CONDITIONING AND RELATIVE ERROR PROPAGATION IN LINEAR AUTONOMOUS ORDINARY DIFFERENTIAL EQUATIONS

Stefano Maset

Dipartimento di Matematica e Geoscienze
Università di Trieste
Via Valerio 12/1, 34127, Trieste, Italy

Abstract. In this paper, we study the relative error propagation in the solution of linear autonomous ordinary differential equations with respect to perturbations in the initial value. We also consider equations with a constant forcing term and a nonzero equilibrium. The study is carried out for equations defined by normal matrices.

1. Introduction. Consider the scalar real Ordinary Differential Equation (ODE)

\[
\begin{aligned}
&y'(t) = ay(t), \quad t \geq 0, \\
y(0) = y_0
\end{aligned}
\]

where \(a\) is a given real number. If the initial value \(y_0 \neq 0\) is perturbed to \(\tilde{y}_0\) with relative error

\[
\varepsilon = \frac{|\tilde{y}_0 - y_0|}{|y_0|},
\]

the value \(y(t) = e^{at}y_0\) of the solution at the time \(t\) is perturbed to \(\tilde{y}(t) = e^{at}\tilde{y}_0\) with relative error

\[
\delta(t) = \frac{|\tilde{y}(t) - y(t)|}{|y(t)|}.
\]

Whereas the relation between the absolute errors \(|\tilde{y}(t) - y(t)|\) and \(|\tilde{y}_0 - y_0|\) is

\[
|\tilde{y}(t) - y(t)| = e^{at} |\tilde{y}_0 - y_0|, \quad t \geq 0,
\]

the relation between the relative errors \(\delta(t)\) and \(\varepsilon\) is

\[
\delta(t) = \varepsilon, \quad t \geq 0.
\]

So, whereas the absolute error of the perturbation grows or decays with time in dependence of the sign of \(a\), the relative error does not change with time, for all \(a\), \(y_0 \neq 0\) and \(\tilde{y}_0\).

This shows that the absolute and relative errors of the perturbed solution behave very differently.

2010 Mathematics Subject Classification. Primary: 65F35, 65F60; Secondary: 34D10.

Key words and phrases. Linear ordinary differential equations, relative error, condition numbers.

The first author is supported in part by the GNCS of the italian “Istituto Nazionale di Alta Matematica”.

2879
Now, we want to consider the same question for the \( n \)-dimensional real linear autonomous ODE
\[
\begin{aligned}
y'(t) &= Ay(t), \quad t \geq 0, \\
y(0) &= y_0,
\end{aligned}
\tag{1}
\]
where \( A \) is a given \( n \times n \) real matrix. Let \( \| \cdot \| \) be a norm on \( \mathbb{R}^n \). If the initial value \( y_0 \in \mathbb{R}^n \setminus \{0\} \) is perturbed to \( \tilde{y}_0 \) with relative error
\[
\varepsilon = \frac{\|y_0 - \tilde{y}_0\|}{\|y_0\|},
\]
the value \( y(t) = e^{tA}y_0 \) of the solution at the time \( t \) is perturbed to \( \tilde{y}(t) = e^{tA}\tilde{y}_0 \) with relative error
\[
\delta(t) = \frac{\|\tilde{y}(t) - y(t)\|}{\|y(t)\|}.
\]
Observe that \( y_0 \neq 0 \) implies, for all \( t \geq 0 \), \( y(t) \neq 0 \) and so \( \delta(t) \) is defined.

As for the absolute errors \( \|\tilde{y}(t) - y(t)\| \) and \( \|\tilde{y}_0 - y_0\| \), the following two inequalities are well known (see e.g. [8]):
\[
\|\tilde{y}(t) - y(t)\| \leq M(t)e^{\alpha(A)t}\|\tilde{y}_0 - y_0\|, \quad t \geq 0,
\]
where \( M(t) \) grows polynomially with \( t \) and \( \alpha(A) \) is the spectral abscissa of \( A \) defined as the maximum real part of the eigenvalues of \( A \), and
\[
\|\tilde{y}(t) - y(t)\| \leq e^{\mu(A)t}\|\tilde{y}_0 - y_0\|, \quad t \geq 0,
\]
where \( \mu(A) \) is the logarithmic norm of \( A \) defined as
\[
\mu(A) = \lim_{h \to 0^+} \frac{\|I + hA\| - 1}{h}.
\]

There are specific ways for computing \( \mu(A) \) in the usual cases where the norm \( \| \cdot \| \) is the 1-norm, the 2-norm or the \( \infty \)-norm (see e.g. [5]). For example, in case of the 2-norm, the relevant logarithmic norm \( \mu_2(A) \) is the maximum eigenvalue of the symmetric part \( \frac{1}{2}(A^T + A) \) of \( A \).

Despite its basic importance, it seems that the relation between the relative errors \( \delta(t) \) and \( \varepsilon \) has not been yet investigated in literature. Indeed, any perturbation analysis for the ODE (1) found in literature considers the relative error of the perturbed matrix exponential \( e^{tA} \) with respect to \( e^{tA} \), when \( A \) is perturbed to \( \tilde{A} \): see [8], [4], [6], [1] and [7]. Moreover, in these papers the role of the particular initial value \( y_0 \) is ignored since they deal with the matrix \( e^{tA} \) rather than the vector \( e^{tA}y_0 \).

In Section 2 below, we study the relation between \( \delta(t) \) and \( \varepsilon \) by using the 2-norm as norm \( \| \cdot \| \) and by assuming that \( A \) is a normal matrix. In the successive Section 3, the same analysis is accomplished for a linear autonomous ODE with a forcing term and a nonzero equilibrium.

Observe that the relative errors \( \delta(t) \) and \( \varepsilon \) are measured on vectors, not on their components. However, for a vector \( u \in \mathbb{R}^n \) with nonzero components and a perturbed vector \( \tilde{u} \in \mathbb{R}^n \), we have
\[
\frac{\|\tilde{u} - u\|}{\|u\|} \leq \max_{i=1,\ldots,n} \frac{|\tilde{u}_i - u_i|}{|u_i|}
\]
and
\[
\frac{|\tilde{u}_i - u_i|}{|u_i|} \leq \frac{\|u\|}{|u_i|} \cdot \frac{\|\tilde{u} - u\|}{\|u\|}, \quad i = 1, \ldots, n,
\]
whenever the norm \( \| \cdot \| \) is such that
\[
|v_i| \leq \|v\|, \ v \in \mathbb{R}^n \text{ and } i = 1, \ldots, n,
\]
\[
\|v\| = \| (|v_1|, \ldots, |v_n|) \|, \ v \in \mathbb{R}^n,
\]
\[
\|v\| \leq \|w\|, \ v, w \in \mathbb{R}^n \text{ with } |v_i| \leq |w_i| \text{ for } i = 1, \ldots, n.
\]

This holds for a \( p \)-norm.

Thus, relative error on vectors and relative errors on components are strictly related. For an initial value \( y_0 \) with nonzero components \( y_{0i} \) perturbed to \( \tilde{y}_0 \) with components \( \tilde{y}_{0i} \): if
\[
\max_{i=1, \ldots, n} \left| \frac{\tilde{y}_{0i} - y_{0i}}{y_{0i}} \right| \leq \text{TOL},
\]
i.e. all components of \( y_0 \) are perturbed with relative errors within a tolerance \( \text{TOL} \), then \( \varepsilon \leq \text{TOL} \); if (an estimate of) \( \delta(t) \) is known, then we have information on the relative errors of the components of \( y(t) \) by
\[
\delta(t) \leq \max_{i=1, \ldots, n} \left| \frac{\tilde{y}_i(t) - y_i(t)}{y_i(t)} \right| \cdot \delta(t), \ i = 1, \ldots, n.
\]

In numerical analysis, the term “conditioning” refers to how a perturbation in the data of a problem influences the result of the problem. Conditioning studies consider relative errors, rather than absolute errors, because relative errors quantify the loss of accuracy, in terms of significant figures, introduced by a perturbation. The magnification of the relative error in the result with respect to the relative error in the data is ruled by the so-called “condition numbers”.

In the next sections, we present a conditioning analysis with respect to the initial value of linear autonomous ODEs.

2. Condition numbers with respect to the initial value. We begin by presenting an illustrating example.

Example 2.1. Consider the ODE (1) with the symmetric matrix
\[
A = \begin{bmatrix} -2.505 & 2.495 \\ 2.495 & -2.505 \end{bmatrix},
\]
the initial value
\[
y_0 = (0.9, -0.7)
\]
and the perturbed initial value
\[
\tilde{y}_0 = y_0 + (0.01, -0.01) = (0.91, -0.71).
\]

In Figure 1 left, we see the graph of the absolute error \( \| \tilde{y}(t) - y(t) \|_2 \) for \( t \in [0, 3] \). The values \( y(t) \) and \( \tilde{y}(t) \) are computed by using the MATLAB function expm for the matrix exponential. We observe a fast decrease of the absolute error in time. The decrease can be explained by noticing that the eigenvalues \( \lambda_1 \) and \( \lambda_2 \) of \( A \) are negative:
\[
\lambda_1 = -0.01 \text{ and } \lambda_2 = -5.
\]
Observe that we have
\[
\| \tilde{y}(t) - y(t) \|_2 \leq e^{-0.01t} \| \tilde{y}_0 - y_0 \|_2, \ t \geq 0,
\]
since $\mu_2(A) = \lambda_1$, but (3) is not sharp in this situation.

In Figure 1 right, we see the graph of the relative error $\delta(t) = \frac{\|\tilde{y}(t) - y(t)\|_2}{\|y(t)\|_2}$ for $t \in [0, 3]$. Also the relative error has a fast decrease in time.

In Figure 2, we see the graphs of the absolute and relative errors on the interval $[0, 3]$ for the new perturbed initial value 

\[
\tilde{y}_0 = y_0 + (0.01, 0.01) = (0.91, -0.69).
\]
The left side shows a slow decrease of the absolute error, more coherent with the inequality (3). On the other hand, the right side shows a fast grow of one order of magnitude for the relative error. Indeed, at $t = 3$ we have

$$y(t) = (0.0970, 0.0970).$$
Figure 3. Propagation of absolute and relative errors for $y_0 = (1, -1)$ and $\tilde{y}_0 = (1.01, -0.99)$.

So, whereas the absolute error is more or less equal to the initial one as it is shown in Figure 2, $y(t)$ is almost 10 times smaller than the initial value $y_0$. This explains the growth of one order of magnitude for the relative error.

The relative error can behave even worse. Consider the new initial value $y_0 = (1, -1)$
and the perturbed initial value
\[ \tilde{y}_0 = y_0 + (0.01, 0.01) = (1.01, -0.99). \]
The graphs of the absolute and relative errors on the interval \([0, 3]\) are presented in Figure 3. Along with a very slow decrease of the absolute error very similar to that in Figure 2, we have an explosion of six order of magnitude of the relative error. Indeed, at \(t = 3\), we have
\[ y(t) = 10^{-6} \cdot (0.3059, 0.3059) \]
with an absolute error more or less equal to the initial one.

In conclusion, we could say that although the matrix (2) is universally considered “stable”, in the sense that it dumps a perturbation in the initial value, at the same time it can be considered “unstable” because it can magnify the same perturbation when we look at the relative error rather than the absolute error.

The previous example shows that the time propagation of the relative error in the initial value of the ODE (1) is an interesting phenomenon, which merits to be investigated. To this aim, now we introduce some more or less known notions about the conditioning of linear problems (see e.g. [2]).

Consider the problem of computing the result \(v = Bu\) for the data \(u \in \mathbb{R}^n \setminus \{0\}\), where \(B \in \mathbb{R}^{n \times n}\) is given and non-singular. Let \(\| \cdot \|\) be a fixed vector norm on \(\mathbb{R}^n\) and with the same symbol we denote the matrix norm on \(\mathbb{R}^{n \times n}\) induced by this vector norm. If the data \(u\) is perturbed to \(\tilde{u}\) with relative error
\[ \varepsilon = \frac{\| \tilde{u} - u \|}{\| u \|}, \]
then the result \(v = Bu\) is perturbed to \(\tilde{v} = B \tilde{u}\) with relative error
\[ \delta = \frac{\| \tilde{v} - v \|}{\| v \|}. \]
By writing
\[ \tilde{u} = u + \varepsilon \| u \| \hat{z}, \]
where
\[ \hat{z} = \frac{\tilde{u} - u}{\| u - u \|} \]
is the direction of the perturbation (\(\hat{z}\) is an arbitrary unit vector when \(\tilde{u} = u\), i.e. \(\varepsilon = 0\)), we have
\[ \delta = \kappa(B, u, \hat{z}) \cdot \varepsilon, \]
where
\[ \kappa(B, u, \hat{z}) := \frac{\| B \hat{z} \|}{\| B u \|} \quad \text{with} \quad \hat{u} = \frac{u}{\| u \|} \]
is called the condition number of \(B\) at the data \(u\) with direction of the perturbation \(\hat{z}\).

In general, we do not know the specific direction of the perturbation \(\hat{z}\), but we have information only on \(\varepsilon\). So, we can only write
\[ \delta \leq \kappa(B, u) \cdot \varepsilon, \quad (4) \]
where
\[ \kappa(B, u) := \max_{\hat{z} \in \mathbb{R}^n, \| \hat{z} \| = 1} \kappa(B, u, \hat{z}) = \frac{\| B \|}{\| B \hat{u} \|} \]
is called the condition number of $B$ at the data $u$. If the direction of the perturbation $\hat{z}$ is such that

$$\|B\| = \max_{v \in \mathbb{R}^n \setminus \{0\}} \|Bv\| = \|B\hat{z}\|,$$

we have $\kappa(B,u) = \kappa(B,u,\hat{z})$ and equality in (4) holds.

If we do not know the specific data $u$, we can only write

$$\delta \leq \kappa(B) \varepsilon,$$

where

$$\kappa(B) := \max_{u \in \mathbb{R}^n \setminus \{0\}} \kappa(B,u) = \|B\| \|B^{-1}\|,$$

is called the condition number of $B$. If the data $u$ is $B^{-1}w$, where $w \in \mathbb{R}^n \setminus \{0\}$ is such that

$$\|B^{-1}\| = \frac{\|B^{-1}w\|}{\|w\|},$$

we have $\kappa(B) = \kappa(B,u)$ and if, in addition, the direction of the perturbation $\hat{z}$ satisfies (5), we have $\kappa(B) = \kappa(B,u) = \kappa(B,u,\hat{z})$ and equality in (6) holds.

We define the following condition numbers of the ODE (1) with respect to the initial value by looking at the result $y(t) = e^{tA}y_0$ relevant to the data $y_0$:

$$K(t,A,y_0,\hat{z}_0) := \kappa(e^{tA},y_0,\hat{z}_0) = \frac{\|e^{tA}\hat{z}_0\|}{\|e^{tA}y_0\|},$$

$$K(t,A,y_0) := \kappa(e^{tA},y_0) = \max_{\hat{z}_0 \in \mathbb{R}^n \setminus \{0\}} K(t,A,y_0,\hat{z}_0) = \frac{\|e^{tA}\|}{\|e^{tA}y_0\|},$$

$$K(t,A) := \kappa(e^{tA}) = \max_{y_0 \in \mathbb{R}^n \setminus \{0\}} K(t,A,y_0) = \|e^{tA}\| \cdot \|e^{-tA}\|,$$

where

$$\hat{z}_0 = \frac{\tilde{y}_0 - y_0}{\|\tilde{y}_0 - y_0\|} \quad \text{and} \quad \hat{y}_0 = \frac{y_0}{\|y_0\|}.$$

Apart from the well-known condition number (9) of the matrix exponential, which turns out to be not too much important since it describes the propagation of the relative error in a nongeneric situation for the initial value, at the best of our knowledge conditions numbers for ODEs like (7) and (8) have not been previously considered in literature.

2.1. 2-norm and normal matrices. In the following, we use the 2-norm and we assume that $A$ is a normal matrix. When the 2-norm is used, the previous condition numbers $K(\cdot)$ are written $K_2(\cdot)$.

Under the assumption that $A$ is diagonalizable, $A$ has the spectral decomposition

$$A = \sum_{i=1}^{p} \lambda_i P_i$$

and, for an analytic complex function $g$ and $u \in \mathbb{R}^n$, we have

$$g(A)u = \sum_{i=1}^{p} g(\lambda_i) P_i u,$$
where $\lambda_1, \ldots, \lambda_p$ are the distinct complex eigenvalues of $A$ and, for $i = 1, \ldots, p$, $P_i$ is the projection on the eigenspace of the eigenvalue $\lambda_i$. Under the stronger assumption that $A$ is normal, the eigenspaces are orthogonal and so we have

$$\|g(A)u\|^2 = \sum_{i=1}^{p} |g(\lambda_i)|^2 \cdot \|P_i u\|^2.$$  \hfill (11)

The big advantage of considering normal matrices is that then we have (11) at our disposal. On the other hand, by confining our attention to normal matrices is not too much restrictive since the class of normal matrices includes important families of matrices as symmetric matrices, skew-symmetric matrices and orthogonal matrices. Also observe that the choice of considering only normal matrices is adequate for a first paper on the subject as the present one.

The following theorem gives expressions for the three condition numbers $K_2(t,A,y_0,\tilde{z}_0)$, $K_2(t,A,y_0)$ and $K_2(t,A)$ in case of a normal matrix $A$.

**Theorem 2.2.** Let $A \in \mathbb{R}^{n \times n}$ be a normal matrix. Assume that the spectrum $\{\lambda_1, \ldots, \lambda_p\}$ of $A$ is partitioned in the subsets

$$\{\lambda_{i_0, j+1}, \lambda_{i_0, j+2}, \ldots, \lambda_{i_0, j}\}, \quad j = 1, \ldots, q,$$

where $i_0 = 0 < i_1 < i_2 < \cdots < i_q = p$, by decreasing real parts, i.e. we have

$$\text{Re} (\lambda_{i_0, j+1}) = \text{Re} (\lambda_{i_0, j+2}) = \cdots = \text{Re} (\lambda_{i_0, j}) =: r_j, \quad j = 1, \ldots, q,$$

with

$$r_1 > r_2 > \cdots > r_q$$

(see Figure 4). For $i = 1, \ldots, p$, let $P_i$ be the orthogonal projection on the eigenspace of the eigenvalue $\lambda_i$. Moreover, for $j = 1, \ldots, q$, let

$$Q_j = \sum_{i=j_{j-1}+1}^{i_j} P_i$$

be the orthogonal projection on the sum of the eigenspaces of $\lambda_{i_{j-1}+1}, \lambda_{i_{j-1}+2}, \ldots, \lambda_{i_j}$. Finally, let $y_0 \in \mathbb{R}^n \setminus \{0\}$ be an initial value of (1) and let $\tilde{y}_0 \in \mathbb{R}^n$ be a perturbed initial value.
For any $t \geq 0$, we have

$$K_2(t, A, y_0, \hat{z}_0) = \frac{\|Q_1 \hat{z}_0\|_2^2 + \sum_{j=2}^q e^{2(r_j - r_1)t} \|Q_j \hat{z}_0\|_2^2}{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^q e^{2(r_j - r_1)t} \|Q_j \hat{y}_0\|_2^2}$$  \hspace{1cm} (12)$$

$$K_2(t, A, y_0) = \frac{1}{\sqrt{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^q e^{2(r_j - r_1)t} \|Q_j \hat{y}_0\|_2^2}}$$

$$K_2(t, A) = e^{(r_1 - r_q)t},$$

where

$$\hat{z}_0 = \frac{\hat{y}_0 - y_0}{\|\hat{y}_0 - y_0\|_2} \text{ and } \hat{y}_0 = \frac{y_0}{\|y_0\|_2}.$$

Moreover:

- if $\hat{z}_0$ lies in the sum of the eigenspaces of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_i$, with maximum real part, then, for any $t \geq 0$,

$$K_2(t, A, y_0) = K_2(t, A, y_0, \hat{z}_0);$$

- if $y_0$ lies in the sum of the eigenspaces of the eigenvalues $\lambda_{i-1}, \lambda_{i-1}, \ldots, \lambda_i = \lambda_{i_q}$ with minimum real part, then, for any $t \geq 0$,

$$K_2(t, A) = K_2(t, A, y_0)$$

and if, in addition, $\hat{z}_0$ lies in the sum of the eigenspaces of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_i$, then, for any $t \geq 0$,

$$K_2(t, A) = K_2(t, A, y_0) = K_2(t, A, y_0, \hat{z}_0).$$

Proof. By (11) we have, for $u \in \mathbb{R}^n$,

$$\|e^{tA}u\|_2^2 = \sum_{i=1}^p |e^{\lambda_i t}|^2 \cdot \|P_i u\|_2^2 = \sum_{i=1}^p e^{2\text{Re}(\lambda_i)t} \cdot \|P_i u\|_2^2$$

$$= \sum_{j=1}^q e^{2r_j t} \sum_{i=i_{j-1} + 1}^{i_j} \|P_i u\|_2^2 = \sum_{j=1}^q e^{2r_j t} \|Q_j u\|_2^2.$$

Thus, for $t \geq 0$, we have

$$K_2(t, A, y_0, \hat{z}_0) = \frac{\|e^{tA} \hat{z}_0\|_2^2}{\|e^{tA} \hat{y}_0\|_2^2} = \frac{\sum_{j=1}^q e^{2r_j t} \|Q_j \hat{z}_0\|_2^2}{\sum_{j=1}^q e^{2r_j t} \|Q_j \hat{y}_0\|_2^2}$$

$$= \frac{\sum_{j=1}^q e^{2(r_j - r_1)t} \|Q_j \hat{z}_0\|_2^2}{\sum_{j=1}^q e^{2(r_j - r_1)t} \|Q_j \hat{y}_0\|_2^2}.$$
\[
\|Q_1 \hat{z}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{z}_0\|_2^2 = \frac{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2}{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2}.
\]

Now, we give the expression for \(K_2 (t, A, y_0)\). Observe that

\[
\|Q_1 \hat{z}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{z}_0\|_2^2 \leq \sum_{j=1}^{q} \|Q_j \hat{z}_0\|_2^2 = 1.
\]

Moreover, if \(\|Q_j \hat{z}_0\| = 0, j = 2, \ldots, q\), i.e. \(\hat{z}_0\) lies in the sum of the eigenspaces of the eigenvalues \(\lambda_1, \lambda_2, \ldots, \lambda_i\), then

\[
\|Q_1 \hat{z}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{z}_0\|_2^2 = \|Q_1 \hat{z}_0\|_2^2 = 1.
\]

Thus

\[
K_2 (t, A, y_0)^2 = K_2 (t, A, y_0, \hat{z}_0)^2 = \frac{1}{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2}
\]

for such \(\hat{z}_0\).

Finally, we give the expression for \(K_2 (t, A)\). Observe that

\[
\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2 \geq e^{2(r_q-r_1)t} \sum_{j=1}^{q} \|Q_j \hat{y}_0\|_2^2 = e^{2(r_q-r_1)t}.
\]

Moreover, if \(\|Q_j \hat{y}_0\| = 0, j = 1, \ldots, q-1\), i.e. \(y_0\) lies in the sum of the eigenspaces of the eigenvalues \(\lambda_{i_1}, \lambda_{i_1+1}, \lambda_{i_1+2}, \ldots, \lambda_{i_p} = \lambda_p\), then

\[
\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2 = e^{2(r_q-r_1)t} \sum_{j=2}^{q} \|Q_j \hat{y}_0\|_2^2 = e^{2(r_q-r_1)t}.
\]

Thus

\[
K_2 (t, A)^2 = K_2 (t, A, y_0)^2 = e^{2(r_1-r_q)t}
\]

for such \(y_0\).

For a normal matrix \(A\), the previous theorem says that if the initial value \(y_0 \in \mathbb{R}^n \setminus \{0\}\) of (1) is perturbed to \(\hat{y}_0\) with direction of the perturbation \(\hat{z}_0\) and relative error \(\varepsilon\), then, regarding the relative error \(\delta\) of the perturbed solution, we have the following three facts.

- For any \(t \geq 0\), we have
  \[
  \delta (t) = K_2 (t, A, y_0, \hat{z}_0) \cdot \varepsilon = \frac{\sqrt{\|Q_1 \hat{z}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{z}_0\|_2^2}}{\sqrt{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2}} \cdot \varepsilon.
  \]

- For any \(t \geq 0\), independently of \(\hat{z}_0\), we have
  \[
  \delta (t) \leq K_2 (t, A, y_0) \cdot \varepsilon = \frac{1}{\sqrt{\|Q_1 \hat{y}_0\|_2^2 + \sum_{j=2}^{q} e^{2(r_j-r_1)t} \|Q_j \hat{y}_0\|_2^2}} \cdot \varepsilon.
  \]
with \( \leq \) replaced by \( = \) for any \( t \geq 0 \) if \( \hat{z}_0 \) lies in the sum of the eigenspaces of the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_i \).

- For any \( t \geq 0 \), independently of \( y_0 \) and \( \hat{z}_0 \), we have
  \[
  \delta (t) \leq K_2 (t, A) \cdot \varepsilon = e^{(r_1 - r_\gamma) t} \cdot \varepsilon,
  \]
  with \( \leq \) replaced by \( = \) for any \( t \geq 0 \) if \( y_0 \) lies in the sum of the eigenspaces of the eigenvalues \( \lambda_{i-1+1}, \lambda_{i-1+2}, \ldots, \lambda_{i_q} = \lambda_p \) and \( \hat{z}_0 \) lies in the sum of the eigenspaces of the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_{i_q} \).

Regarding the condition number \( K_2 (t, A, y_0) \) given in (13), we can say that it is an increasing function of \( t \) with \( K_2 (0, A, y_0) = 1 \) whose asymptotic behavior is given by

\[
K_2 (t, A, y_0) \sim e^{(r_1 - r_\gamma) t} R_j \| Q_j y_0 \|_2, \quad t \to +\infty,
\]
where \( j^* \) is the minimum index \( j = 1, \ldots, q \) such that \( Q_j y_0 \neq 0 \) and \( \sim \) means limit 1 as \( t \to +\infty \).

Regarding the condition number \( K_2 (t, A, y_0, \hat{z}_0) \) given in (12), we can write

\[
K_2 (t, A, y_0, \hat{z}_0) = K_2 (t, A, y_0) \cdot R_2 (t, A, \hat{z}_0), \quad t \geq 0,
\]
where

\[
R_2 (t, A, \hat{z}_0) = \sqrt{\| Q_1 \hat{z}_0 \|_2^2 + \sum_{j=2}^q e^{2(r_j - r_\gamma) t} \| Q_j \hat{z}_0 \|_2^2}.
\]

\( R_2 (t, A, \hat{z}_0) \) is a decreasing function of \( t \) with \( R_2 (0, A, \hat{z}_0) = 1 \) whose asymptotic behavior is given by

\[
R_2 (t, A, \hat{z}_0) \sim e^{(r_j^* - r_\gamma) t} \| Q_j \hat{z}_0 \|_2, \quad t \to +\infty,
\]
where \( j^{**} \) is the minimum index \( j = 1, \ldots, q \) such that \( Q_j \hat{z}_0 \neq 0 \). So, the asymptotic behavior of \( K_2 (t, A, y_0, \hat{z}_0) \) is given by

\[
K_2 (t, A, y_0, \hat{z}_0) \sim e^{(r_j^* - r_\gamma) t} \frac{\| Q_j \hat{z}_0 \|_2}{\| Q_j y_0 \|_2}, \quad t \to +\infty.
\]

Moreover, we have

\[
K_2 (t, A, y_0, \hat{z}_0) = 1, \quad t \geq 0,
\]
when \( q = 1 \), since

\[
K_2 (t, A, y_0) = 1 \quad \text{and} \quad R_2 (t, A, \hat{z}_0) = 1, \quad t \geq 0,
\]
when \( q = 1 \).

Now, some conclusions can be drawn.

If

\[
\text{Re} (\lambda_1) = \text{Re} (\lambda_2) = \cdots = \text{Re} (\lambda_p),
\]
i.e. \( q = 1 \), then

\[
\delta (t) = \varepsilon, \quad t \geq 0,
\]
for all \( y_0 \neq 0 \) and \( \bar{y}_0 \) as in the scalar case. Observe that skew-symmetric matrices fulfill (16). It is not a surprise that (17) holds for skew-symmetric matrices: if \( A \) is skew-symmetric, then, for \( t \geq 0 \), \( e^{tA} \) is orthogonal and so

\[
\delta (t) = \frac{\| e^{tA} (\bar{y}_0 - y_0) \|_2}{\| e^{tA} \bar{y}_0 \|_2} = \frac{\| \bar{y}_0 - y_0 \|_2}{\| y_0 \|_2} = \varepsilon.
\]
If (16) does not hold, then we have an unbounded magnification of the relative error \( \delta \) with respect to \( \varepsilon \) only in the non-generic situation \( Q_1 \bar{y}_0 = 0 \), i.e. \( j^* > 1 \), and when \( j^{**} < j^* \). In the generic situation \( Q_1 \bar{y}_0 \neq 0 \), i.e. \( j^* = 1 \), the magnification can be up to \( \|Q_1 \bar{y}_0\|_2 \).

Now, we can understand what happens in the example at the beginning of this section.

**Example 2.3.** (Example 2.1 revisited). The matrix (2) has eigenvalues \( \lambda_1 = -0.01 \) and \( \lambda_2 = -5 \) with relevant eigenvectors \( (1,1) \) and \( (1,-1) \), respectively. The projection \( Q_1 \) and \( Q_2 \) are given by

\[
Q_1 u = \frac{1}{2} (u_1 + u_2) \cdot (1,1) \quad \text{and} \quad Q_2 u = \frac{1}{2} (u_1 - u_2) \cdot (1,-1), \; u \in \mathbb{R}^2.
\]

In Figure 1, we have

\[
y_0 = (0.9, -0.7) \quad \text{and} \quad \bar{y}_0 = y_0 + (0.01,-0.01).
\]

Thus \( j^* = 1 \) with \( \|Q_1 \bar{y}_0\|_2 = \frac{0.2}{\sqrt{2.60}} \) and \( j^{**} = 2 \) with \( \|Q_2 \hat{z}_0\|_2 = 1 \). We conclude that

\[
\delta(t) \sim e^{(r_{j^*} - r_{j^{**}}) t} \frac{\|Q_{j^{**}} \hat{z}_0\|_2}{\|Q_{j^*} \bar{y}_0\|_2} \cdot \varepsilon = 5\sqrt{2.60} e^{-4.99 t} \cdot \varepsilon, \; t \to +\infty,
\]

and this explains the fast decrease of the relative error in the right side of Figure 1.

In Figure 2, we have the same \( y_0 \) of Figure 1 and

\[
\bar{y}_0 = y_0 + (0.01,0.01).
\]

Thus \( j^* \) and \( \|Q_1 \bar{y}_0\|_2 \) are the same as above and now \( j^{**} = 1 \) with \( \|Q_1 \hat{z}_0\|_2 = 1 \). We conclude that

\[
\delta(t) \sim 5\sqrt{2.60} \cdot \varepsilon, \; t \to +\infty,
\]

and this explains the growth of one order of magnitude for the relative error in the right side of Figure 2. Observe that in this case we have

\[
K_2(t,A,y_0,\hat{z}_0) = K_2(t,A,y_0), \; t \geq 0,
\]

since \( \hat{z}_0 \) lies in the eigenspace of \( \lambda_1 \).

Finally, in Figure 3 we have

\[
y_0 = (1, -1) \quad \text{and} \quad \bar{y}_0 = y_0 + (0.01,0.01).
\]

Thus \( j^* = 2 \) with \( \|Q_1 \bar{y}_0\|_2 = 1 \) and \( j^{**} = 1 \) with \( \|Q_1 \hat{z}_0\|_2 = 1 \). We conclude that

\[
\delta(t) \sim e^{4.99 t} \cdot \varepsilon, \; t \to 0,
\]

and this explains the explosion of the relative error in the right side of Figure 3. Observe that in this case we have

\[
K_2(t,A,y_0,\hat{z}_0) = K_2(t,A,y_0) = K_2(t,A) = e^{4.99 t}, \; t \geq 0,
\]

since \( y_0 \) lies in the eigenspace of \( \lambda_2 \) and \( \hat{z}_0 \) lies in the eigenspace of \( \lambda_1 \) and then

\[
\delta(t) = e^{4.99 t} \cdot \varepsilon, \; t \geq 0.
\]

What happens in this last case is that the unperturbed solution includes only the fast decreasing term \( e^{\lambda_2 t} = e^{-5 t} \), but any small perturbation in the initial value moving it outside the eigenspace of \( \lambda_2 \) gives rise to the slow decreasing term \( e^{\lambda_1 t} = e^{-0.01 t} \) in the perturbed solution that soon dominates completely the fast decreasing term.
By concluding this section, we remark that, for a normal matrix \( A \in \mathbb{R}^{n \times n} \),

“all the eigenvalues of \( A \) have real part \( \leq 0 \)” \((18)\)
is a necessary and sufficient condition for not amplifying the absolute error of perturbations in the initial value, i.e.

\[
\|y(t) - y(t)\|_2 \leq \|\tilde{y}_0 - y_0\|_2 \quad \text{for all } y_0, \tilde{y}_0 \in \mathbb{R}^n \text{ and } t \geq 0.
\]

On the other hand, above we have seen that

“all the eigenvalues of \( A \) have the same real part” \((19)\)
is a necessary and sufficient condition for not amplifying the relative error of perturbations in the initial value, i.e.

\[
\frac{\|y(t) - y(t)\|_2}{\|y(t)\|_2} \leq \frac{\|\tilde{y}_0 - y_0\|_2}{\|y_0\|_2} \quad \text{for all } y_0 \in \mathbb{R}^n \setminus \{0\}, \tilde{y}_0 \in \mathbb{R}^n \text{ and } t \geq 0.
\]

Indeed, under \((19)\), \( \leq \) is replaced with \( = \).

Observe that the condition \((18)\) for the absolute error is a condition on the absolute position of the eigenvalues in the complex plane, whereas the condition \((19)\) for the relative error is a condition on their relative position.

Since the real parts of the eigenvalues of the normal matrix \( A \) are the eigenvalues of the symmetric part \( \frac{1}{2} (A^T + A) \) of \( A \) (see [3]), we obtain that \((19)\) holds if and only if

\[
\frac{1}{2} (A^T + A) = \alpha I
\]

for some \( \alpha \in \mathbb{R} \), i.e.

\[
a_{ii} = \alpha, \ i = 1, \ldots, n, \\
a_{ij} = -a_{ij}, \ i,j = 1, \ldots, n \text{ with } i \neq j.
\]

So, a normal matrix does not amplify the relative error of perturbations in the initial value if and only if it is the sum of a skew-symmetrical matrix and a multiple of the identity.

We also remark that all our previous results are not confined to normal real matrices but they are also valid for normal complex matrices.

3. Condition numbers for equations with nonzero equilibrium. Many mathematical models based on linear autonomous ordinary differential equations have, unlike \((1)\), a nonzero equilibrium. Therefore, now we consider the ODE

\[
\begin{align*}
\begin{cases}
y'(t) = Ay(t) + b, \ t \geq 0, \\
y(0) = y_0,
\end{cases}
\end{align*}
\]

\((20)\)

where \( A \in \mathbb{R}^{n \times n} \) is non-singular and \( b \in \mathbb{R}^n \setminus \{0\} \). The solution is given by

\[
y(t) = x + e^{tA} (y_0 - x), \ t \geq 0,
\]

where \( x = -A^{-1}b \neq 0 \). For \( y_0 = x \), the ODE \((20)\) has the equilibrium solution

\[
y(t) = x, \ t \geq 0,
\]

which is the unique equilibrium.

A relative error analysis similar to \((1)\) can be done. We consider an initial value \( y_0 \neq 0 \) and a perturbed initial value \( \tilde{y}_0 \). Let

\[
\varepsilon = \frac{\|\tilde{y}_0 - y_0\|}{\|y_0\|}.
\]
Let $\tilde{y}$ be the solution relevant to the perturbed initial value $y_0$ and let

$$\delta(t) = \frac{\|y(t) - y(t)\|}{\|y(t)\|}, \quad t \geq 0.$$ 

We study the relation between the relative errors $\delta(t)$ and $\varepsilon$.

Unlike (1), it is not true that $y(t) \neq 0$, $t \geq 0$, for any $y_0 \neq 0$. We consider the relative error $\delta(t)$ not defined for $t \geq 0$ such that $y(t) = 0$.

Now, we introduce the two condition numbers $J(t, A, b, y_0, \tilde{z}_0)$ and $J(t, A, b, y_0)$, whose role is similar to $K(t, A, y_0, \tilde{z}_0)$ and $K(t, A, y_0)$, respectively. We have

$$\delta(t) = J(t, A, b, y_0, \tilde{z}_0) : \varepsilon, \quad t \geq 0,$$

where

$$J(t, A, b, y_0, \tilde{z}_0) := \|e^{tA}\| \cdot \frac{\|y_0\|}{\|y(t)\|},$$

with

$$\tilde{z}_0 = \frac{\tilde{y}_0 - y_0}{\|\tilde{y}_0 - y_0\|},$$

being the direction of perturbation ($\tilde{z}_0$ is an arbitrary unit vector when $\tilde{y}_0 = y_0$, i.e. $\varepsilon = 0$). Moreover, we set

$$J(t, A, b, y_0) := \max_{\tilde{z}_0 \in \mathbb{R}^n : \|\tilde{z}_0\| = 1} J(t, A, b, y_0, \tilde{z}_0) = \|e^{tA}\| \cdot \frac{\|y_0\|}{\|y(t)\|}, \quad t \geq 0.$$

Of course, $J(t, A, b, y_0, \tilde{z}_0)$ and $J(t, A, b, y_0)$ are defined for $t \geq 0$ such that $y(t) \neq 0$ and, for $s \geq 0$ such that $y(s) = 0$, we have

$$\lim_{t \to s} J(t, A, b, y_0, \tilde{z}_0) = +\infty \quad \text{and} \quad \lim_{t \to s} J(t, A, b, y_0) = +\infty.$$

**Remark 1.** The absolute error analysis for (20) is identical to the absolute error analysis for (1): in both cases we have

$$y(t) - y(t) = e^{tA}(\tilde{y}_0 - y_0).$$

On the other hand, the relative error analysis for (20) is different from the relative error analysis for (1): in both cases we have

$$\delta(t) = \|e^{tA}\| \cdot \frac{\|y_0\|}{\|y(t)\|} \cdot \varepsilon,$$

where $\tilde{z}_0$ is the direction of the perturbation, but $y(t) = e^{tA}y_0$ for (1) and $y(t) = x + e^{tA}(y_0 - x)$ for (20).

Let us introduce the unit vector

$$\hat{x} = \frac{x}{\|x\|},$$

and then we write an initial value as

$$y_0 = \|x\| \left( \hat{x} + c\hat{d}_0 \right), \quad (21)$$

where the unit vector $\hat{d}_0$ and $c \geq 0$ are given by

$$\hat{d}_0 = \frac{y_0 - x}{\|y_0 - x\|} \quad \text{and} \quad c = \frac{\|y_0 - x\|}{\|x\|}$$

($\hat{d}_0$ is an arbitrary unit vector when $y_0 = x$, i.e. $c = 0$). Since $y_0 \neq 0$, the situation $\hat{d}_0 = -\hat{x}$ and $c = 1$ is excluded.
We have
\[ y(t) = \|x\| \left( \widehat{x} + ce^{tA}\widehat{d}_0 \right), \quad t \geq 0, \]  
and so
\[ J(t, A, b, y_0, \widehat{z}_0) = \|e^{tA}\widehat{z}_0\| \cdot \frac{\|\widehat{x} + cd_0\|}{\|\widehat{x} + ce^{tA}d_0\|}, \quad t \geq 0, \]  
and
\[ J(t, A, b, y_0) = \|e^{tA}\| \cdot \frac{\|\widehat{x} + cd_0\|}{\|\widehat{x} + ce^{tA}d_0\|}, \quad t \geq 0. \]  

3.1. 2-norm and normal matrices. Now, as in the relative error analysis for (1), we use the 2-norm (by writing the condition numbers \( J(\cdot) \) as \( J_2(\cdot) \)) and we assume that \( A \) is a normal matrix.

Next theorem gives an expression for the two condition numbers \( J_2(t, A, b, y_0, \widehat{z}_0) \) and \( J_2(t, A, b, y_0) \). In the theorem and in the following, \( \langle \cdot, \cdot \rangle \) denotes the standard scalar product on \( \mathbb{R}^n \).

**Theorem 3.1.** Let \( A \in \mathbb{R}^{n \times n} \) be a non-singular normal matrix. Assume that the spectrum of \( A \) is partitioned by decreasing real parts as in Theorem 2.2. We use the same notations of that theorem. Let \( y_0 \in \mathbb{R}^n \setminus \{0\} \) be an initial value of (20) and let \( \widehat{y}_0 \in \mathbb{R}^n \) be a perturbed initial value. For any \( t \geq 0 \), we have
\[ J_2(t, A, b, y_0, \widehat{z}_0) = J_2(t, A, b, y_0) \cdot R_2(t, A, \widehat{z}_0) \]  
where
\[ \widehat{z}_0 = \frac{\widehat{y}_0 - y_0}{\|\widehat{y}_0 - y_0\|_2}, \quad \widehat{d}_0 = \frac{y_0 - x}{\|y_0 - x\|_2}, \]  
\( R_2(t, A, \widehat{z}_0) \) is defined in (14) and, for \( j = 1, \ldots, q \),
\[ \mu_j(t) = \sum_{i=1 \atop i \neq 1}^{q} e^{\sqrt{-1} \text{Im}(\lambda_i) t} \left( \widehat{x}, P_i \widehat{d}_0 \right) \]  
with \( \sqrt{-1} \) denoting the imaginary unit.

Moreover, if \( \widehat{z}_0 \) lies in the sum of the eigenspaces of the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_i \), with maximum real part, then, for any \( t \geq 0 \),
\[ J_2(t, A, b, y_0) = J_2(t, A, b, y_0, \widehat{z}_0). \]  

**Proof.** As for the term \( \|e^{tA}\widehat{z}_0\|_2 \) in (23), observe that
\[ \|e^{tA}\widehat{z}_0\|_2 = \sqrt{\sum_{j=1}^{q} e^{2r_j t} \|Q_j \widehat{z}_0\|_2^2} = e^{r_1 t} R_2(t, A, \widehat{z}_0), \quad t \geq 0. \]  

(27)
Moreover, if \( Q_j \hat{z}_0 = 0, \ j = 2, \ldots, q, \) i.e. \( \hat{z}_0 \) lies in the sum of the eigenspaces of the eigenvalues of maximum real part \( \lambda_1, \lambda_2, \ldots, \lambda_i, \) then
\[
\| e^{tA} \|_2 = \| e^{tA} \hat{z}_0 \|_2 = e^{r_1 t}, \ t \geq 0,
\]
and so
\[
J_2 (t, A, b, y_0) = J_2 (t, A, b, \hat{z}_0), \ t \geq 0.
\]
By (23), (27), (28) and (24), we obtain (25).

As for the term \( \| \hat{x} + cA_0 \|_2 \) in (24), we have
\[
\| \hat{x} + cA_0 \|_2 = \sqrt{1 + 2c \langle \hat{x}, d_0 \rangle + c^2}.
\]
(29)

As for the other term \( \| \hat{x} + c e^{tA} d_0 \|_2 \) in (24), by (10) and (11) we have
\[
\| \hat{x} + c e^{tA} d_0 \|_2 = \sqrt{1 + 2c \sum_{j=1}^q e^{r_j t} \mu_j(t) + c^2 \sum_{j=1}^q e^{2r_j t} \| Q_j d_0 \|_2^2}
\]
\[
\leq \| Q_j \|_2 \| \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} P_i \|_2 \| P_i \|_2 = \| Q_j d_0 \|_2.
\]
(30)

By (28), (29) and (30), we obtain (26).

Regarding the function \( \mu_j, \ j = 1, \ldots, q, \) in the previous theorem, the following bound holds.

**Theorem 3.2.** For \( j = 1, \ldots, q, \) we have
\[
| \mu_j (t) | \leq \| Q_j \|_2 \cdot \| Q_j d_0 \|_2, \ t \geq 0.
\]
(31)

**Proof.** Since \( \hat{x} - Q_j \hat{x} \) is orthogonal to the sum of the eigenspaces of \( \lambda_{ij-1+1}, \ldots, \lambda_{ij}, \) we have
\[
| \mu_j (t) | = \left| \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} \langle \hat{x}, P_i d_0 \rangle \right|
\]
\[
= \left| \langle \hat{x}, \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} P_i d_0 \rangle \right|
\]
\[
= \left| \langle Q_j \hat{x}, \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} P_i d_0 \rangle \right|
\]
\[
\leq \| Q_j \|_2 \| \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} P_i d_0 \|_2
\]
and, since \( P_i P_k = P_k P_i = 0 \) for \( i \neq k \) holds,
\[
\left| \sum_{i=ij-1+1}^{ij} e^{\sqrt{-1} \text{Im}(\lambda_i) t} P_i d_0 \right|^2 = \sum_{i=ij-1+1}^{ij} \left| e^{\sqrt{-1} \text{Im}(\lambda_i) t} \right|^2 \| P_i d_0 \|_2 = \| Q_j d_0 \|_2.
\]
(32)
3.2. Asymptotic behavior of \( J_2(t, A, b, y_0, \hat{z}_0) \) and \( J_2(t, A, b, y_0) \).

By (25) and (15), we obtain

\[
J_2(t, A, b, y_0, \hat{z}_0) \sim J_2(t, A, b, y_0) \cdot e^{(r_{j^{**}} - r_1)t} \|Q_{j^{**}} \hat{z}_0\|_2, \ t \to +\infty,
\]

where \( j^{**} \) is the minimum index \( j = 1, \ldots, q \) such that \( Q_j \hat{z}_0 \neq 0 \).

Observe that (33) describes the asymptotic behavior of the condition number \( J_2(t, A, b, y_0, \hat{z}_0) \) in terms of the asymptotic behavior of the condition number \( J_2(t, A, b, y_0) \).

In the case \( y_0 = x \), i.e. \( c = 0 \), we have

\[
J_2(t, A, b, y_0) = e^{rt}, \ t \geq 0.
\]

This is in accordance with the fact that the equilibrium \( x \) is unstable, stable and asymptotically stable, when \( r_1 > 0 \), \( r_1 = 0 \) and \( r_1 < 0 \), respectively. Moreover, in this case we have

\[
J_2(t, A, b, y_0, \hat{z}_0) \sim e^{r_{j^{**}}t} \|Q_{j^{**}} \hat{z}_0\|_2, \ t \to +\infty.
\]

In the following, we assume \( y_0 \neq x \), i.e. \( c \neq 0 \). Moreover, let \( j^* \) be the minimum index \( j = 1, \ldots, q \) such that \( Q_j \hat{d}_0 \neq 0 \).

By (26), we obtain the asymptotic behavior of the condition number \( J_2(t, A, b, y_0) \) and then, by (33), the asymptotic behavior of the condition number \( J_2(t, A, b, y_0, \hat{z}_0) \) in the three cases \( r_{j^*} > 0 \), \( r_{j^*} < 0 \) and \( r_{j^*} = 0 \).

- For \( r_{j^*} > 0 \), we have

\[
J_2(t, A, b, y_0) \sim e^{(r_{j^*} - r_1)t} \sqrt{1 + 2c \left< \hat{x}, \hat{d}_0 \right> + c^2} \frac{1}{c \|Q_{j^*} \hat{d}_0\|_2}, \ t \to +\infty.
\]

and

\[
J_2(t, A, b, y_0, \hat{z}_0) \sim e^{(r_{j^{**}} - r_1)t} \sqrt{1 + 2c \left< \hat{x}, \hat{d}_0 \right> + c^2} \frac{1}{c \|Q_{j^{**}} \hat{z}_0\|_2}, \ t \to +\infty.
\]

- For \( r_{j^*} < 0 \), we have

\[
J_2(t, A, b, y_0) \sim e^{r_{j^*}t} \sqrt{1 + 2c \left< \hat{x}, \hat{d}_0 \right> + c^2}, \ t \to +\infty.
\]

and

\[
J_2(t, A, b, y_0, \hat{z}_0) \sim e^{r_{j^{**}}t} \sqrt{1 + 2c \left< \hat{x}, \hat{d}_0 \right> + c^2} \|Q_{j^{**}} \hat{z}_0\|_2, \ t \to +\infty.
\]

- For \( r_{j^*} = 0 \), we have

\[
\lim_{t \to +\infty} e^{-r_{j^*}t} J_2(t, A, b, y_0) = \frac{\sqrt{1 + 2c \left< \hat{x}, \hat{d}_0 \right> + c^2}}{\sqrt{1 + 2cm_{j^*} + c^2 \|Q_{j^*} \hat{d}_0\|_2}}
\]
\[ \liminf_{t \to +\infty} e^{-r_{j^*}t} J_2 (t, A, b, y_0) = \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2}}{\sqrt{1 + 2cM_{j^*} + c^2 \|Q_{j^*}\hat{d}_0\|_2^2}}. \]  
(38)

and

\[ \limsup_{t \to +\infty} e^{-r_{j^*}t} J_2 (t, A, b, y_0, z_0) = \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2}}{\sqrt{1 + 2cM_{j^*} + c^2 \|Q_{j^*}\hat{d}_0\|_2^2}} \|Q_{j^*}z_0\|_2, \]

\[ \liminf_{t \to +\infty} e^{-r_{j^*}t} J_2 (t, A, b, y_0, z_0) = \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2}}{\sqrt{1 + 2cM_{j^*} + c^2 \|Q_{j^*}\hat{d}_0\|_2^2}} \|Q_{j^*}\hat{d}_0\|_2, \]

where

\[ m_{j^*} := \liminf_{t \to +\infty} \mu_{j^*} (t) \text{ and } M_{j^*} := \limsup_{t \to +\infty} \mu_{j^*} (t). \]

Observe that

\[ - \|Q_{j^*}\hat{x}\|_2 \cdot \|Q_{j^*}\hat{d}_0\|_2 \leq m_{j^*} \leq M_{j^*} \leq \|Q_{j^*}\hat{x}\|_2 \cdot \|Q_{j^*}\hat{d}_0\|_2 \]  
(39)

by Theorem 3.2.

Now, we consider some questions concerning the case \( r_{j^*} = 0 \). Then, we analyze the asymptotic behavior in the cases \( r_1 > 0, r_1 < 0 \) and \( r_1 = 0 \).

3.2.1. About the case \( r_{j^*} = 0 \). Let \( r_{j^*} = 0 \). Since \( A \) is nonsingular, the eigenvalues of \( A \) with real part \( r_{j^*} = 0 \) are conjugate pure imaginary pairs of eigenvalues and \( r_{j^*} = i_{j^*} - i_{j^*} = 1 \), the number of such eigenvalues, is even. Let

\[ \lambda_{i_{j^*} - 1 + 1} = -\sqrt{-1} \omega_2, \quad \lambda_{i_{j^*} - 2 + 2} = \sqrt{-1} \omega_2, \]
\[ \lambda_{i_{j^*} - 1 + 3} = -\sqrt{-1} \omega_4, \quad \lambda_{i_{j^*} - 1 + 4} = \sqrt{-1} \omega_4, \]
\[ \vdots \]
\[ \lambda_{i_{j^*} - 1} = -\sqrt{-1} \omega_{i_{j^*} - i_{j^*} - 1}, \quad \lambda_{i_{j^*}} = \sqrt{-1} \omega_{i_{j^*} - i_{j^*} - 1}, \]

where \( \omega_2, \omega_4, \ldots, \omega_{i_{j^*} - i_{j^*} - 1} > 0 \), such conjugate pure imaginary pairs of eigenvalues.

If \( Q_j \hat{x} = 0 \), (39) says that \( m_{j^*} = M_{j^*} = 0 \) and then (38) and (39) becomes

\[ J_2 (t, A, b, y_0) \sim e^{r_{j^*}t} \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2}}{\sqrt{1 + c^2 \|Q_{j^*}\hat{d}_0\|_2^2}}, \quad t \to +\infty, \]

and

\[ J_2 (t, A, b, y_0, z_0) \sim e^{r_{j^*}t} \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2}}{\sqrt{1 + c^2 \|Q_{j^*}\hat{d}_0\|_2^2}} \|Q_{j^*}z_0\|_2, \quad t \to +\infty. \]
Now, we assume $Q_j^* \hat{x} \neq 0$. Theorem 3.2 says that $\mu_{j^*}(t), \ t \geq 0,$ lies between $\pm \|Q_j^* \hat{x}\|_2 \cdot \|Q_j^* \hat{d_0}\|_2$. Next theorem considers the reachability of these two bounds.

**Theorem 3.3.** Assume $Q_j^* \hat{x} \neq 0$. Let $d_{j^*}$ be the dimension of the sum the eigenspaces of $\lambda_{i_{j^*}-1}, \ldots, \lambda_{i_{j^*}}$. If $d_{j^*} = 2$, i.e. there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues, then

$$\mu_{j^*}(t) = \pm \|Q_j^* \hat{x}\|_2 \cdot \|Q_j^* \hat{d_0}\|_2$$

(40) holds for infinitely many $t \geq 0$ which repeat themselves periodically. If $d_{j^*} > 2$, the condition "(40) holds for some $t \geq 0"$ is a nongeneric condition on $\hat{x}$ and $\hat{d_0}$.

**Proof.** Let $t \geq 0$. Observe that

$$\left\| \sum_{i=i_{j^*}-1+1}^{i_{j^*}} e^{\sqrt{-1} \text{Im}(\lambda_{i})t} P_{i} \hat{d_0} \right\|_2 = \left\|Q_j^* \hat{d_0}\right\|_2$$

(41) (see (32)) and then

$$\sum_{i=i_{j^*}-1+1}^{i_{j^*}} e^{\sqrt{-1} \text{Im}(\lambda_{i})t} P_{i} \hat{d_0} \neq 0.$$

Moreover, we have

$$\sum_{i=i_{j^*}-1+1}^{i_{j^*}} e^{\sqrt{-1} \text{Im}(\lambda_{i})t} P_{i} \hat{d_0} = 2 \sum_{k=2}^{i_{j^*}-i_{j^*}-1} k \text{ even} \left( e^{i \omega_k t} P_{i_{j^*}-1+k} \hat{d_0} \right)$$

$$= 2 \sum_{k=2}^{i_{j^*}-i_{j^*}-1} \left( \cos \omega_k t \cdot \text{Re} \left( P_{i_{j^*}-1+k} \hat{d_0} \right) - \sin \omega_k t \cdot \text{Im} \left( P_{i_{j^*}-1+k} \hat{d_0} \right) \right)$$

with

$$\text{Re}(P_{i_{j^*}-1+2} \hat{d_0}), \text{Im}(P_{i_{j^*}-1+2} \hat{d_0}),$$

$$\text{Re}(P_{i_{j^*}-1+4} \hat{d_0}), \text{Im}(P_{i_{j^*}-1+4} \hat{d_0}),$$

$$\vdots$$

$$\text{Re}(P_{i_0} \hat{d_0}), \text{Im}(P_{i_0} \hat{d_0})$$

orthogonal vectors of $\mathbb{R}^n$. Here, $\text{Re}(w)$ and $\text{Im}(w)$ denote the vectors whose components are the real parts and the imaginary parts, respectively, of the components of the vector $w$.

Now, by looking at the Cauchy-Schwarz inequality used in the proof of Theorem 3.2, we see that

$$\mu_{j^*}(t) = \pm \|Q_j^* \hat{x}\|_2 \cdot \|Q_j^* \hat{d_0}\|_2$$
if and only if

there exists $\alpha > 0$ such that

$$Q_j \hat{x} = \pm \alpha \sum_{i = i_j, j - 1}^{i_j, j} e^{\sqrt{-1} \cdot t \cdot \text{Im}(\lambda_i) \cdot t} P_i \hat{d}_0$$

$$= \pm 2\alpha \sum_{k = 2}^{i_j, j - 1} \left( \cos^{\omega t} \cdot \text{Re} \left( P_{i_j, j - 1} + k \hat{d}_0 \right) - \sin^{\omega t} \cdot \text{Im} \left( P_{i_j, j - 1} + k \hat{d}_0 \right) \right).$$

(42)

Let $V_{j^*}$ be the sum the eigenspaces of $\lambda_{i_j, j - 1}, \ldots, \lambda_{i_j, j}$, and let $d_{j^*}$ be its dimension. If $d_{j^*} = 2$, then (42) holds for infinitely many $t \geq 0$ which repeat themselves periodically. In fact, $Q_{j^*} \hat{x} \in V_{j^*}$ and, in case of $d_{j^*} = 2$, we have $i_{j^*} - i_{j^* - 1} = 2$ and $\text{Re}(P_{i_{j^*}} \hat{d}_0)$ and $\text{Im}(P_{i_{j^*}} \hat{d}_0)$ constitute a base for $V_{j^*}$. On the other hand, observe that

$$t \rightarrow \sum_{i = i_{j^*}, j - 1}^{i_{j^*}, j} e^{\sqrt{-1} \cdot t \cdot \text{Im}(\lambda_i) \cdot t} P_i \hat{d}_0 = 2 \sum_{k = 2}^{i_{j^*}, j - 1} \text{Re} \left( e^{\sqrt{-1} \cdot t \cdot P_{i_j, j - 1} + k \hat{d}_0} \right), \quad t \geq 0,$$

(43)

is a curve on the intersection $W$ of $V_{j^*}$ and the sphere of radius $\|Q_{j^*} \hat{d}_0\|_2$ (see (41)). Thus, we have “(42) holds for some $t \geq 0$” if and only if the point

$$\pm \frac{\|Q_{j^*} \hat{d}_0\|_2}{\|Q_{j^*} \hat{x}\|_2} Q_{j^*} \hat{x}$$

(44)

of $W$ is on the curve (43). If $d_{j^*} > 2$, since $W$ is a manifold of dimension $d_{j^*} - 1 > 1$ and (44) can be any point in $W$ by varying the equilibrium $x$, we conclude that the condition “the point (44) is on the curve (43)” is a nongeneric condition on $\hat{x}$, fixed $\hat{d}_0$. □

By the previous theorem, we obtain that if there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues, then

$$m_{j^*} = -\|Q_{j^*} \hat{x}\|_2 \cdot \left\|Q_{j^*} \hat{d}_0\right\|_2 \quad \text{and} \quad M_{j^*} = \|Q_{j^*} \hat{x}\|_2 \cdot \left\|Q_{j^*} \hat{d}_0\right\|_2.$$

The quantity

$$\sqrt{1 + 2cm_{j^*} + c^2 \left\|Q_{j^*} \hat{d}_0\right\|_2^2} = \lim_{t \rightarrow +\infty} \frac{\|y(t)\|_2}{\|x\|_2},$$

(45)

(see (22) and (30)) appears at the denominator in (38) and (39). It is a measure of how much the solution $y$ is asymptotically away from zero. Next theorem says when it is nonzero and when it becomes zero.

**Theorem 3.4.** If $\|Q_{j^*} \hat{x}\|_2 < 1$ and $c\|Q_{j^*} \hat{d}_0\|_2 \neq 1$, then

$$1 + 2cm_{j^*} + c^2 \left\|Q_{j^*} \hat{d}_0\right\|_2^2 > 0.$$

If $\|Q_{j^*} \hat{x}\|_2 = 1$ and $c\|Q_{j^*} \hat{d}_0\|_2 = 1$, then

$$1 + 2cm_{j^*} + c^2 \left\|Q_{j^*} \hat{d}_0\right\|_2^2 \geq 0.$$
and
\[ 1 + 2c m_j^* + c^2 \||Q_j \hat{d}_0||^2 = 0 \]
if and only if
\[ m_j^* = -\||Q_j \hat{d}_0||. \quad (46) \]

Proof. We have
\[ 1 + 2c m_j^* + c^2 \||Q_j \hat{d}_0||^2 = \left( 1 - c \||Q_j \hat{d}_0|| \right)^2 + 2c \left( m_j^* + \||Q_j \hat{d}_0|| \right) \]
with (see (39))
\[ m_j^* + \||Q_j \hat{d}_0|| \geq \||Q_j \hat{d}_0|| \left( 1 - \|Q_j \hat{x}\|_2 \right). \]

Observe that if \( \|Q_j \hat{x}\|_2 = 1 \) and \( c \|Q_j \hat{d}_0\|_2 = 1 \), then the condition (46) is satisfied when there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues. So, when there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues, we have
\[ 1 + 2c m_j^* + c^2 \||Q_j \hat{d}_0||^2 \geq 0 \]
and
\[ 1 + 2c m_j^* + c^2 \||Q_j \hat{d}_0||^2 = 0 \]
if and only if
\[ \|Q_j \hat{x}\|_2 = 1 \] and \( c \|Q_j \hat{d}_0\|_2 = 1. \quad (47) \]

3.2.2. Asymptotic behavior in the case \( r_1 > 0 \). Suppose \( r_1 > 0 \).
In the case \( Q_1 \hat{d}_0 \neq 0 \), i.e. \( j^* = 1 \), by (34) we obtain
\[ \lim_{t \to + \infty} J_2(t, A, b, y_0) = \sqrt{1 + 2c \left\langle \hat{x}, \hat{d}_0 \right\rangle + c^2} \]
with \( c \|Q_1 \hat{d}_0\|_2 \).

Observe that:
1. For fixed \( \hat{d}_0 \), if \( c \to 0 \), i.e. \( y_0 \to x \) along the fixed direction \( \hat{d}_0 \), then
\[ \lim_{t \to + \infty} J_2(t, A, b, y_0) \to + \infty \]
and
\[ \lim_{t \to + \infty} J_2(t, A, b, y_0) \sim \frac{1}{c \|Q_1 \hat{d}_0\|_2}; \]
2. For fixed \( \hat{d}_0 \), if \( c \to + \infty \), i.e. \( y_0 \to \infty \) along the fixed direction \( \hat{d}_0 \), then
\[ \lim_{t \to + \infty} J_2(t, A, b, y_0) \to \frac{1}{\|Q_1 \hat{d}_0\|_2}; \]
3. Assume \( Q_1 \hat{x} \neq 0 \). If \( c \to 1 \) and \( \hat{d}_0 \to -\hat{x} \), i.e. \( y_0 \to 0 \) (see (21) and (29)), then
\[ \lim_{t \to + \infty} J_2(t, A, b, y_0) \to 0. \]
4. Assume $Q_1 \hat{x} = 0$. We have
\[
\liminf_{c \to 1} \lim_{t \to +\infty} J_2 (t, A, b, y_0) \geq 1.
\]
Moreover, for any $v \in [1, +\infty]$, there exist sequences $\{c^{(n)}\}$ and $\{\hat{d}_0^{(n)}\}$ with $c^{(n)} \to 1$ and $\hat{d}_0^{(n)} \to -\hat{x}$, $n \to \infty$, such that
\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) \to v, \; n \to \infty.
\]

The asymptotic behavior of $J_2 (t, A, b, y_0, \hat{z}_0)$ when $j^* = 1$ is described by (35) and we have
\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0, \hat{z}_0) = \begin{cases} 
\sqrt{1+2c(x, d_0)+c^2} \|Q_{j^{**}} \hat{z}_0\|_2 & \text{if } j^{**} = 1 \\
0 & \text{if } j^{**} > 1.
\end{cases}
\]

In the case $Q_{j^{*}} \hat{d}_0 = 0$, i.e. $j^* > 1$, by (34) or (36) or (38) we have
\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) = +\infty.
\]
The asymptotic behavior of $J_2 (t, A, b, y_0, \hat{z}_0)$ when $j^* > 1$ is described by (35) or (37) or (39). For example, in the case $r^*_j < 0$ we have
\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0, \hat{z}_0) = \begin{cases} 
+\infty & \text{if } j^{**} < j^* \\
\sqrt{1+2c(x, d_0)+c^2} \|Q_{j^{**}} \hat{z}_0\|_2 & \text{if } j^{**} = j^* \\
0 & \text{if } j^{**} > j^*.
\end{cases}
\]

3.2.3. Asymptotic behavior in the case $r_1 < 0$. Suppose $r_1 < 0$. Since $r^*_j < 0$, the asymptotic behaviors of $J_2 (t, A, b, y_0)$ and $J_2 (t, A, b, y_0, \hat{z}_0)$ are described by (36) and (37), respectively. In (37), we have $r_{j^{**}} \leq r_1 < 0$.

The relative error in the initial value is asymptotically vanishing and this is in accordance with the fact that the equilibrium $x$ is globally asymptotically stable when $r_1 < 0$: for any initial value, the absolute error is asymptotically vanishing and the solution approaches to $x \neq 0$ as $t \to +\infty$.

3.2.4. Asymptotic behavior in the case $r_1 = 0$. Suppose $r_1 = 0$. Since $A$ is nonsingular, the eigenvalues of $A$ with zero real part are conjugate pure imaginary pairs of eigenvalues and $i_1$, the number of such eigenvalues, is even. Let
\[
\lambda_1 = -\sqrt{-1} \omega_2, \; \lambda_2 = \sqrt{-1} \omega_2, \\
\lambda_3 = -\sqrt{-1} \omega_4, \; \lambda_4 = \sqrt{-1} \omega_4, \\
\vdots \\
\lambda_{i_1-1} = -\sqrt{-1} \omega_{i_1}, \; \lambda_{i_1} = \sqrt{-1} \omega_{i_1},
\]
where $\omega_2, \omega_4, \ldots, \omega_{i_1} > 0$, be such conjugate pure imaginary pairs of eigenvalues.

For an initial value $y_0 \in \mathbb{R}^n$ such that $y_0 - x$ is in the sum of the eigenspaces of $\lambda_{k-1}$ and $\lambda_k$, $k = 2, 4, \ldots, i_1$, we obtain as solution the equilibrium plus a periodic
function of period $\frac{2\pi}{\omega_k}$:

\[
y(t) = x + e^{tA} (y_0 - x) = x + e^{\lambda_k - \xi_t} P_{k-1} (y_0 - x) + e^{\lambda_k t} P_k (y_0 - x)
\]

\[
= x + 2\text{Re} \left( e^{\sqrt{-1} \omega_k t} P_k (y_0 - x) \right), \quad t \geq 0.
\]

For an initial value $y_0 \in \mathbb{R}^n$ such that $y_0 - x$ is in the sum of the eigenspaces of the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{i_1 - 1}, \lambda_{i_1}$, we obtain as solution the equilibrium plus a sum of $\frac{1}{2}$ periodic functions of periods $\frac{2\pi}{\omega_1}, \frac{2\pi}{\omega_4}, \ldots, \frac{2\pi}{\omega_{11}}$:

\[
y(t) = x + e^{tA} (y_0 - x) = x + 2 \sum_{k=2}^{i_1} \text{Re} \left( e^{\sqrt{-1} \omega_k t} P_k (y_0 - x) \right), \quad t \geq 0.
\]

For a generic initial value $y_0 \in \mathbb{R}^n$, we obtain as solution the equilibrium point plus a sum of $\frac{1}{2}$ periodic functions of periods $\frac{2\pi}{\omega_2}, \frac{2\pi}{\omega_4}, \ldots, \frac{2\pi}{\omega_{11}}$ plus a function asymptotically vanishing.

In the case $Q_1 \hat{d}_0 = 0$, i.e. $j^* > 1$, we have $r_{j^*} < 0$ and so by (36) we obtain

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) = \sqrt{1 + 2c \left( \hat{x}, \hat{d}_0 \right) + c^2}.
\]

Observe that:

1. If $c \to 0$, i.e. $y_0 \to x$, then

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) \to 1.
\]

2. If $c \to +\infty$, i.e. $y_0 \to +\infty$, then

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) \to +\infty
\]

and

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) \sim c; \quad (48)
\]

3. If $c \to 1$ and $\hat{d}_0 \to -\hat{x}$, i.e. $y_0 \to 0$, then

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0) \to 0.
\]

The asymptotic behavior of $J_2 (t, A, b, y_0, \hat{z}_0)$ when $j^* > 1$ is described by (37) and we have

\[
\lim_{t \to +\infty} J_2 (t, A, b, y_0, \hat{z}_0) = \begin{cases} 
\sqrt{1 + 2c \left( \hat{x}, \hat{d}_0 \right) + c^2} \cdot \| Q_1 \hat{z}_0 \|_2 \quad & \text{if } j^{**} = 1 \\
0 \quad & \text{if } j^{**} > 1.
\end{cases}
\]
Now, we assume the case $Q\hat{d_0} \neq 0$, i.e. $j^* = 1$. We have $r_{j^*} = 0$ and by (38) we obtain

\[
\begin{align*}
\limsup_{t \to +\infty} J_2(t, A, b, y_0) &= \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d_0} \rangle + c^2}}{\sqrt{1 + 2cm_1 + c^2 \| Q_1 \hat{d_0} \|^2_2}} \\
\liminf_{t \to +\infty} J_2(t, A, b, y_0) &= \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d_0} \rangle + c^2}}{\sqrt{1 + 2cM_1 + c^2 \| Q_1 \hat{d_0} \|^2_2}}.
\end{align*}
\]

(49)

The asymptotic behavior of $J_2(t, A, b, y_0, \hat{z}_0)$ when $j^* = 1$ is described by (39) and we have

\[
\begin{align*}
\limsup_{t \to +\infty} J_2(t, A, b, y_0, \hat{z}_0) &= \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d_0} \rangle + c^2}}{\sqrt{1 + 2cm_1 + c^2 \| Q_1 \hat{d_0} \|^2_2}} \| Q_1 \hat{z}_0 \|_2 \\
\liminf_{t \to +\infty} J_2(t, A, b, y_0, \hat{z}_0) &= \frac{\sqrt{1 + 2c \langle \hat{x}, \hat{d_0} \rangle + c^2}}{\sqrt{1 + 2cM_1 + c^2 \| Q_1 \hat{d_0} \|^2_2}} \| Q_1 \hat{z}_0 \|_2
\end{align*}
\]

if $j^{**} = 1$ and

\[
\lim_{t \to +\infty} J_2(t, A, b, y_0, \hat{z}_0) = 0
\]

if $j^{**} > 1$.

Regarding (49), we can observe the following.

1. If $c \to 0$, i.e. $y_0 \to x$, then

\[
\limsup_{t \to +\infty} J_2(t, A, b, y_0) \to 1 \\
\liminf_{t \to +\infty} J_2(t, A, b, y_0) \to 1.
\]

2. Fixed $\hat{d}_0$, if $c \to +\infty$, i.e. $y_0 \to \infty$ along the fixed direction $\hat{d}_0$, then

\[
\begin{align*}
\limsup_{t \to +\infty} J_2(t, A, b, y_0) &= \frac{1}{\| Q_1 \hat{d_0} \|_2} \\
\liminf_{t \to +\infty} J_2(t, A, b, y_0) &= \frac{1}{\| Q_1 \hat{d_0} \|_2}
\end{align*}
\]

(50)

3. Assume $\| Q_1 \hat{x} \|_2 < 1$. If $c \to 1$ and $\hat{d}_0 \to -\hat{x}$, i.e. $y_0 \to 0$, then

\[
\limsup_{t \to +\infty} J_2(t, A, b, y_0) \to 0.
\]

Now, assume that there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues. Moreover, assume $\| Q_1 \hat{x} \|_2 = 1$. 
4. We have
\[ \limsup_{t \to +\infty} J_2(t, A, b, y_0) = \sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2} \geq 1. \]

5. For any \( v \in [1, +\infty] \), there exist sequences \( \{c^{(n)}\} \) and \( \{\hat{d}_0^{(n)}\} \) with \( c^{(n)} \to 1 \) and \( \hat{d}_0^{(n)} \to -\hat{x}, n \to \infty \), such that
\[ \limsup_{t \to +\infty} J_2(t, A, b, y_0) \to v, n \to \infty. \]

6. If \( c \|Q_1 \hat{d}_0\|_2 \to 1 \), i.e. the solution becomes more and more asymptotically close to zero (see (47)), with \( c \) outside of a neighborhood of 1 or \( \hat{d}_0 \) outside of a neighborhood of \(-\hat{x}\), then
\[ \limsup_{t \to +\infty} J_2(t, A, b, y_0) \to +\infty. \]

Next example illustrates three situations, called A, B and C, where we have a large condition number \( J_2(t, A, b, y_0) \) for large \( t \).

**Example 3.5.** Consider the ODE (20) with the normal matrix
\[ A = \begin{bmatrix} 0 & 1 & -1 \\ -1 & -\sqrt{2} & -\sqrt{2} \\ 1 & -\sqrt{2} & -\sqrt{2} \end{bmatrix}. \]

The matrix \( A \) has the conjugate pure imaginary pair of eigenvalues \( \lambda_1 = -\sqrt{2}i \) and \( \lambda_2 = \sqrt{2}i \) with relevant eigenvectors \((\sqrt{2}, -i, i)\) and \((\sqrt{2}, i, -i)\), respectively, and the other eigenvalue \( \lambda_3 = -2\sqrt{2} \) with relevant eigenvector \((0, 1, 1)\). So, we have \( q = 2 \) with \( r_1 = 0 \) and \( r_2 = -2\sqrt{2} \). The projections \( Q_1 \) and \( Q_2 \) are given by
\[ Q_1 u = \left( u_1, -\frac{1}{2} (u_2 - u_3), \frac{1}{2} (u_2 - u_3) \right) \]
and \( Q_2 u = \frac{1}{2} (0, u_2 + u_3, u_2 + u_3), \ u \in \mathbb{R}^3. \)

Observe that this is the case where there is only one pair of conjugate pure imaginary eigenvalues and they are simple eigenvalues.

As a forcing term for (20), we consider \( b = -Ax \), where
\[ x = (1, 1, -1) \]
is the equilibrium. Note that \( x \) belongs to the sum of the eigenspaces of \( \lambda_1 \) and \( \lambda_2 \) and so \( \|Q_1 x\|_2 = 1 \).

Situation A. Consider the initial value \( y_0 \) determined by
\[ \hat{d}_0 = \frac{1}{\sqrt{3}} (-1, 1, -1) \] and \( c = 0.99. \)

Here \( \hat{d}_0 \) belongs to the sum of the eigenspaces of \( \lambda_1 \) and \( \lambda_2 \) and so the solution is the equilibrium \( x \) plus a periodic function of period \( T = \sqrt{2}\pi \). In the left side of Figure 5, we see the graph of \( \|y(t)\|_2 \) for \( t \in [0, 3T] \). Observe that the solution is periodically quite close to zero. At the times \( t \) where the solution is close to zero,
$J_2(t, A, b, y_0)$ is large (see (22), (24) and (28)). We see this in the right side of Figure 5, where the graph of $J_2(t, A, b, y_0)$ for $t \in [0, 3T]$ is shown. We have

$$\limsup_{t \to +\infty} J_2(t, A, b, y_0) = \sqrt{1 + 2c \langle \hat{x}, \hat{d}_0 \rangle + c^2 \left| 1 - c \| Q_1 \hat{d}_0 \|_2 \right|} = 162.4838$$

and this value is also the maximum value periodically reached by $J_2(t, A, b, y_0)$ (since $\hat{d}_0$ belongs to the sum of the eigenspaces of $\lambda_1$ and $\lambda_2$).

This confirms the point 6 above for the case $Q_1 \hat{d}_0 \neq 0$: we have $c$ large and $\limsup_{t \to +\infty} J_2(t, A, b, y_0)$ is large.

Situation B. Consider the initial value relevant to $\hat{d}_0 = \frac{1}{\sqrt{2}} (0, 1, 1)$ and $c = 100$. Here $\hat{d}_0$ belongs to the eigenspace of $\lambda_3$ and so the solution is the equilibrium $x$ plus an asymptotically vanishing function. The graph of the 2-norm of the solution on $[0, 3T]$ is shown in the left side of Figure 6. Here, we are in the case $Q_1 \hat{d}_0 = 0$ and (see (26))

$$J_2(t, A, b, y_0) = \frac{\sqrt{1 + c^2}}{\sqrt{1 + c^2 e^{2\pi^2 T}}}$$

asymptotically grows up to

$$\lim_{t \to +\infty} J_2(t, A, b, y_0) = \sqrt{1 + c^2} \approx 100$$

as illustrated in the right side of Figure 6, which shows the graph of $J_2(t, A, b, y_0)$ for $t \in [0, 3T]$.

This confirms the point 2 and (48) above for the case $Q_1 \hat{d}_0 = 0$: we have $c$ large and $\lim_{t \to +\infty} J_2(t, A, b, y_0)$ large close to $c$.

Situation C. Consider the initial value relevant to $\hat{d}_0 = \frac{1}{\sqrt{2.0202}} (0.01, 1.01, 1)$ and $c = 1000$. Here $\| Q_1 \hat{d}_0 \|_2 = 0.008617$ and so the solution is the equilibrium $x$ plus a small periodic function of period $T$ plus an asymptotically vanishing function. The graph of the 2-norm of the solution on $[0, 3T]$ is shown in the left side of Figure 7. The graph of $J_2(t, A, b, y_0)$ on $[0, 3T]$ is shown in the right side of Figure 7.

This confirms the point 2 and (50) above for the case $Q_1 \hat{d}_0 \neq 0$: we have $c$ large and

$$\limsup_{t \to +\infty} J_2(t, A, b, y_0) \approx \frac{1}{\| Q_1 \hat{d}_0 \|_2} = 116.05$$

$$\liminf_{t \to +\infty} J_2(t, A, b, y_0) \approx \frac{1}{\| Q_1 \hat{d}_0 \|_2} = 116.05.$$
4. Conclusions. In this paper, we have studied how a relative error in the initial value of the ODE (1) with zero equilibrium or the ODE (20) with a nonzero equilibrium is propagated along the solution. In particular, given a norm $\| \cdot \|$ on $\mathbb{R}^n$, figure 5. 2-norm $\|y(t)\|_2$ and condition number $J_2(t, A, b, y_0)$ for $t \in [0, 3T]$ in case of $\hat{d}_0 = \frac{1}{\sqrt{3}} (-1, 1, -1)$ and $c = 0.99$ (situation A).
Figure 6. 2-norm $\|y(t)\|_2$ and condition number $J_2(t, A, b, y_0)$ for $t \in [0, 3T]$ in case of $d_0 = \frac{1}{\sqrt{2}}(0, 1, 1)$ and $c = 100$ (situation B).

we have considered the relation between the relative error

$$\delta(t) = \frac{\|\bar{y}(t) - y(t)\|}{\|y(t)\|}, \quad t \geq 0,$$

and the relative error

$$\varepsilon = \frac{\|\bar{y}_0 - y_0\|}{\|y_0\|},$$
where \( y \) is the solution relevant to the initial value \( y_0 \neq 0 \) and \( \tilde{y} \) is the solution relevant to the perturbed initial value \( \tilde{y}_0 \).

We have confined our study to the 2-norm and this is not restrictive. In fact, given an arbitrary norm \( \| \cdot \|' \) different from the 2-norm, the norm equivalence says that there exists constants \( c,C > 0 \) depending on the dimension \( n \) of the ODE such that

\[
\| y(t) \|_2 \leq C \| y(t) \|' \quad \text{and} \quad \| \Delta y(t) \|_2 \leq C \| \Delta y(t) \|'
\]
that, for any $u \in \mathbb{R}^n$, we have
\[ c\| u \|_2 \leq \| u \|' \leq C\| u \|_2. \]

Then
\[ \frac{1}{D} K_2(t, A, y_0, \hat{z}_0) \leq K'(t, A, y_0, \hat{z}_0) \leq DK_2(t, A, y_0, \hat{z}_0) \]
and
\[ \frac{1}{D} J_2(t, A, b, y_0, \hat{z}_0) \leq J'(t, A, b, y_0, \hat{z}_0) \leq DJ_2(t, A, b, y_0, \hat{z}_0), \]
where
\[ D = \left( \frac{C}{c} \right)^2, \]
holds for the condition numbers $K_2(t, A, y_0, \hat{z}_0)$ and $J_2(t, A, b, y_0, \hat{z}_0)$ relevant to the 2-norm and the condition numbers $K'(t, A, y_0, \hat{z}_0)$ and $J'(t, A, b, y_0, \hat{z}_0)$ relevant to the norm $\| \cdot \|'$. If $D$ is not a large number and this happens if the dimension $n$ is not large, then the condition numbers $K_2(t, A, y_0, \hat{z}_0)$ and $K'(t, A, y_0, \hat{z}_0)$ have the same order of magnitude, as well as the condition numbers $J_2(t, A, b, y_0, \hat{z}_0)$ and $J'(t, A, b, y_0, \hat{z}_0)$. This means that, for a dimension $n$ not large, it is sufficient to study the propagation of the relative error by using the 2-norm. On the other hand, if $D$ is large and this happens when $n$ is large, the previous condition numbers can have different order of magnitude and then the choice of any particular norm becomes important.

Moreover, we have taken into account only normal matrices $A$ in (1) and (20).

So, as a future work, it is necessary to extend this analysis of the relative error propagation to the case of nonnormal matrices $A$. Moreover, also the extension to time-dependent matrices $A(t)$ and nonlinear equations is of interest.

REFERENCES

[1] A. H. Al-Mohy and N. J. Higham, Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation, SIAM J. Matrix Anal. Appl., 30 (2009), 1639–1657.
[2] F. Burgisser and F. Cucker, Condition, Springer 2013.
[3] R. Grone, C. R. Johnson, E. M. Sa and H. Wolkowicz, Normal matrices, Linear Algebra and its Applications, 87 (1987), 213–225.
[4] G. Golub and C. F. Van Loan, Matrix Computations, The John Hopkins University Press, third edition 1996.
[5] E. Hairer, S. Norsett and G. Wanner, Solving Ordinary Differential Equations I, Nonstiff Problems, Springer-Verlag Berlin Heidelberg, Second Revised Edition, 1993.
[6] N. J. Higham, Functions of Matrices, Theory and Computation, Siam, 2008.
[7] N. J. Higham and A. H. Al-Mohy, Computing matrix functions, Acta Numerica, 19 (2010), 159–208.
[8] B. Kagstrom, Bounds and perturbations for the matrix exponential, BIT, 17 (1977), 39–57.

Received April 2017; revised January 2018.

E-mail address: maset@units.it