On Shallit’s Minimization Problem

S. Yu. Sadov

1*Mathematical Notes*, Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, 119991 Russia

Received May 7, 2019; in final form, May 29, 2021; accepted May 30, 2021

Abstract—In Shallit’s problem (SIAM Review, 1994), it was proposed to justify a two-term asymptotics of the minimum of a rational function of \( n \) variables defined as the sum of a special form whose number of terms is of order \( n^2 \) as \( n \to \infty \). Of particular interest is the second term of this asymptotics (“Shallit’s constant”). The solution published in SIAM Review presented an iteration algorithm for calculating this constant, which contained some auxiliary sequences with certain properties of monotonicity. However, a rigorous justification of the properties, necessary to assert the convergence of the iteration process, was replaced by a reference to numerical data. In the present paper, the gaps in the proof are filled on the basis of an analysis of the trajectories of a two-dimensional dynamical system with discrete time corresponding to the minimum points of \( n \)-sums. In addition, a sharp exponential estimate of the remainder in Shallit’s asymptotic formula is obtained.

DOI: 10.1134/S0001434621090078

Keywords: Shallit’s constant, minimization, hyperbolic point, local linearization, rate of convergence.

1. INTRODUCTION

The subject here is a problem proposed by J. Shallit in 1994 [1] in the problem section (later abolished) of SIAM Review.

Problem. Let \( \mathbf{x} = (x_1, \ldots, x_n) \) be a vector with components \( x_i > 0 \). Denote

\[
  f_n(\mathbf{x}) = \sum_{i=1}^{n} x_i + \sum_{1 \leq i \leq j \leq n} \prod_{k=i}^{j} \frac{1}{x_k}. \tag{1}
\]

Prove that there exists a constant \( C > 0 \) such that

\[
  \min_{\mathbf{x} > 0} f_n(\mathbf{x}) = 3n - C + o(1) \quad \text{as} \quad n \to \infty. \tag{2}
\]

Here and below, we write \( \mathbf{x} > 0 \) or \( \mathbf{x} \in \mathbb{R}_n^+ \) if \( x_i > 0, i = 1, \ldots, n \).

We put

\[
  A_n = \min_{\mathbf{x} > 0} f_n(\mathbf{x}), \quad C_n = 3n - A_n. \tag{3}
\]

In (2) and (3), it is justified to use the symbol min rather than inf, because the continuous function \( f_n(\cdot) \) attains minimum on the cube

\[
  t^{-1} \leq x_i \leq t, \quad 1 \leq i \leq n,
\]

where \( t = f_n(1,1,\ldots,1) \) and \( f_n(\mathbf{x}) > t \) outside this cube.

*E-mail: serge.sadov@gmail.com
The constant

\[ C = \lim_{n \to \infty} C_n = 1.3694514039 \ldots \]

is known as \textit{Shallit's constant} [2, Sec. 3.1], [3].

Professor Shallit informed the author (on April 18, 2018) that he came to the described problem, which was vaguely reminiscent of the problem published at about the same time in [4], when he was looking for similar inequalities of a nonstandard type.

In turn, the author learned about Shallit’s problem from the book [2], where it is mentioned in the section devoted to the Shapiro–Drinfeld constant; a generalization of the Shapiro problem concerning the lower bound of a cyclic sum was studied by the author in [5]. In [6, Sec. 7], the author interpreted the problem of minimizing the Shallit sums \( f_n(\mathbf{x}) \) as a special case of the problem of minimizing a sum of positive quantities when some of their partial products have fixed values.

Shallit’s problem was analyzed by Grosjean and De Meyer in [7], where they introduced auxiliary sequences in terms of which the quantities \( C_n \) were expressed. As a result, an algorithm for calculating the constant \( C \) with an arbitrary accuracy was obtained. The proof of convergence of the algorithm is based on the formulated properties of monotonicity of the sequences under study; these properties have been confirmed numerically but have not been proved rigorously.

The assertions in [7] which until now had the status of numerical heuristics, are proved in this paper. Moreover, we show that the remainder \( o(1) \) in the asymptotic formula (2) is

\[ \Theta(\rho^{-n}), \quad \text{where} \quad \rho = 2 + \sqrt{3}. \]

As usual, \( f = \Theta(g) \) or \( f \asymp g \) is an abbreviated form of the two-sided estimate \( m|f| \leq |g| \leq M|f| \) with constants \( 0 < m < M \) whose values we do not specify.

The paper is organized as follows.

In Section 2, partially following [7], we introduce a dynamical system with discrete time whose trajectories satisfying special boundary conditions correspond to the minimizing vectors for \( f_n( \cdot ) \).

Section 3 contains a chain of short lemmas that cover the assertions from [7] that were not proved there rigorously. The proof of convergence of the sequence \( (C_n) \) is completed by means of elementary analysis, in Section 4 (see Theorem 1).

In Section 5, the exact rate of convergence of the sequence \( (C_n) \) and some related estimates are obtained. Part (c) of Theorem 2 is a proper refinement of Theorem 1. Section 5 is less elementary; the key role here is played by a rather subtle Hartman theorem on smooth local linearization of a two-dimensional hyperbolic diffeomorphism.

We here keep the notation introduced in [7] with one exception: the quantity called \( \lambda_j \) in [7] here will be called \( \lambda_j^* \) (Sec. 3.6).

2. DERIVATION OF THE BOUNDARY-VALUE PROBLEM

2.1. Equation for the Extremum Point

The change of variables

\[ x_1 = \frac{1}{u_1}, \quad x_j = \frac{1 + u_{j-1}}{u_j}, \quad j = 2, 3, \ldots, n, \]  

proposed in [7] leads to the objective function \( g_n(\mathbf{u}) = f_n(\mathbf{x}) \) of a form simpler than (1), i.e., with \( \Theta(n) \) rather than with \( \Theta(n^2) \) terms,

\[ g_n(\mathbf{u}) = \sum_{j=1}^{n} L(u_{j-1}, u_j), \quad L(t, s) = s + \frac{1 + t}{s}, \]  

where \( \mathbf{u} = (0, u_1, u_2, \ldots, u_n) \). We put \( \hat{\mathbf{u}} = (u_1, u_2, \ldots, u_n) \). The transformation (5), \( \mathbf{x} \mapsto \hat{\mathbf{u}} \), is bijective on \( \mathbb{R}_+^n \). Therefore,

\[ A_n = \min_{\hat{\mathbf{u}} > 0} g_n(\mathbf{u}). \]
Let us write a necessary condition for the extremum, \( \nabla g_n(u) = 0 \), in expanded form:

\[
\frac{1 + u_j-1}{u_j^2} = \frac{1}{u_{j+1}} + 1, \quad j = 1, \ldots, n - 1, \\
\frac{1 + u_{n-1}}{u_n^2} = 1, \quad u_0 = 0.
\]

### 2.2. Dynamical System and Boundary-Value Problem

We define a partial mapping of \( \mathbb{R}^2 \) into itself, \( \Phi: (p, u) \mapsto (p', u') \), by the formulas

\[
p' = p^2(u + 1) - 1, \quad u' = 1/p, \quad p \neq 0.
\]

The trajectory of the dynamical system determined by the mapping \( \Phi \) with the initial point \((p_0, u_0)\) is a finite or infinite sequence of iterations \((p_j, u_j) = \Phi^j(p_0, u_0)\). In what follows, we use only the trajectories with coordinates \(p_j \geq 0, u_j > -1\).

For a given positive integer \(n\), we define a finite trajectory \(T_n = \{(p_j, u_j), j = 0, 1, \ldots, n\} \) by the conditions

\[
u_0 = p_n = 0 \quad \text{and} \quad p_j > 0, \quad 0 \leq j \leq n - 1.
\]

It is easy to see that the coordinates \(u_j\) of points of such a trajectory satisfy Eqs. (8). Conversely, a positive solution of the system of equations (8) (such a solution exists, because minimum (7) is attained) determines a trajectory \(T_n\): we put

\[
p_{j-1} = \frac{1}{u_j}, \quad j = 1, \ldots, n, \quad p_n = 0.
\]

The uniqueness of the trajectory \(T_n\), i.e., of a positive solution of system (8), is shown below in Lemma 2(c).

**Notation.** We denote the points of a trajectory of general form or a trajectory \(T_n\) with a fixed (in the context) \(n\) by \((p_j, u_j)\). If we simultaneously consider a family of trajectories \(T_n\), then we denote the points of the trajectory \(T_n\) by \((p_{j,n}, u_{j,n}), 0 \leq j \leq n\). In [7], either the complete \(u_{j,n}\) or brief \(u_j\) notations were also used.

### 3. ANALYSIS OF TRAJECTORIES

#### 3.1. Qualitative Picture

The only fixed point \(P_0 = (1, 1)\) of the mapping \(\Phi\) in \(\mathbb{R}^2\) is a hyperbolic point: the eigenvalues of the Jacobi matrix

\[
D\Phi(1, 1) = \begin{bmatrix} 4 & 1 \\ -1 & 0 \end{bmatrix}
\]

are \(\rho = 2 + \sqrt{3} > 1\) and \(\rho^{-1} = 2 - \sqrt{3} < 1\). It follows from the general theory [8, Theorem 6.2.3] that the following invariant curves are determined in a neighborhood of the point \(P_0\): a stable curve \(\gamma_s\) and an unstable curve \(\gamma_u\). The curve \(\gamma_s\) (appropriately, \(\gamma_u\)) is tangent at \(P_0\) to the stable (appropriately, unstable) separatrix \(\tau_s\) (\(\tau_u\)) of the linearized mapping.

The existence of the asymptotics (2) is related to the fact that the invariant curves can be extended at least as far as the coordinate axes: the stable curve \(\gamma_s\) intersects the ray \(u = 0, p > 0\) at a point \(P_s\) and the unstable curve \(\gamma_u\) intersects the ray \(p = 0, u > 0\) at a point \(P_u\). The trajectory \(T_n\) determined by the boundary conditions (10) starts near \(P_s\) and terminates near \(P_u\), see the figure. The initial iterations rapidly approach \(P_0\) along the arc of the curve \(\gamma_s\) from \(P_s\) to \(P_0\). The final iterations rapidly move away from \(P_0\) along the arc of the curve \(\gamma_u\) from \(P_0\) to \(P_u\). An arbitrarily small prescribed neighborhood of the point \(P_0\) contains all but \(O(1)\) points of the trajectory \(T_n\). Most of the terms \(L(u_{j-1}, u_j)\) in (6) at the extremum point are close to \(L(1, 1) = 3\).

The outlined qualitative considerations are concretized below as quantitative estimates. The invariant curves will be used explicitly only in Section 5.
Fig. 1. Invariant curves and the trajectory \( T_n \).

### 3.2. Identities

**Lemma 1.** The following assertions hold:

(a) if \((p', u') = \Phi(p, u)\), then \(p' u' + u' = pu + p\);

(b) in particular, \(p_j u_j + u_j = p_j-1 u_j-1 + p_j-1\) for any trajectory;

(c) for the trajectory \( T_n \) and \(1 \leq \ell \leq n\),

\[
\sum_{j=1}^{\ell} (p_{j-1} - u_j) = p_\ell u_\ell.
\]

**Proof.** (a) Straightforward verification.

(c) This assertion follows from (b) and the relation \(u_0 = 0\). \(\square\)

### 3.3. Monotonicity with Respect to Initial Data

For the values \(a > -1\), we define rational functions \(P_{a,j}(t), U_{a,j}(t), j = 0, 1, 2, \ldots\), by setting \(P_{a,0}(t) = t\) and \(U_{a,0}(t) = a\), and further, we define them recurrently as

\[
(P_{a,j}(t), U_{a,j}(t)) = \Phi(P_{a,j-1}(t), U_{a,j-1}(t)) = \Phi^j(t, a).
\]

For example, \(P_{a,1}(t) = t^2(a + 1) - 1\), \(U_{a,1}(t) = 1/t\).

We will mainly use the functions \(P_j(\cdot) = P_{0,j}(\cdot)\) and \(U_j(\cdot) = U_{0,j}(\cdot)\), but in the proof of Lemma 4(a), we use the functions \(P_{1,j}(\cdot)\), and in Sec. 5.5, these functions with an arbitrary \(a\).

**Lemma 2.** (a) \(P_{a,j}(t) \sim ct^{\nu}\) as \(t \to \infty\), where \(c = c(a, j) > 0\) and \(\nu = \nu(j) \geq 1\).

(b) There exist values \(0 = t_{a,0} < t_{a,1} < \ldots < t_{a,n} < \ldots\) such that \(P_{a,j}(t_{a,j}) = 0\), \(P_{a,j}(t) > 0\) for \(t > t_{a,j}\), and the functions \(P_{a,j+1}(t), P_{a,j+1}(t)U_{a,j+1}(t), \) and \(-U_{a,j+1}(t)\) increase for \(t > t_{a,j}\).

(c) The trajectory \(T_n, n \geq 1\), satisfying condition (10) is unique and corresponds to the initial conditions \(u_0 = 0, p_0 = t_0, n\).

(d) The abscissas of points on the trajectories \(T_n\) and \(T_{n+1}\) satisfy the inequalities

\[
p_{j,n+1} > p_{j,n}, \quad 0 \leq j \leq n.
\]
Finally, the de...\[ \sigma(a) \]

Lemma 4. On the trajectory \( T_n \), the inequalities \( p_j u_j < 1 \) are satisfied for \( 0 \leq j \leq n \).

(b) The trajectory \( T_n \) goes in the northwest direction, i.e.,

\[ p_0 > p_1 > \cdots > p_{n-1} > p_n = 0; \quad 0 = u_0 < u_1 < \cdots < u_n. \]

(c) The inequalities \( p_j \geq 1 \) for \( 0 \leq j \leq (n-1)/2 \), \( u_j \geq 1 \) for \( (n+1)/2 \leq j \leq n \), \( u_j < 1 \) for \( 0 \leq j \leq n/2 \), and \( p_j < 1 \) for \( n/2 < j \leq n \) are satisfied on the trajectory \( T_n \).
Thus, \( p_{j,n} u_{j,n} = P_{j}(t_{0,n})U_{0,j}(t_{0,n}) < P_{j}(t_{0,2j+1})U_{0,j}(t_{0,2j+1}) = p_{j,2j+1} u_{j,2j+1} \).

By Lemma 3(b, c), we have \( p_{j,2j+1} u_{j,2j+1} = 1 \cdot p_{j+1,2j+1} \). Now we note that

\[
\Phi(j)(p_{j+1,2j+1}, 1) = \Phi(j)(p_{j+1,2j+1}, u_{j,2j+1}) = (p_{2j+1,2j+1}, u_{2j+1,2j+1}).
\]

This means that \( P_{j}(p_{j+1,2j+1}) = p_{2j+1,2j+1} = 0 \) (by the definition of the trajectory \( T_{2j+1} \)). Because \( t_{1,j} \) is the greatest real root of the polynomial \( P_{j}(\cdot) \) (Lemma 2(b)), we have the inequality \( p_{j+1,2j+1} \leq t_{1,j} \).

Thus, \( p_{j,n} u_{j,n} < p_{j+1,2j+1} \leq t_{1,j} \).

We have \( P_{1,j}(t_{1,j}) = 0 \), while \( P_{1,j}(1) = 1 \), because \( \Phi(j)(1, 1) = (1, 1) \). Proceeding by induction on \( j \), by the monotonicity of the functions \( P_{j} \) on the interval \((t_{1,j-1}, \infty)\) and due to the basis of induction \( t_{1,1} = 1/\sqrt{2} \), we conclude that \( t_{1,j} < 1 \) for all \( j \geq 2 \). As a result, we obtain \( p_{j,n} u_{j,n} < 1 \).

(b) We have \( p_{j}/p_{j-1} = p_{j+1} u_{j} < 1 \) due to (a), and the inequality \( u_{j} < u_{j+1} \) follows from (13).

(c) If \( n = 2k - 1 \), then \( p_{k-1} = 1 \) (Lemma 3(c)); the inequalities \( p_{j} \geq 1 \) for \( 0 \leq j \leq k-1 \) and \( p_{j} < 1 \) for \( k \leq j \leq n \) follow from the monotonicity property (b).

If \( n = 2k \), then \( p_{k} = \sqrt{p_{k-1}/u_{k}} \) by (a). So, \( p_{j} < 1 \) for \( k \leq j \leq n \). We also have \( p_{k-1} = 1/u_{k} > 1 \), and hence \( p_{j} > 1 \) for \( 0 \leq j \leq k-1 \).

The inequalities for \( u_{j} \) follow from the inequalities for \( p_{j} \) by Lemma 3(b). \( \square \)

3.6. Boundedness and the Limits

We put \( \phi = (1 + \sqrt{5})/2 \).

Lemma 5. The sequence \((p_{0,n})\) is bounded: \( p_{0,n} < \phi \) for all \( n \geq 1 \).

Proof. Fixing \( n \), we have \( p_{0} > p_{1} \) by Lemma 4(b). On the other hand, \( p_{1} = p_{0}^{2} - 1 \). This means that \( p_{0}^{2} - p_{0} - 1 < 0 \), whence \( p_{0} < \phi \). \( \square \)

Notation. It follows from Lemmas 2(d) and 5 that the monotone limits

\[
p_{j}^{*} = \lim_{n \to \infty} p_{j,n}, \quad j \geq 0, \\
u_{j}^{*} = \lim_{n \to \infty} u_{j,n} = \frac{1}{p_{j-1}}, \quad j \geq 1,
\]

exist for each \( j = 0, 1, 2, \ldots \). By Lemmas 4(b, c) and 5, we have \( \phi \geq p_{0}^{*} \geq p_{1}^{*} \geq p_{2}^{*} \geq \cdots \geq 1 \) and \( 1/p_{0}^{*} = u_{1}^{*} \leq u_{2}^{*} \leq \cdots \leq 1 \).

Remark. It is intuitively obvious that the points \( P_{s} \) and \( P_{s} \) (see the figure) have coordinates \((p_{0}^{*}, 0)\) and \((0, p_{0}^{*})\), respectively. To prove this, it is first required to have a rigorous definition of invariant curves. This will be done in Sec. 5.4 (see Lemma 9).

Remark. The upper bound \( p_{0,n} < \phi = 1.618 \ldots \) in Lemma 5 is not too far off the numerical upper bound (see (9) in [7]), which is equal to \( p_{0}^{*} = 1/u_{1}^{*} = 1.447 \ldots \).

Notation. We introduce deviations from the fixed point

\[
\lambda_{j,n} = 1 - u_{j,n}.
\]

We will use the brief notation \( \lambda_{j} = \lambda_{j}^{*} \) for a fixed \( n \) and put \( \lambda_{j} = 1 - u_{j}^{*} \). (In [7], precisely these limit values were denoted by \( \lambda_{j} \).)
Eliminating $p_j = p_{j,n}$ from the relations that hold on the trajectory $T_n$,

$$p_j = \frac{1}{u_{j+1}} \quad \text{and} \quad u_{j-1} = u_j^2(p_j + 1) - 1,$$

we obtain the second-order recurrent relations

$$\lambda_{j+1} = \frac{4\lambda_j - \lambda_{j-1} - 2\lambda_j^2}{1 + 2\lambda_j - \lambda_{j-1} - \lambda_j^2}. \quad (16)$$

Passing to the limit as $n \to \infty$, we obtain recurrent relations for $\lambda^*_j$ which coincide with those given in [7] (formula (21)).

The parameters $\lambda_{j,n}$ will be particularly useful in Sec. 5.6.

3.7. Exponential Proximity to the Fixed Point

**Lemma 6.** The coordinates of points of the trajectory $T_n$ satisfy the inequalities

$$|p_j - 1| \leq 2^{-j}, \quad (17)$$
$$0 < 1 - p_j u_j \leq \phi 2^{-j}. \quad (18)$$

Here $0 \leq j \leq n$ and $\langle j \rangle = \min(j, n - j)$.

**Proof.** We consider three cases to prove (17).

(i) Let $n/2 \leq j \leq n - 1$. Then $p_j u_j < 1$ and $p_j < 1$ by Lemma 4 (a, c). This means that

$$1 - p_{j+1} = 2 - p_j^2(u_j + 1) > 2 - p_j - p_j^2 = (1 - p_j)(2 + p_j) > 2(1 - p_j).$$

Because $1 - p_n = 1$, we obtain $1 - p_j < 2^{j-n} = 2^{-\langle j \rangle}$ by induction.

(ii) Let $n$ be odd, and let $j = (n - 1)/2$. In this case, $p_j - 1 = 0$ by Lemma 3 (c).

(iii) Let $0 \leq j < (n - 1)/2$. Then $1 < p_j = u_{n-j} = 1/p_{n-j-1}$. By (i), we have

$$p_j - 1 = p_j(1 - p_{n-j-1}) < 2^{-j-1} p_j.$$

Since $p_j \leq p_0 < \phi < 2$ by Lemmas 4 (b) and 5, the required estimate is proved.

Let us prove (18). By symmetry, it suffices to consider the case $j \leq n/2$; and by (17), $1 \leq p_j \leq \phi$ and $0 \leq 1 - u_j \leq 2^{-j}$. We have

$$0 < 1 - p_j u_j = p_j(1 - u_j) + (1 - p_j) \leq p_j(1 - u_j) \leq \phi 2^{-j},$$

as required.

**Corollary.** The following limits exist:

$$\lim_{j \to \infty} p_j^* = \lim_{j \to \infty} u_j^* = 1 \quad \text{and} \quad \lim_{j \to \infty} \lambda_j^* = 0.$$

**Remark.** These assertions appear in [7] (formulas (13) and (15)) as numerical facts.

**Proof.** The estimates (17) for $|1 - p_{j,n}|$ are uniform in $n$ as $n \geq 2j$. 

MATHEMATICAL NOTES Vol. 110 No. 3 2021
4. CONVERGENCE OF THE SEQUENCE \((C_n)\)

4.1. Constants \(C_n\) in Terms of the Coordinates of Trajectory Points

We fix \(n\), so that \((p_j, u_j)\) denotes a point of the trajectory \(T_n\).

**Lemma 7.** Let \(k = \lfloor n/2 \rfloor\). The constant (4) can be represented as

\[
C_n = 2 \sum_{j=0}^{k-1} (3 - 2p_j - p_ju_j) + p_k^2.
\]

**Proof.** Expressing the minimizing vector \(u\) of the function \(g_n(\cdot)\) in terms of the coordinates of points of the trajectory \(T_n\) (Sec. 2.2) and taking into account the fact that \(1/u_j = p_j - 1\), we obtain

\[
\begin{align*}
A_n &= g_n(u) = \sum_{j=0}^{n} (p_j + u_j + p_ju_j).
\end{align*}
\]

By symmetry (see Lemma 3(b)), we have

\[
A_n = 2 \sum_{j=0}^{k-1} (2p_j + p_ju_j) - \mu,
\]

where \(\mu = 2 = 3 - p_k^2\) for odd \(n\) and \(\mu' = -p_k^2\) for even \(n\). This implies (19). \(\Box\)

4.2. Existence of Shallit’s Constant

We put (using notation (14))

\[
S_N = \sum_{j=0}^{N} (3 - 2p_j^* - p_j^*u_j^*).
\]

**Theorem 1.** The finite limits

\[
S = \lim_{N \to \infty} S_N \quad \text{and} \quad C = \lim_{n \to \infty} C_n
\]

exist, and \(C = 2S + 1\).

**Proof.** We put \(h_j^* = 3 - 2p_j^* - p_j^*u_j^*, \ j \geq 0\). Let \(k_n = \lfloor n/2 \rfloor\). For \(j = 0, \ldots, k_n - 1\), we define \(h_{j,n} = 3 - 2p_j,n - p_j,nu_{j,n}\) and \(z_{j,n} = h_{j,n} - h_j^*\). We also put \(z_{j,n} = 0\) for \(j \geq k_n\).

Writing

\[
h_{j,n} = 2(1 - p_j,n) + (1 - p_j,nu_{j,n}),
\]

and taking Lemma 6 and the inequalities \(1 - p_j,n < 0 < 1 - p_j,nu_{j,n}\) and \(\phi < 2\) into account, we obtain \(|h_{j,n}| < 2^{1-j}\). Therefore, \(|h_j^*| \leq 2^{1-j}\) and the limit \(\lim_{N \to \infty} S_N\) exists.

For each \(j\), we have \(\lim_{n \to \infty} z_{j,n} = 0\). The estimate

\[
\sum_{j=0}^{\infty} |z_{j,n}| \leq \sum_{j=0}^{k_n-1} |h_{j,n}| + \sum_{j=0}^{\infty} |h_j^*| \leq 8
\]

holds uniformly in \(n\).

By the Dominated Convergence Theorem, we conclude that \(\lim_{n \to \infty} \sum_{j=0}^{k_n-1} z_{j,n} = 0\). It remains to refer to Lemma 7 and take into account that \(\lim_{n \to \infty} p_{k_n,n} = 1\) by (17). \(\Box\)
4.3. Monotonicity of the Sequence \((C_n)\)

For completeness, we prove that the sequence \(C_n\) is monotone, which is an observation in [7] (Table 1).

**Proposition 1.** The sequence \((C_n)\) increases monotonically.

**Proof.** We fix \(n\) and, just as elsewhere in this paper, use the one-index notation \(u_j = u_{j,n}\), etc. bearing in mind that \((p_j, u_j)\) are points of the trajectory \(T_n\). The vector at which the function \(g_n(\cdot)\) attains minimum in (6) is \(u = (0, u_1, \ldots, u_n)\). To prove the inequality \(C_{n+1} > C_n\), it suffices to specify a vector \(u^\dagger \in \mathbb{R}^{n+2}\) such that \(u_0^\dagger = 0, u^\dagger = (u_1^\dagger, \ldots, u_{n+1}^\dagger) \in \mathbb{R}^{n+1}\), and \(\delta = g_n(u) + 3 - g_{n+1}(u^\dagger) > 0\).

We put

\[
u_j^\dagger = \begin{cases} u_j, & 1 \leq j \leq k, \\
r, & j = k + 1, \\
u_{j-1}, & k + 2 \leq j \leq n + 1,
\end{cases}
\]

where \(k \in \{1, \ldots, n - 1\}\) and \(r > 0\) are the parameters to be determined. Then

\[
\delta = 3 + \frac{u_k}{u_{k+1}} - \frac{u_k}{r} - \frac{1}{r} - \frac{r}{u_{k+1}} = 3 + p_ku_k - \frac{u_k + 1}{r} - (p_k + 1)r.
\]

We choose \(r = \sqrt{(1 + u_k)/(1 + p_k)}\) so as to maximize the right-hand side. In this case,

\[
\delta = 3 + p_ku_k - 2\sqrt{(1 + p_k)(1 + u_k)}.
\]

The inequality \(\delta > 0\) is equivalent to the inequality

\[
(p_ku_k - 1)^2 + 4(1 - p_k)(1 - u_k) > 0.
\]

We put \(k = \lfloor n/2 \rfloor\). For even \(n\), we have \(p_k = u_k < 1\), and for odd \(n\), we have \(p_k = 1\) (Lemma 3(c)) and \(u_k < 1\) (Lemma 4(c)). The required condition is satisfied in both cases.

**Remark.** From the above proof, we obtain the estimate

\[
C_{n+1} - C_n > \begin{cases} (1 - u_k)^2, & n = 2k, \\
(1 - u_k)^2 \left(1 + o(1)\right), & n = 2k + 1.
\end{cases}
\]

(21)

Theorem 2 states that indeed \(C_n - C_n \sim (1 - u_k)^2\), cf. (25) and (27).

5. SHARP ESTIMATES OF THE RATE OF CONVERGENCE

5.1. Survey

The numerical data given in [7] (Tables 1 and 2) suggest the following conjecture about the asymptotic behavior, and the goal of this section is to justify this conjecture:

\[
p_0 - p_{0,n} \sim \rho^{-n}, \quad C - C_n \sim \rho^{-n}, \quad \lambda_j^* = 1 - u_j^* \sim \rho^{-j}.
\]

Recall (cf. Sec. 3.1) that \(\rho = 2 + \sqrt{3}\) is the larger eigenvalue of the Jacobian matrix \(DF(1,1)\).

We note that if we pass to the limit as \(m \to \infty\) in the inequality

\[
|p_{j,n} - p_{j,m}| \leq |p_{j,n} - 1| + |p_{j,m} - 1|,
\]

then it follows from Lemma 6 that the estimate \(p_{j,n} - p_j = O(2^{-j})\) holds uniformly for \(n \geq 2j\). The similar uniform estimate \(\lambda_{j,n} - \lambda_j^* = O(2^{-j})\) also holds. A slight modification of the argument (by using a linear approximation of the mapping \(F\) near the fixed point) would allow one to replace the base 2 with \(\rho\) in these estimates. In order to obtain a qualified upper bound for the difference \(C_n - C_n\), estimates for
\(|p_{j,n} - p_j^*|\) and \(|u_{j,n} - u_j^*|\) are needed, and it is much more difficult to obtain such estimates. An adequate tool for this purpose is Hartman’s theorem on two-dimensional \(C^1\)-linearization. It provides coordinates in which the mapping \(\Phi\) is linear, while the distortion of the metric in the coordinate transformation admits a two-sided estimate. Using such a linearization, it is not difficult to derive the necessary uniform estimates, both upper and lower, for the coordinate differences along the trajectories if it is known that the “initial” coordinate in the unstable direction is not too small. Geometrically, the latter condition means that the invariant curves intersect the coordinate axes transversally. The transversality is proved in Sec. 5.5.

A proof of the estimate of the difference \(C - C_n\) that would be exact in order still requires some additional effort. The use of expression (19) for \(C_n\) would yield only the estimate \(C - C_n = O(\rho^{-n/2})\). The terms of the series (20) (with \(N = \infty\)) decrease as \(O(\rho^{-j})\). A simple local re-expansion (see (22) in Sec. 5.6) gives a series whose terms decrease as \(O(\rho^{-2j})\), which was already indicated in [7] (formula (20)). Using this representation, one can derive (see the remark after the proof of Theorem 2) the estimate \(C - C_n = O(n\rho^{-n})\) which is still somewhat worse than the rate of convergence observed experimentally. In Lemma 16, we derive the representation of \(C_n\) as a sum whose general term decreases as \(O(\rho^{-3j})\). (The possibility of numerical acceleration of the convergence, which is mentioned in the concluding part of [7], has nothing in common with this.) This is already sufficient for the proof of the main theorem in Sec. 5.7.

5.2. Local Linearization

Hartman’s theorem [9] states that a diffeomorphism of class \(C^2\) defined on an open set in \(\mathbb{R}^2\) with a hyperbolic fixed point is \(C^1\)-conjugate to its linear part in a certain neighborhood of this point. This assertion, in which the fact of being two-dimensional is significant in contrast to the more universal and widely known Grobman–Hartman theorem on the \(C^0\)-linearization, ensures that the conjugating mapping is quasi-isometric, i.e., the ratio of the distances in the old and new coordinate systems lies between two finite positive constants. For more information about local linearization, see [8, Sec. 6.6].

For the mapping \(\Phi\) defined by formulas (9), there exists a neighborhood \(\Omega\) of the fixed point \(P_0 = (1, 1)\) and a \(C^1\)-diffeomorphism \(h\) that takes \(\Omega\) onto a neighborhood of the point \((0, 0) \in \mathbb{R}^2\), \(h\) satisfying \(h(P_0) = (0, 0)\) and

\[h \circ \Phi(p, u) = \Psi \circ h(p, u),\]

where \(\Psi\) is a linear mapping with matrix \(\text{diag}(\rho^{-1}, \rho)\). We write the transformation of coordinates in the form \((\xi, \eta) = h(p, u)\).

If \((\xi', \eta') = h(p', u')\), where \((p', u') = \Phi(p, u)\), then \(\xi' = \rho^{-1}\xi, \eta' = \rho\eta\), i.e., \(\xi\) is a coordinate in the stable direction, and \(\eta\), in the unstable direction.

5.3. Boundary-Value Problem for Trajectories of a Linearized Mapping

By Lemma 6 and its corollary, there exists an \(n_0 \geq 1\) such that if \(n/2 \geq j \geq n_0\), then \((p_{j,n}, u_{j,n}) \in \Omega\) and \((p_{n_0}^{*}, u_{n_0}^{*}) \in \Omega\). We put

\[(\xi_{j,n}, \eta_{j,n}) = h(p_{j,n}, u_{j,n}),\]

\(n_0 \leq j \leq \left\lfloor \frac{n}{2} \right\rfloor\).

(We do not define the values of \(\xi_{j,n}\) and \(\eta_{j,n}\) with \(j \notin \{n_0, \ldots, \lfloor n/2 \rfloor\}\), because they are not needed.)

Now we formulate the boundary-value problem determining the trajectory \(T_n\) in the coordinates \((\xi, \eta)\).

Let \(I_1\) be an interval on the axis \(u = 0\) that contains the point \((p_0^*, 0)\) and is small enough so that \(\Phi^{(n_0)}(I_1) \subset \Omega\). Let \(I_2^p\) and \(I_2^u\) be intervals on the straight lines \(p = 1\) and \(p = u\) that contain the point \((1, 1)\) and lie in \(\Omega\). We put

\[\Gamma_1 = h \circ \Phi^{(n_0)}(I_1),\]

\[\Gamma_2^p = h(I_2^p),\]

\[\Gamma_2^u = h(I_2^u),\]

so that \(\Gamma_1 \cup \Gamma_2^p \cup \Gamma_2^u \subset h(\Omega)\).
Lemma 8. Let \( n > 2n_0 \). We write \( k = \lfloor n/2 \rfloor \) and \( k' = k - n_0 \). The point \( Q = (\xi_{n_0,n}, \eta_{n_0,n}) \in h(\Omega) \) is uniquely determined by the following conditions:

(i) \( Q \in \Gamma_1 \);

(ii) if \( n \) is odd, then \( \Psi^{(k')}(Q) \in \Gamma_2' \); if \( n \) is even, then \( \Psi^{(k')}(Q) \in \Gamma_2'' \).

Proof. The point \( Q^# = (p_{n_0,n}, u_{n_0,n}) = h^{-1}(Q) \) lies on the trajectory \( T_n \) and is uniquely determined by the following conditions:

(i) \( Q^# \) lies in the \( \Phi^{(n_0)} \)-image of the semiaxis \( u = 0, p > 0 \);

(ii) if \( n \) is odd, then \( \Phi^{(k')}(Q^#) \) lies on the line \( p = 1 \); if \( n \) is even, then \( \Phi^{(k')}(Q^#) \) lies on the line \( p = u \) (by Lemma 3(c)).

Applying the mapping \( h \), we obtain the conditions listed in the lemma. \( \square \)

5.4. Invariant Curves

For simplicity, we assume that the neighborhood of zero \( h(\Omega) \) is the square \(|\xi| < \varepsilon, |\eta| < \varepsilon\) and consider segments of the coordinate axes lying in \( h(\Omega) \): \( \gamma_s^# = \{(\xi, \eta) \mid |\eta| = 0, |\xi| < \varepsilon\} \) and \( \gamma_u^# = \{(\xi, \eta) \mid |\xi| = 0, |\eta| < \varepsilon\} \). We see that \( \gamma_s^# \) is a stable and \( \gamma_u^# \) is an unstable invariant curve of the linear mapping \( \Psi \). This means that if \( Q \in \gamma_s^# \) and \( Q' \in \gamma_u^# \), then \( \Psi^{(j)}(Q) \to (0, 0) \) and \( \Psi^{(-j)}(Q') \to (0, 0) \) for \( j \to +\infty \).

Now we can rigorously define invariant curves of the mapping \( \Phi \) which were mentioned in Sec. 3.1. We will only need the stable curve \( \gamma_s \). (The curve \( \gamma_u \) is symmetric to it with respect to the diagonal \( p = u \).)

Definition. The curve \( \gamma_s \) in the \((p, u)\)-plane is the image \((h \circ \Phi^{(n_0)})^{-1}(\gamma_s^#)\).

Lemma 9. The points \((p_j^*, u_j^*)\), \( j = 0, 1, 2, \ldots \), with coordinates (14) lie on \( \gamma_s \).

Proof. By Lemma 6 and its corollary, we have \((p_j^*, u_j^*) \to P_0 \) for \( j \to \infty \). It follows from (14) and the linearity of the mapping \( \Phi \) that \((p_j^*, u_j^*) = \Phi^{(j)}(p_0^*, u_0^*) \). For any \( n > n_0 \), the point \( Q = h(p_n^*, u_n^*) \) lies in \( h(\Omega) \) and has the property \( \lim_{j \to \infty} \Psi^{(j)}(Q) = (0, 0) \) which characterizes the curve \( \gamma_s^# \). Now the assertion of the lemma follows from the definition of the curve \( \gamma_s \). \( \square \)

Remark. By construction, \( \gamma_s \) is a curve of smoothness class at least \( C^1 \). In particular, it has the tangent at the point of intersection with the abscissa axis \( P_s = (p_0^*, 0) \).

Definition. We define extensions of the curve \( \gamma_s \) by iterations of the mapping \( \Phi^{-1} \):

\[ \gamma_s^{(j)} = \{(p, u) \mid p > 0, u > -1, \Phi^{(j)}(p, u) \in \gamma_s\}, \quad j \geq 1. \]

It is clear that \( \gamma_s \subset \gamma_s^{(1)} \subset \gamma_s^{(2)} \subset \cdots \) and \( \Phi(\gamma_s^{(j+1)}) \subset \gamma_s^{(j)} \).

The maximal continuous extension of the curve \( \gamma_s \) is the \( \Phi \)-invariant curve \( \gamma_s^* = \bigcup_{j \geq 1} \gamma_s^{(j)} \) (see the figure). We note that the restriction of the mapping \( \Phi \) to \( \gamma_s^* \) is not surjective: the points with ordinates \( u \leq 0 \) do not belong to \( \Phi(\gamma_s^*) \).
Now we describe an alternative method for constructing the curve $\gamma^*_n$. Starting from the quadrant $D_0 = \{(p, u) \mid p > 0, u > -1\}$, we recurrently define $D_n = \Phi^{-1}(D_{n-1}) \cap D_0$. For example, $D_1 = \{(p, u) \mid p > 0, u > p^{-2} - 1\}$. It is clear that $D_0 \supset D_1 \supset D_2 \supset \ldots$. It turns out that all domains $D_n$ are convex. The set $D_\infty = \bigcap_{n=1}^\infty D_n$ is convex, invariant under $\Phi$, and nonempty: for example, $P_0 = (1,1) \in D_\infty$. The curve $\gamma^*_n$ is its boundary. The following three lemmas justify this construction.

**Lemma 10.** (a) The domain $D_n (n \geq 0)$ is given by the inequalities $u > -1$ and $p > t_{u,n}$, where $t_{u,n}$ is a root of the equation $P_{u,n}(t) = 0$ (Sec. 3.3).

(b) The boundary curve $\partial D_n$ of the domain $D_n$, $n \geq 1$, admits the parametrization $p = U_n(t)$, $u = P_n(t)$, $t > t_{0,n-1}$.

(c) The functions $u \mapsto t_{u,n}$ are decreasing and convex; therefore, the domains $D_n$ are convex.

**Proof.** (a) The case $n = 0$ is trivial. Let $n \geq 1$. For any $u > -1$, the system of $n$ conditions

$$\Phi^{(j)}(p, u) = (P_{u,j}(p), U_{u,j}(p)) \in D_0, \quad 1 \leq j \leq n,$$

is equivalent to the inclusion $(p, u) \in D_n$. But by Lemma 4(b), the same system of conditions is characterized by the inequality $p > t_{u,n}$.

(b) The relation

$$\partial D_1 = \left\{ \left( \frac{1}{t}, t^2 - 1 \right) \mid t > 0 \right\} = \{(U_1(t), P_1(t)) \mid t > t_{0,0}\}$$

is verified directly. Assuming that $n \geq 2$ and the parametrization in question holds for $\partial D_{n-1}$ and taking Lemma 3(a) into account, we obtain

$$\Phi^{-1}(\partial D_{n-1}) = \{(U_n(t), P_n(t)) \mid t > t_{0,n-2}\}.$$  

The boundary of the domain $D_n$, which is a subset of this set, is determined by the conditions $U_n(t) \geq 0$ and $P_n(t) \geq -1$ that follow from the requirement $D_n \subset D_0$. Because $U_n(t) = 1/P_{n-1}(t)$, the condition for the first coordinate has the form $P_{n-1}(t) > 0$, where $t > t_{0,n-2}$. By Lemma 4(b), we conclude that $t > t_{0,n-1}$. The condition for the second coordinate $P_n(t) \geq -1$ is automatically satisfied in this case due to (12).

(c) Using the parametrization given in (b) and the monotonicity of the functions $P_n(\cdot)$ and $U_n(\cdot)$ (Lemma 4(b)), we conclude that $u \mapsto t_{u,n}$ is a decreasing function. The convexity of its graph follows from the inequality

$$\dot{U}_n(t)\ddot{P}_n(t) - \dot{U}_n(t)\dddot{P}_n(t) \geq 0, \quad t > t_{0,n-1},$$

which can be proved by induction. For $n = 0$ and any $t$, we have an equality. Calculating

$$\dot{U}_{n+1} = -\frac{\dot{P}_n}{P_n^2}, \quad \ddot{U}_{n+1} = \frac{2(\dot{P}_n)^2 - P_n \dddot{P}_n}{P_n^3},$$

$$\dddot{P}_{n+1} = P_n^2 \dot{U}_n + 2P_n(U_n + 1)\dddot{P}_n,$$

$$\dddot{P}_{n+1} = 4P_n \dot{P}_n \dot{U}_n + P_n^2 \dddot{U}_n + 2(U_n + 1)(\dot{P}_n)^2 + 2P_n(U_n + 1)\dddot{P}_n,$$

we obtain

$$\dddot{U}_{n+1} \dddot{P}_{n+1} - \dddot{U}_{n+1} \dddot{P}_{n+1} = (\dddot{U}_n \dddot{P}_n - \dddot{U}_n \dddot{P}_n) + \frac{6(\dot{P}_n)^2}{P_n^2}(P_n \dddot{U}_n + (U_n + 1)\dddot{P}_n).$$

By Lemma 4(b), $(P_n U_n)^\cdot + \dot{P}_n > 0$ for $t > t_{0,n-1}$, and the induction step is complete. \hfill $\Box$

**Lemma 11.** The boundary $\partial D_\infty$ of the set $D_\infty$ is the graph of the decreasing convex function defined on $(0, \infty)$ and inverse to the function $p = t_{u,\infty}$, where $t_{u,\infty} = \sup_n t_{u,n}$.
Proof. We will show that \( u \mapsto t_{u,\infty} \) is a finite nonincreasing function and

\[ D_\infty = \{(p, u) \mid u > -1, p \geq t_{u,\infty}\}. \]

By Lemma 10(a), we have: \((p, 0) \in D_n\) if and only if \(p > p_{0,n}\). Passing to the limit as \(n \to \infty\), we conclude that \((p_0^*, 0) \in \partial D_\infty\). It follows from Lemma 10(c) that \(\sup_n t_{u,n} \leq p_0^*\) for \(u \geq 0\).

Now we assume that \(-1 < u < 0\) and \(\tau = (1/(u+1))^{1/2}\). Let us show that \(t_{u,n} < \tau\). Assume the converse. Then \(\Phi(t_{u,n}, u) = (t_{u,n}^2/(u+1) - 1, 1/t_{u,n})\) lies in the domain \(\{(p, u) \mid p \geq p_0^*, u > 0\}\) and hence in the domain \(D_\infty\) (by the already analyzed case \(u > 0\)). Therefore, the iterations of the point \((t_{u,n}, u)\) never reach the line \(p = 0\), and this is a contradiction. So \(\sup_n t_{u,n} < \infty\) for any \(u > -1\).

By Lemma 4(b), we have \(t_{u,\infty} = \lim_{n \to \infty} t_{u,n}\). The monotonicity of the function \(u \mapsto t_{u,\infty}\) follows from Lemma 10(c) by passing to the limit. The same applies to its convexity. It remains to prove that the inverse function is defined on \((0, \infty)\), and for this, it suffices to show that (i) \(t_{u,\infty} \to 0\) as \(u \to +\infty\) and (ii) \(t_{u,\infty} \to +\infty\) as \(u \to -1^+\).

Assume that (i) is false. Then \(t_{u,\infty} > \tau\) for some \(\tau > 0\) and all \(u > -1\). Let \(u\) be large enough so that \(\tau^2(u+1) - 1 > t_{1,\tau,n}\). Then \(\tau^2(u+1) - 1 > t_{1,\tau,n}\) for all \(n\). Therefore,

\[ \Phi(\tau, u) = (\tau^2(u+1) - 1, 1/\tau) \in D_n \]

by Lemma 10(a). But then \((\tau, u) \in D_{n+1}\) for any \(n\), and hence \(\tau \geq \sup_{n \geq 1} t_{u,n} = t_{u,\infty}\), which is a contradiction.

Assume that (ii) is false. Then \(t_{u,\infty} < \tau\) for some finite \(\tau > 0\) and all \(u > -1\). If \(u\) is close to \(-1\) so that \(\tau^2(u+1) - 1 < t_{1,\tau,n}\), then \((\tau, u) \in D_\infty\) but \(\Phi(\tau, u) \notin D_\infty\), which is a contradiction. \(\square\)

**Lemma 12.** The curve \(\partial D_\infty\) is invariant, i.e., \(\Phi(\partial D_\infty) \subset \partial D_\infty\). The points \((p_j^*, u_j^*)\), \(j = 0, 1, 2, \ldots\), defined in Sec. 3.6 lie on \(\partial D_\infty\).

**Proof.** We have \(\Phi(D_n) \subset D_{n-1}\). Therefore, \(\Phi(D_\infty) \subset D_\infty\). Let \((p, u) \in \partial D_\infty\). Then \(p = t_{u,\infty} = \lim_{n \to \infty} t_{u,n}\).

Let us put \((q, v) = \Phi(p, u)\) and \((q_n, v_n) = \Phi(t_{u,n}, u)\). Then we have

\[ \Phi^{(n-1)}(q_n, v_n) = \Phi^{(n)}(t_{u,n}, u) = (0, *) \notin D_0, \]

which means that \((q_n, v_n) \notin D_\infty\). Therefore, \((q, v) = \lim_{n \to \infty} (q_n, v_n)\) is not an interior point of the domain \(D_\infty\). We conclude that \(\Phi(\partial D_\infty) \subset \partial D_\infty\).

Since \(p_0^* = t_{0,\infty}\), we have \((p_0^*, 0) \in \partial D_\infty\). It follows from the above-proved invariance that \((p_j^*, u_j^*) \in \partial D_\infty\) for all \(j \geq 0\). \(\square\)

**Lemma 13.** The curve \(\partial D_\infty\) coincides with the invariant curve \(\gamma_\ast^s\).

**Proof.** Let \(\Omega\) be the neighborhood of the point \((1, 1)\) chosen in Sec. 5.4, and let \(\Omega' \subset \Omega\) be a neighborhood so small that

\[ \Phi(\Omega') \cup \Phi^{-1}(\Omega') \subset \Omega. \]

In \(\Omega'\), there exist precisely two invariant curves, namely, parts of the curves \(\gamma_s\) and \(\gamma_u\). (More precisely, \(\Phi(\gamma_s \cap \Omega') \subset \gamma_s \cap \Omega'\) but \(\Phi(\gamma_u \cap \Omega') \subset \gamma_u \cap \Omega\).)

It follows from the continuity of the mapping \(\Phi\) at the point \((1, 1)\) and the \(\Phi\)-invariance of the curve \(\partial D_\infty\) that there exists a neighborhood \(\Omega'' \subset \Omega'\) of the point \((1, 1)\) such that

\[ \Phi(\partial D_\infty \cap \Omega'') \subset \partial D_\infty \cap \Omega. \]

By Lemma 12, the curve \(\partial D_\infty \cap \Omega''\), just as \(\gamma_s \cap \Omega''\), contains all points \((p_j^*, u_j^*)\) with sufficiently large subscripts \(j \geq j_0\). By convexity, the curve \(\partial D_\infty\) has a right tangent at the point \((1, 1)\) which coincides with the tangent to \(\gamma_s\). The left tangent to \(\partial D_\infty\) cannot coincide with the tangent to \(\gamma_u\), because the slope
of the latter is less steep, and this contradicts the convexity. Therefore, the segments of the curves \( \partial D_\infty \) and \( \gamma_s \) in the neighborhood \( \Omega' \) coincide on both sides. Acting by iterations of the mapping \( \Phi^{-1} \) on the coinciding segments as long as the resulting points remain in the quadrant \( D_0 \) (this formally means that we construct \( \Phi \)-invariant extensions as in the definition of the curve \( \gamma_s^* \)), we conclude that \( \partial D_\infty = \gamma_s^* \).

From Lemmas 11 and 13, we immediately obtain a result that is technically important for estimating the proximity between the trajectories \( T_n \) and the invariant curves and is used in the proof of assertion (b) of Theorem 2.

**Lemma 14.** The invariant curve \( \gamma_s \) transversally intersects the abscissa axis at the point \( P_s \). Appropriately, the curves \( \Gamma_1 \) and \( \gamma_s^\# \) defined in Sec. 5.3 transversally intersect at the point \( P_s^\# \).

### 5.6. Representation of \( C_n \) by a Sum with Rapidly Decreasing Terms

Here we deal with a fixed trajectory \( T_n \). First, we recall the notation \( \lambda_{j,n} = 1 - u_{j,n} \) introduced in (15). Throughout this section, \( k = \lfloor n/2 \rfloor \).

We rewrite formula (19) in terms of deviations,

\[
C_n = 2 \sum_{j=1}^{k} \frac{\lambda_{j-1} - 3\lambda_j}{1 - \lambda_j} + p_k^2. \tag{19'}
\]

Here the terms decrease at a linear rate with respect to the quantities \( \lambda_j \) (for \( j \gg 1 \)), which are small. As the first step, in the next lemma, we derive a similar expression, where the terms decrease as \( \lambda_j^2 \). Such a rate of convergence is not yet sufficient to prove Theorem 2 (see the remark at the end of Sec. 5.7), and we will subsequently obtain an expression whose terms decrease at a cubic rate.

**Lemma 15.** Representation (19) of the constant \( C_n \) can be transformed as

\[
C_n = 2 \sum_{j=0}^{k-1} (3 - p_j + u_j - 2u_{j+1} - p_j u_j) + \delta = 2 \sum_{j=1}^{k} \frac{\lambda_j (\lambda_{j-1} - 2\lambda_j)}{u_j} + \delta, \tag{22}
\]

where \( \delta = 1 - (p_k - 1)^2 \).

**Proof.** Summing identity (19), the telescopic identity

\[
2 \sum_{j=0}^{k-1} (u_j - u_{j+1}) + 2u_k = 0,
\]

and the identity in Lemma 1(c) with \( \ell = k \), we obtain the first line in (22) with the term \( \delta \) in the form

\[
\delta = p_k^2 - 2p_k u_k + 2u_k.
\]

For even \( n \), we have \( u_k = p_k \), and for odd \( n \), we have \( p_k = 1 \); the relation \( \delta = 1 - (p_k - 1)^2 \) combines both of these cases.

To obtain the second line in (22), we substitute \( p_j = 1/u_{j+1} \) which yields

\[
3 - \frac{1}{u_{j+1}} + u_j - 2u_{j+1} - \frac{u_j}{u_{j+1}} = \left(\frac{u_{j+1}}{u_j} - u_j - 2u_{j+1} + 1\right) u_{j+1}.
\]

The desired expression differs only in the shift of the summation index \( j + 1 \mapsto j \). 

In the following lemma, which is the main lemma in this section, the structure of the expression is important, but not the explicit form of the terms. We introduce a version of asymptotic \( O \)-symbols, which allows us to avoid cumbersome explicit formulas.
**Notation.** If \( f \) is a rational function of three variables such that

\[
f(\lambda_{j-1}, \lambda_j, \lambda_{j+1}) \leq \text{const} \cdot \max(\lambda_{j-1}^d, \lambda_j^d, \lambda_{j+1}^d)
\]

for \( 0 \leq \lambda_{j-1} \leq \lambda_j^*, 0 \leq \lambda_j \leq \lambda_j^*, \) and \( 0 \leq \lambda_{j+1} \leq \lambda_{j+1}^* \), then we write \( f(\vec{\lambda}_j) = O(\lambda_j^d) \), where \( \vec{\lambda}_j = (\lambda_{j-1}, \lambda_j, \lambda_{j+1}) \).

This notation will be used below with the values \( \lambda_j = \lambda_{j,n} = 1 - u_{j,n} \) corresponding to the trajectories \( T_n \). We note that \( \lambda_{j,n} \leq \lambda_j^* \) for all \( n \), cf. (14).

**Lemma 16.** There exist functions \( f \) and \( r \) of three variables such that

\[
f(\vec{\lambda}_j) = O(\lambda_j^3) \quad r(\vec{\lambda}_j) = O(\lambda_j^3) \quad \text{and} \quad C_n = \lambda_0 \lambda_1 + 1 + \sum_{j=1}^{k} f(\vec{\lambda}_j) + r(\vec{\lambda}_k),
\]

where the values \( \lambda_j = \lambda_{j,n} \) are taken on the trajectory \( T_n \).

**Proof.** It follows from the recurrent relations (16) that

\[
\lambda_{j-1} + \lambda_{j+1} = 4\lambda_j + O(\lambda_j^2).
\]

This implies

\[
2 \sum_{j=1}^{k} \lambda_{j-1} \lambda_j = \sum_{j=1}^{k} \lambda_{j-1} \lambda_j + \sum_{j=0}^{k-1} \lambda_j \lambda_{j+1} = \lambda_0 \lambda_1 + \sum_{j=1}^{k} (\lambda_{j-1} + \lambda_{j+1}) \lambda_j - \lambda_k \lambda_{k+1}
\]

\[
= \lambda_0 \lambda_1 + \sum_{j=1}^{k} (4\lambda_j^2 + O(\lambda_j^3)) - O(\lambda_k^2).
\]

Since \( 1/u_j = 1 + O(\lambda_j) \), we can rewrite formula (22) as

\[
C_n = 2 \sum_{j=1}^{k} (\lambda_{j-1} \lambda_j - 2\lambda_j^2 + O(\lambda_j^3)) + 1 + O(\lambda_k^2).
\]

Substituting the preceding identity into this relation, we obtain the desired formula. \[\square\]

**Remark.** The replacement of \( O \)-estimates by explicit expressions in the above proof leads in the limit as \( n \to \infty \) to the following formula for \( C \) with general term of order \( O(\rho^{-3j}) \):

\[
C = p_0^* + \sum_{j=1}^{\infty} \frac{\lambda_j^*^2 (\lambda_{j+1}^* - 2\lambda_j^* + \lambda_{j+1}^* \lambda_{j+1}^*)}{w_j^* w_{j+1}^*}.
\]

As a matter of fact, the \( j \)th term is independent of \( \lambda_j^* \), but this does not follow from Lemma 16.

5.7. **Main Theorem**

**Theorem 2.** The following asymptotic relations hold (where the hidden constants are independent of \( n \) and \( j \)):

(a) the distance from the points of the trajectory \( T_n \) to the fixed point is estimated as

\[
|p_{j,n} - 1| + |u_{j,n} - 1| \approx \rho^{-j}, \quad 0 \leq j \leq \frac{n}{2}, \tag{24}
\]

and for \( k = [n/2] \),

\[
|u_{k,n} - 1| \approx \rho^{-n/2}; \tag{25}
\]

\[\text{MATHEMATICAL NOTES} \quad \text{Vol. 110 No. 3} \quad 2021\]
(b) the distance from the points of the trajectory $T_n$ to the invariant curve $\gamma_s$ is estimated as
\[
|p_{j,n} - p_{j}^*| + |u_{j,n} - u_j^*| \asymp \rho^{j-n}, \quad 0 \leq j \leq \frac{n}{2};
\]

in particular,
\[
|p_{0,n} - p_0^*| \asymp \rho^{-n};
\]

(c) the rate of convergence of the sequence $(C_n)$ has the asymptotics
\[
C - C_n \asymp \rho^{-n}.
\]

**Proof.** It suffices to prove (24) and (26) for $j \geq n_0$ and under the assumption that $n > 2n_0$, where $n_0$ is as in Sec. 5.3.

Since the linearizing homeomorphism $h$ is of class $C^1$, one can, as was noted in Sec. 5.2, perform the estimates in the coordinates $(\xi, \eta)$, where the mapping $\Phi$ is linear.

Throughout the proof, we assume that $k = \lfloor n/2 \rfloor$.

(a) Let $d_{j,n} = |\xi_{j,n}| + |\eta_{j,n}|$. Then $d_{j,n} \asymp |p_{j,n} - 1| + |u_{j,n} - 1|$. It follows from Lemma 8(ii) that the point $(\xi_{k,n}, \eta_{k,n})$ lies either on the curve $\Gamma_2$ or on $\Gamma_2'$ (depending on whether $n$ is even or not). Both of these curves are transversal to the axes of the coordinate plane $(\xi, \eta)$ at the origin, because their images $I_2$, $I_2'$ on the plane $(p, u)$ are transversal to the stable and unstable directions of the mapping $\Phi$ at the fixed point. Therefore, there exists a constant $m_1 > 0$ such that $|\xi_{k,n}| \geq m_1 d_{k,n}$.

By the linearity of the mapping $\Psi$, we have
\[
d_{j,n} = \rho^{k-j} |\xi_{k,n}| + \rho^{j-k} |\eta_{k,n}|.
\]

Therefore (we put $M_1 = 1$ to make both sides of the inequalities look alike),
\[
m_1 \rho^{k-j} d_{k,n} \leq d_{j,n} \leq M_1 \rho^{k-j} d_{k,n}, \quad n_0 \leq j \leq k.
\]

Further, since the limit $\lim_{n \to \infty} (\xi_{n_0,n}, \eta_{n_0,n}) = h(p_{n_0,n}, u_{n_0,n}) \neq (0, 0)$ exists, it follows that we have $m'_1 < d_{n_0,n} < M'_1$ with some $0 < m'_1 < M'_1$ independent of $n$. Thus,
\[
m'_2 \rho^{-j} < d_{j,n} < M''_2 \rho^{-j},
\]

where $m''_1 = (m_1 m'_1 / M_1) \rho^{n_0}$ and $M''_1 = (M_1 M'_1 / m_1) \rho^{n_0}$. The estimate (24) is proved.

The estimate (25) holds due to the inequality $|p_k - 1| \leq |u_k - 1|$. (Cf. Lemmas 3(c) and 4.)

(b) We put $d_{j,n}^* = |\xi_{j,n} - \xi_j^*| + |\eta_{j,n}|$. Recall that (see Sec. 5.4) $x, 0) = h(p_j^*, u_j^*) \in \gamma_s^#$ and $(\xi_{n_0,n}, \eta_{n_0,n}) \in \Gamma_1$ by Lemma 8(i). It follows from the transversality of $\Gamma_1$ and $\gamma_s^#$ at the point $(\xi^*, 0)$ (Lemma 14) that $|\eta_{n_0,n}| \geq m_2 d_{n_0,n}^*$ with a certain constant $m_2 > 0$. Since
\[
d_{j,n}^* = \rho^{n_0-j} |\xi_{n_0,n} - \xi_{n_0}^*| + \rho^{j-n_0} |\eta_{n_0,n}|,
\]

putting $M_2 = 1$, we obtain
\[
m_2 \rho^{j-n_0} d_{n_0,n}^* \leq d_{j,n}^* \leq M_2 \rho^{j-n_0} d_{n_0,n}^*, \quad n_0 \leq j \leq k.
\]

Repeating the argument used at the beginning of the proof of assertion (a), we conclude that the ratio $|\eta_{k,n}| / d_{k,n}$ is bounded away from zero. By (24), we have $d_{k,n} \asymp \rho^{-k}$. Therefore, $d_{k,n} \geq |\eta_{k,n}| \geq m'_2 \rho^{-k}$ with a certain constant $m'_2 > 0$.

Further, $\xi_k^* = \xi_{n_0} \rho^{n_0-k}$. Therefore, $d_{k,n}^* \leq d_{k,n} + |\xi_k^*| \leq M'_2 \rho^{-k}$ with a certain constant $M'_2 > 0$. So
\[
m'_2 \rho^{-k} \leq d_{k,n}^* \leq M'_2 \rho^{-k}.
\]
Taking (28) into account, we obtain
\[ m_2'' \rho^{j-2k} \leq d_{j,n}^* \leq M_2'' \rho^{j-2k}, \]
where \( m_2'' = m_2m_2'/M_2 \) and \( M_2'' = M_2M_2'/m_2 \). Since
\[ n - 1 \leq 2k \leq n \quad \text{and} \quad |p_{j,n} - p_j^0| + |u_{j,n} - u_j^0| \approx d_{j,n}^*, \]
the estimate (26) is proved.

(c) The lower bound for \( C - C_n \) follows from inequalities (21) and (25). We will prove that
\[ C - C_n = O(\rho^{-n}). \]

Let \( f \) and \( r \) be the functions in Lemma 16. We take \( N > n \) and \( K = \lfloor N/2 \rfloor \). We put
\[ \Delta_{n,N}^{(1)} = \lambda_{0,N} \lambda_{1,N} - \lambda_{0,n} \lambda_{1,n}, \quad \Delta_{n,N}^{(2)} = \sum_{j=1}^{k} (f(\bar{\lambda}_{j,N}) - f(\bar{\lambda}_{j,n})), \]
\[ \Delta_{n,N}^{(3)} = \sum_{j=k+1}^{K} f(\bar{\lambda}_{j,N}), \quad \Delta_{n,N}^{(4)} = r(\bar{\lambda}_{K,N}) - r(\bar{\lambda}_{k,n}). \]

By Lemma 16, we have the identity
\[ C_N - C_n = \Delta_{n,N}^{(1)} + \Delta_{n,N}^{(2)} + \Delta_{n,N}^{(3)} + \Delta_{n,N}^{(4)}. \]

Let us estimate the right-hand side.

(1) From (26) and the triangle inequality, we obtain \( \lambda_{j,N} - \lambda_{j,n} = O(\rho^{j-n}) \) for \( 0 \leq j \leq k \). Taking \( j = 0 \) and \( j = 1 \), we have
\[ \Delta_{n,N}^{(1)} = (\lambda_{0,N} - \lambda_{0,n}) \lambda_{1,N} + \lambda_{0,n} (\lambda_{1,N} - \lambda_{1,n}) = O(\rho^{-n}). \]

(2) Since \( f \) is a rational function of a vector argument and \( f(\bar{\lambda}_{j,n}) = O(\lambda_{j,N}^3) \), we have
\[ \| \nabla f(\bar{\lambda}_{j,n}) \| = O(\lambda_{j,N}^2), \]
and hence
\[ |f(\bar{\lambda}_{j,N}) - f(\bar{\lambda}_{j,n})| = O(\lambda_{j,N}^2) + O(\lambda_{j,n}^2) \| \bar{\lambda}_{j,N} - \bar{\lambda}_{j,n} \|. \]

By (24), the first factor in the last term on the right is \( O(\rho^{j-n}) \) and, as was noted above, the second factor is \( O(\rho^{-n}) \). Summing over \( j \) from 1 to \( k \), we obtain \( \Delta_{n,N}^{(2)} = O(\rho^{-n}) \).

(3) Again using (24) and estimating the sum of decreasing geometric progression, we obtain
\[ \Delta_{n,N}^{(3)} = O(\rho^{-3k}) = O(\rho^{-3n/2}). \]

(4) By (25) and the estimate for \( r(\cdot) \) in Lemma 16, we have \( r(\bar{\lambda}_{k,n}) = O(\rho^{-2k}) = O(\rho^{-n}) \) and, similarly, \( r(\bar{\lambda}_{K,N}) = O(\rho^{-N}) \). Therefore, \( \Delta_{n,N}^{(4)} = O(\rho^{-n}) \).

The obtained estimates are uniform in \( N \) for \( N > n \). Theorem 1 implies that \( C = \lim_{N \to \infty} C_N \). Therefore,
\[ |C - C_n| \leq \sup_{N > n} |C_N - C_n| = O(\rho^{-n}). \]

The proof is complete. \( \square \)

Remark. When using Lemma 15 instead of Lemma 16, instead of \( f \) we consider a function from (22) (we denote it by \( \widetilde{f} \)) satisfying the estimate \( \widetilde{f}(\bar{\lambda}_{j,n}) = O(\lambda_{j,n}^2) \). The terms in the definition of the quantity \( \tilde{\Delta}_{n,N}^{(2)} \) corresponding to the quantity \( \Delta_{n,N}^{(2)} \) are of order \( O(\rho^{-n}) \) (rather than \( O(\rho^{-n-j}) \)) for \( j = 1, \ldots, k \). This implies the estimate \( \tilde{\Delta}_{n,N}^{(2)} = O(n\rho^{-n}) \) and the same nonoptimal asymptotic estimate for the difference \( C - C_n \).
ACKNOWLEDGMENTS

The author wishes to express his gratitude to the referee for the careful reading of the manuscript, valuable comments, and the suggestions for improving the text.

REFERENCES

1. J. Shallit, “A minimization problem,” SIAM Rev. 36 (3), 490–491 (1994).
2. S. Finch, Mathematical Constants (Princeton Univ. Press, Princeton, 2003).
3. The On-Line Encyclopedia of Integer Sequences A086276, https://oeis.org/A086276.
4. R. L. Brown, K. R. Davidson, and J. Shallit, “Problem 10433,” Amer. Math. Monthly 102 (2), 170 (1995).
5. S. Sadov, “Lower bound for cyclic sums of Diananda type,” Arch. Math. (Basel) 106 (2), 135–144 (2016) doi 10.1007/s00013-015-0853-3.
6. S. Sadov, Minimization of the Sum Under Product Constraints, arXiv:2012.15517(2020).
7. C. C. Grosjean and H. E. De Meyer, “Solution of Problem 94-15: A minimization problem,” SIAM Rev. 37 (3), 451–458 (1995).
8. A. B. Katok and B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems (Cambridge Univ. Press, Cambridge, 1995).
9. P. Hartman, “On local homeomorphisms of Euclidean spaces,” Bol. Soc. Mat. Mexicana (2) 5, 220–241 (1960).