of the results. The mechanical part of the setup consists of a number of identical sections with pendulums, diffusively connected by torsion springs. The computer-controlled electric motor is connected with the first pendulum of the chain via the spring, applying the torque to the “left” end of the chain. The “right” end of the chain is unconnected. Axes of the neighboring sections are connected by torsion springs, arranging force interaction between pendulums. In principle, any number of sections can be connected. At the moment mechanical parts of 50 sections are manufactured. The setup is described in more details in [20].

B. Modeling the chain of pendulums

Following [20], the rotation angle of the drive shaft, connected with the first pendulum of the chain, is considered as the system input. The last (Nth) pendulum in the chain is mechanically connected only with the previous one, no boundary conditions for Nth pendulum are specified. This
leads to the following model of the chain dynamics:

\[
\begin{align*}
\dot{\phi}_1 + \rho \dot{\phi}_1 + \Omega^2 \sin \varphi_1 - k(\varphi_2 - 2\varphi_1) &= ku(t), \\
\dot{\phi}_i + \rho \dot{\phi}_i + \Omega^2 \sin \varphi_i - k(\varphi_{i+1} - 2\varphi_i + \varphi_{i-1}) &= 0, \\
(i &= 2, 3, \ldots, N-1), \\
\dot{\phi}_N + \rho \dot{\phi}_N + \Omega^2 \sin \varphi_N - k(\varphi_N - \varphi_{N-1}) &= 0,
\end{align*}
\]

(6)

where \(\varphi_i = \phi_i(t)\) (i = 1, 2, \ldots, N) are the pendulum deflection angles; \(u = u(t)\) is the controlling action (the rotation angle of the drive shaft). The values \(\rho, \omega_0, k\) are the system parameters: \(\rho\) is the viscous friction parameter; \(\Omega\) is the natural frequency of small oscillations of the isolated pendulum; \(k\) is the coupling strength parameter, which depends on the softness of the connecting spring. The model (6) parameters have been preliminary estimated based on the mass-geometry properties of the mechanical system, and then were specified by means of the trial-and-error procedure, applied to the experimental data sets. The following parameter estimates were finally obtained: \(\Omega = 5.5\ \text{s}^{-1}\), \(\rho = 0.95\ \text{s}^{-1}\), \(k = 5.8\ \text{s}^{-2}\).

C. Data transmission and state estimation algorithms for the multi-pendulum set-up

In our experiments, the chain of four pendulum sections and the motor, attached via the spring to pendulum #1, were used. The outside left rotary angle (the angle of the drive shaft) may be referred to as exogenous action, applied to the plant (the chain of the pendulums), it was coded by means of the first-ordered coder (8), (11), (10), where \(z \equiv \varphi_m\) is taken. The model (6) is used for designing the remote state estimator (2), (8), (11), (10) for the pendulum angles \(\varphi_i(t)\). Namely, in (8), (11), (10) \(z = \varphi_i\) for \(i = 1, \ldots, 4\) is taken.

For the considered problem, observer (2) has been designed by means of (6) decomposition into four interconnected subsystems of the second order. This leads to the following observer equations:

\[
\begin{align*}
\dot{\hat{\phi}}_1 &= \hat{\omega}_1 + l_1 \hat{\varphi}_1, \\
\dot{\hat{\omega}}_1 &= -\rho \hat{\omega}_1 - \Omega^2 \sin \hat{\varphi}_1 + k(\hat{\varphi}_2 - 2\hat{\varphi}_1) + k\hat{\varphi}(m)(t) + l_2 \hat{\varphi}_1, \\
\dot{\hat{\phi}}_2 &= \hat{\omega}_2 + l_1 \hat{\varphi}_2, \\
\dot{\hat{\omega}}_2 &= -\rho \hat{\omega}_2 - \Omega^2 \sin \hat{\varphi}_2 + k(\hat{\varphi}_3 - 2\hat{\varphi}_2 + \hat{\varphi}_1) + l_2 \hat{\varphi}_2, \\
\dot{\hat{\phi}}_3 &= \hat{\omega}_3 + l_1 \hat{\varphi}_3, \\
\dot{\hat{\omega}}_3 &= -\rho \hat{\omega}_3 - \Omega^2 \sin \hat{\varphi}_3 + k(\hat{\varphi}_4 - 2\hat{\varphi}_3 + \hat{\varphi}_2) + l_2 \hat{\varphi}_3, \\
\dot{\hat{\phi}}_4 &= \hat{\omega}_4 + l_1 \hat{\varphi}_4, \\
\dot{\hat{\omega}}_4 &= -\rho \hat{\omega}_4 - \Omega^2 \sin \hat{\varphi}_4 + k(\hat{\varphi}_5 - \hat{\varphi}_4) + l_2 \hat{\varphi}_4,
\end{align*}
\]

(7)

where \(\hat{\phi}_i, \hat{\omega}_i\) (\(i = 1, \ldots, 4\)) stand for the estimates of the rotation angle and the angular velocity of the \(i\)-th pendulum (respectively), \(\hat{\varphi}_i = \hat{\phi}_i - \hat{\phi}_i\) are the observer output errors; \(l_1, l_2\) are the observer gains. To find them, an isolated linear subsystem with the matrices \(A_1 = \begin{bmatrix} 0 & 1 \\ -2k & -\rho \end{bmatrix}\), \(C_1 = [1,0]\) was considered. The gains \(l_1, l_2\) have been found ensuring the prescribed eigenvalues \(s_{1,2} = -14 \pm 14i\) (where \(i\) is an imaginary unit) of the matrix \(A_1 - LC_1\). This gives the following values of the observer gains \(l_1, l_2\): \(l_1 = 27.3, l_2 = 362\). It may be easily verified that for the chosen parameters, the spectrum of the matrix \(A_L\) in (5) is as follows: \([-14.14 \pm 13.75i, -14.14 \pm 13.93i, -14.14 \pm 14.45i, -14.14 \pm 14.21i]\), which leads to exponential stability of observer (2).

D. Experimental results

In course of the experiments, the harmonic waveform voltages have been applied to the motor. The rotary angles of the drive shaft and the pendulums were measured with the sampling rate of 100 Hz and 2° precision by means of the optical sensors. Then the measured signals have been processed by the above coding algorithms for transferring over the channel. During the experiments, the sampling times for the drive motor \(T_m\) and the pendulum angles \(T_{\varphi,i}\) and the quantizer (8) parameters \(v_i\) were taken equal: \(T_m = T_{\varphi,1} = \cdots = T_{\varphi,4}\), \(v_m = v_1 = \cdots = v_4\), varying from one data processing run to another.

Each experiment lasted in \(t_{\text{run}} = 100\) s. The relative output estimation errors \(Q_i(R, v)\) \((i = 1, \ldots, 4)\) have been calculated as

\[
Q_i(R, v) = \frac{\max_{t \in [t_{\text{begin}} + t_{\text{fin}}]} |y_i(t) - \hat{y}_i(t)|}{\max_{t \in [t_{\text{begin}} - t_{\text{fin}}]} |y_i(t)|},
\]

where \(t_{\text{begin}} = 10\) s is taken to eliminate influence of transients on the accuracy indices \(Q_i(R, v)\).

Experimental results are depicted in Figs. 2–6. The time histories of the rotation angle of the first pendulum \(\varphi_1(t)\), obtained by the experiment, and its estimate \(\hat{\varphi}_1(t)\), produced by observer (7) at the decoder’s side, are plotted in Figs. 2, 3 for \(v = 0\) (binary coder) and for \(v = 2\). Corresponding estimates \(\hat{\phi}_1\) of the angular velocity \(\phi_1\) are shown in Figs. 4, 5. It should be noticed, that the angular velocities are not measured by sensors and are subjected to estimation.

The generalized accuracy indexes are plotted in Figs. 6, ??, where dependence of limit output estimation error for the first pendulum \(Q_1\) on the overall transmission rate \(R_C\) and parameter \(v\) is reflected. It is worth mentioning that the “exact” values of \(\varphi_1(t)\), \(\hat{\varphi}_1(t)\) are not known due to the measuring errors in the optical sensors, effecting on experimental evaluation of the data transmission accuracy. Namely, the overall error can not be less than the optical sensor error. It is seen from the plots that dependence of the estimation error, obtained by the experiments, are close to the biased inversely proportional function on the transmission rate. One also may notice that small values of \(v\) are preferable from the viewpoint of the transmission rate for given accuracy, cf. [13], [21].

IV. CONCLUSIONS

We have studied dependence of the error of state estimation for nonlinear Lurie systems over a limited-band communication channel both analytically and numerically. It is demonstrated that upper bound for limit estimation error depends linearly on the transmission error which, in turn, is proportional to the driving signal rate and inversely proportional to the transmission rate. Though these results are obtained for a special type of coder, it reflects peculiarity
of the estimation problem as a nonequilibrium dynamical problem. On the contrary, the stabilisation problem considered previously in the literature on control under information constraints belongs to a class of equilibrium problems.

APPENDIX

A. Coding procedure

Following [13], for a given real number $\varepsilon > 0$ and nonnegative integer $v \in \mathbb{Z}$ define a uniform quantizer to be a discretization interval $\delta = 2^{1-v} \varepsilon$ and the quantizer $q_{v,\varepsilon}(z)$ as

$$q_{v,\varepsilon}(z) = \begin{cases} \delta \cdot \lfloor \delta^{-1} z \rfloor, & \text{if } |z| \leq \varepsilon, \\ \varepsilon \text{sign}(z), & \text{otherwise,} \end{cases} \tag{8}$$

where $z$ denotes the signal to be transmitted over the channel (in our case, $z \in \{0,1\}$), $\lfloor \cdot \rfloor$ denotes round-up to the nearest integer, $\text{sign}(\cdot)$ is the signum function: $\text{sign}(y) = 1$, if $y \geq 0$, $\text{sign}(y) = -1$, if $y < 0$. Therefore, the cardinality of the mapping $q_{v,\varepsilon}$ image is equal to $2^v + 1$, and each codeword symbol contains $\bar{R} = \log_2(2^v + 1) = \log_2(2\varepsilon/\delta + 1)$ bits. Thus, the discretized output of the considered coder is found as $z = q_{v,\varepsilon}(z)$. We assume that the coder and decoder make decisions based on the same information [18], [19].

In the present paper we use $l + m$ independent coders for components $u_i$, $y_j$ ($i = 1, \ldots, m$, $j = 1, \ldots, l$) of the transmitted vectors $u \in \mathbb{R}^m$, $y \in \mathbb{R}^l$. Each coder number $i$, $j$ has its particular sampling period $T_{x,i}$, $T_{x,j}$, ranges $\nu_i$, $\nu_j$ and integers $\nu_i$, $\nu_j$. The corresponding bit-per-second rates $R_i$, $R_j$ are calculated as $R_i = \bar{R}_i/T_{x,i} = \log_2(2^\nu_i + 1)/T_{x,i}$, $R_j = \bar{R}_j/T_{x,j} = \log_2(2^\nu_j + 1)/T_{x,j}$. The overall averaged rate $R$ is a sum of the particular ones, $R = \sum_{i=1}^{l} R_i + \sum_{j=1}^{m} R_j$.

The static quantizer (8) is a part of the time-varying coders with memory, see e.g. [7], [13], [19], [22]. In the first-order (one-step memory) coder the central vectors (the “centroids”) $c[k] \in \mathbb{R}^l$, $k \in \mathbb{Z}$ with initial condition $c[0] = 0$ is utilized. At step $k$ the coder compares the current measured output $z[k]$ with the number $c[k]$, forming the deviation vector $\partial z[k] = z[k] - c[k]$. Then this vector is discretized with given $\varepsilon = |\varepsilon| k$ according to (8). The output signal

$$\partial z[k] = q_{v,\varepsilon}(\partial z[k]) \tag{9}$$

is represented as an $\bar{R}$-bit information symbol from the coding alphabet and transmitted over the communication
channel to the decoder. Then central number \( c[k+1] \) and range parameter \( \varkappa[k] \) are renewed based on the available information about the drive system dynamics. Assuming that transmitted signal \( z(t) \) changes at a slow rate, i.e. that \( z[k+1] \approx z[k] \), the following update algorithms are employed:

\[
c[k+1] = c[k] + \hat{\varepsilon}[k], \quad c[0] = 0, \quad (10)
\]

\[
\varkappa[k] = (\varkappa_0 - \varkappa_{\infty}) \rho^k + \varkappa_{\infty}, \quad k = 0, 1, \ldots, \quad (11)
\]

where \( 0 < \rho \leq 1 \) is the decay parameter, \( \varkappa_{\infty} \) stands for the limit value of \( \varkappa \). The initial value \( \varkappa_0 \) should be large enough to capture all the region of possible initial values of \( z_0 \).

Equations (8), (9), (11) describe the coder algorithm. A similar algorithm is used by the decoder. Namely: the sequence of \( \varkappa[k] \) is reproduced at the receiver node utilizing (11); the values of \( \hat{\varepsilon}[k] \) are restored with given \( \varkappa[k] \) from the received codeword; the central numbers \( c[k] \) are found in the decoder in accordance with (10). Then \( \hat{z}[k] \) is found as a sum \( c[k] + \hat{\varepsilon}[k] \).

It worth to mention that the logarithm quantization [23], [24].

### B. Evaluation of the estimation error

Denoting in the right hand of (5) the sum \( \phi(y) - \phi(y + \delta_i) - B \delta_i - L \delta_i \) as \( \xi(t) \in \mathbb{R}^n \), one obtains the following observer error equation:

\[
\dot{e}(t) = \Lambda e(t) + \xi(t), \quad (12)
\]

where \( \xi(t) \) is \( l_2 \)-norm bounded disturbing vector, \( \|\xi(t)\| \leq C_\xi \), where \( C_\xi \geq 0 \) depends on the magnitudes of \( \delta_i(t), \hat{\delta}_i(t) \), on function \( \phi(\cdot) \) and matrices \( A, B \). Namely, assuming that a nonlinearity \( \phi(\cdot) \) is Lipschitz continuous along all the trajectories of the observed system (1), i.e. that there exists a positive real number \( L_\phi > 0 \) such that \( \|\phi(y) - \phi(y + \delta_i)\| \leq L_\phi \|\delta_i\| \), one may majorize \( C_\xi \) as \( C_\xi \geq (L_\phi + |L_i|)\|\delta_i\| + |B| \|\delta_i\| \).

Let us evaluate the magnitudes \( \|\delta_i\|, \|\delta_i\| \) of the transmission errors \( \delta_i(t), \hat{\delta}_i(t) \). Since all the components \( y_i(t) \) \((i = 1, \ldots, m), y_j(t) \) \((j = 1, \ldots, l)\) are transmitted over the channel by means of scalar coding–decoding procedures (8)–(11), let us denote the scalar signal to be transmitted as \( z(t) \) and evaluate the upper bound \( \Delta_s = \sup \|\delta_i(t)\| \|\delta_i(t)\| \) of the data transmission error \( \delta_i(t) = z(t) - \hat{z}(t) \). To this end let us recall some statements from [13].

Let the growth rate of \( z(t) \) be uniformly bounded, i.e. there exists a positive real number \( L_z \) such as \( \sup \|z(t)\| \leq L_z \).

Let the bit-per-second rate \( R \), parameter \( v \in \frac{1}{L_z} \), and range parameter \( \kappa \) of quantizer (8) be given (design) parameters. Our aim is to find \( \Delta_s \) as a function of \( R, \kappa \) and \( L_z \).

The quantization interval \( \delta \) of quantizer (8) is defined as

\[
\delta = 2^{-v} \kappa. \quad (13)
\]

As follows from (8), the magnitude of the quantization error \( \delta_v = \hat{z} - \partial \) (where \( \partial[k] = z[k] - c[k] \)) for given \( \partial \) does not exceed \( \delta / 2 = 2^{-v} \kappa \) if

\[
|\partial| \leq \kappa + \Partial Derivative / 2 = (1 + 2^{-v}) \kappa. \quad (14)
\]

Inequality (14) corresponds to the nonsaturated quantization, e.g. to the “nominal case”. If (14) is violated, then the error grows linearly on \( \partial \). Therefore the magnitude of the quantization error \( \delta_v \) satisfies the inequality

\[
|\delta_v| = \left\{ \begin{array}{ll}
2^{-v} \kappa, & \text{if } |\partial| \\ |\partial| & \text{otherwise}.
\end{array} \right. \quad (15)
\]

The bit-per-step rate \( \bar{R} \) for quantizer (8) is given by the following expression:

\[
\bar{R} = \log_2(2^v + 1). \quad (16)
\]

Since the bit-per-second rate \( R \) may be found as \( R = \bar{R} / T_z \), where \( T_z \) denotes the sampling period, the following expression for \( T_z \) in the terms of the other coder parameters may be found:

\[
T_z = \frac{\log_2(2^v + 1)}{R}. \quad (17)
\]

Now let us turn to procedure (10) of updating the central number. Firstly, assume, that at \( k \)-th step the measured value \( z[k] = z(t_k), \ t_k = kT_z \), belongs to the interval \([-\kappa - \delta / 2, \kappa + \delta / 2]\) around the central number \( c[k] \):

\[
|z[k] - c[k]| \leq \kappa + \delta / 2. \quad (18)
\]

This means fulfillment of inequality (14). Therefore in this case the quantization error \( \delta_v \) does not exceed \( 2^{-v} \kappa \).

According to update algorithm (10), the next value of the central number is found as

\[
c[k+1] = c[k] + \hat{\varepsilon}[k]. \quad (19)
\]

 Apparently, \( c[k+1] \) represents \( z[k] \) with a maximal error \( 2^{-v} \kappa \), i.e. \( |c[k+1] - z[k]| \leq 2^{-v} \kappa \). To ensure the nonsaturated mode of the coder, inequality (18) should be valid at the next step \( k := k + 1 \). This may be guaranteed if the quantizer range \( \kappa \) is sufficiently large for given \( L_z \geq \sup|z| \). Since

\[
|z(t_{k+1}) - z(t_k)| \leq L_zT_z, \text{ then } \kappa \text{ should satisfy the following inequality:}
\]

\[
\kappa \geq T_zL_z. \quad (20)
\]

Consider now the case when (18) is violated for some \( k = k_0 \). In this case the quantization error may be large depending on \( c[k_0] \), see (15). However, it may be easily shown that if the quantizer range \( \kappa \) satisfies inequality (20) with some margin, i.e. if \( \kappa \) is taken as

\[
\kappa \geq (1 + \theta)T_zL_z, \quad (21)
\]

where \( \theta > 0 \) is an arbitrary small real number, then there exists an integer \( k_0 < k < \infty \) such that (18) is fulfilled for all \( k \), starting from \( k = k_0 \). If \( \theta = 0 \) is taken in (21), then only asymptotic convergence may be ensured. In what follows, we assume that validity of (19) is assured by means of an appropriate choise of \( \kappa[0] \) in zooming procedure (11).

It follows from (17), (20) that \( \kappa \) should satisfy the relation:

\[
\kappa \geq \log_2(2^v + 1) \frac{L_z}{R}. \quad (22)
\]
To analyze the coder-decoder accuracy let us evaluate the upper bound \(\Delta_c = \sup_{t}|\delta_c(t)|\) of the transmission error \(\delta_c(t) = z(t) - \bar{z}(t)\). Consider the sampling interval \([t_k, t_{k+1}]\), \(t_k = kT_s\). It is shown above that \(|\delta_c(t_k)|\) does not exceed \(\delta/2 = 2^{-v}\kappa\). Additionally, the error may increase from \(t_k\) to \(t_{k+1}\) due to a change of \(z(t)\) by a value not exceeding \(L_z T_s\). Therefore the total transmission error for each interval \([t_k, t_{k+1}]\) meets the inequality \(|\delta_c(t_k)| \leq 2^{-v}\kappa + L_z T_s\) and \(\Delta_c\) may be found as

\[
\Delta_c = 2^{-v}\kappa + L_z T_s. \tag{23}
\]

Taking into account expression (17) for the sampling period \(T_s\), we obtain the following relation between the transmission error and other parameters:

\[
\Delta_c = 2^{-v}\kappa + \log_2(2^{v+1}) \frac{L_z}{R}, \tag{24}
\]

where \(\kappa\) should satisfy (22). Since we are aimed to minimize the transmission error for a given rate, it is naturally to choose a minimal admissible value for \(\kappa\). This gives the following expressions for \(\kappa\) and \(\Delta_c\):

\[
\kappa = \log_2(2^{v+1}) \frac{L_z}{R}, \tag{25}
\]

\[
\Delta_c = (1 + 2^{-v}) \log_2(2^{v+1}) \frac{L_z}{R}. \tag{26}
\]

It should be noticed that (25), (26) lead to the following relations between \(\Delta_c\) and \(\kappa\) [13]:

\[
\Delta_c = (1 + 2^{-v}) \kappa, \quad \kappa = \frac{2^v}{1 + 2^v} \Delta_c. \tag{27}
\]

Defining a multiplier \(\lambda(v)\) as \(\lambda(v) = (1 + 2^{-v}) \log_2(2^{v} + 1)\) let us rewrite (24) in the following form

\[
\Delta_c = \lambda(v) \frac{L_z}{R}. \tag{28}
\]

Following [13] let us find \(v\) minimizing \(\Delta_c\) for given \(L_z, R\). The derivative of \(\lambda(v)\) on \(v\)

\[
\frac{d\lambda}{dv} = \frac{2^v - \ln(2^{v} + 1)}{2^v} \ln(2) \tag{29}
\]

is strictly positive. Therefore, \(\lambda(v)\) strictly grows on \(v\), and \(\Delta_c\) is minimized at \(v = 0\). This means optimality of the binary quantizer \(q_v(z) = v \text{sign}(z)\) in the sense of the transmission error for a given rate \(R\) (cf. [13], [21]).

REFERENCES