

Stability Analysis of Degenerate Gradient Flows via the WKB Approximation

Shahar Ben-Menahem and Abraham K. Ishihara

Abstract—In this note we present a powerful new approach to the analysis of a class of linear, degenerate gradient flow systems that frequently arise in adaptive control and system identification. This paper has three main contributions: 1) a stability theorem utilizing a non-integral variant of the persistence of excitation (PE) conditions on the input signal; 2) upper and lower bounds (also using the non-integral PE conditions) which are shown to be superior to those derived in a classical paper on degenerate flow; and 3) construction of a one-term asymptotic approximation that is shown to perform remarkably well when compared to the numerically integrated solution. At the heart of our results is an extension of the WKB method which we name the Iterative Tracking Diagonalization (ITD) procedure. It yields a condition sufficient to ensure exponential stability of the origin. The WKB method utilizes an asymptotic expansion which relies on the existence of a time scale hierarchy. If the time scale separation parameter is sufficiently small, a few iteration steps suffice to derive an accurate estimate for the time constants of exponential stability of the norm of the parameter error vector. An important feature of our stability theorem, bounds and approximations, is that they all involve an analytical treatment of time dependences.

I. INTRODUCTION

In this note, we examine the stability properties of the degenerate flow

$$\dot{z}(t) = A(t)z(t) \quad (1)$$

subject to an initial condition $z(0) = z_0$ and to a slow-variation condition, on $A(t)$, and other conditions. For every $t \geq 0$, $A(t)$ is a positive semi-definite, rank-one matrix of the specific form $A(t) = -\mu x(t)x^T(t)$. We shall assume that $x(t)$ is infinitely differentiable: $x(t) \in C^\infty(\mathbb{R}^+, \mathbb{R}^n)$. This type of degenerate system arises in adaptive control, system identification, noise cancellation, communications, and other areas. Some classical results can be found in references [11], [1], [8], [3]. A more recent treatment can be found in references [4], [2], [9], [7].

A known sufficient condition for exponential stability of this system is stated as follows [11]. Let $x(t)$ satisfy the following 2 conditions:

Assumption 1: There exist (real) numbers $T > 0$ and $\alpha > 0$ such that $\frac{1}{T} \int_t^{t+T} |d^T x(\tau)|^2 d\tau \geq \alpha \|d\|^2$ for all $t \geq 0$ and all constant vectors $d \in \mathbb{R}^{n \times 1}$

Assumption 2: There exists a positive number L such that $\frac{1}{T} \int_t^{t+T} \|x(\tau)\|^2 d\tau \leq L^2$ holds for all $t \geq 0$, where T is the same as in Assumption 1.

S. Ben-Menahem is with the Department of Physics, Stanford University, Stanford CA 94305, U.S.A. and Avago Technologies Inc., San Jose CA USA sanverim@gmail.com

A. Ishihara is with Carnegie Mellon University Silicon Valley, NASA Research Park, Moffett Field, CA 94035 1000, abe.ishihara@west.cmu.edu

Then, the origin $z = 0$ is globally exponentially stable under the dynamics (1), as proven in reference [11]. The first assumption on the input vector, $x(t)$, is known in the literature as Persistence of Excitation (‘PE’). Loosely speaking this condition states that the input vector is not orthogonal to any one particular direction for too long. The second assumption is a boundedness condition; both play important roles in the derivation of the stability bounds. It is to be understood that for an input vector to be PE, there must exist at least one pair (α, T) such that the persistence of excitation condition holds for all $t \geq 0$.

The WKB method is a systematic asymptotic approximation scheme long used in Quantum Mechanics, optics, and other branches of applied mathematics. The asymptotic WKB expansion is in a parameter quantifying a ratio of temporal or spatial scales. Here, we introduce a closely related matrix version of the WKB approach termed ‘Iterative Tracking Diagonalization’ (ITD), and apply it to the dynamic system (1). Specializing to the dimension $n = 2$ and a particular PE input vector signal, and upon replacing the PE conditions by criteria not involving integrals (and far easier to verify), we are able to prove exponential stability. We also greatly improve upon the bounds (both upper and lower) derived in reference [11]. The non integral version of the PE conditions, which we call the *WKB-PE conditions*, incorporates a time-scale separation parameter ($\epsilon \geq 0$). We demonstrate that for low enough values of ϵ , a small number of ITD iteration steps – when combined with a low-order Matrix WKB approximant using our formalism from reference [6] – suffice to yield an excellent closed form analytical approximation to the general solution of (1).

The remainder of this note is organized as follows. In Section II, the WKB approximation and its history are briefly discussed, the time-scale separation (ϵ), and various elements of notation are introduced. In Section III, our ITD procedure along with the relevant linear algebra machinery is introduced; we then specialize to the case $n = 2$ for which two ITD iterations are explicitly worked out, and our main stability theorem is stated. This theorem is then applied to prove exponential stability for small enough ϵ values. We also summarize the upper and lower analytical bounds for the parameter error vector, $z(t)$, that may be derived from the main theorem. In Section IV, we further specialize to a particular $n = 2$ input signal which is PE. For this example, we combine numerical simulations with analytical derivations, (including those of earlier sections) to derive: (1) an excellent WKB approximation to $z(t)$; (2) our own (WKB-PE-conditions based) upper and lower bounds

on the *components* of $z(t)$ and on its norm $\|z(t)\|$; (3) optimized upper and lower bounds (on $\|z(t)\|$ alone) based upon the formulas of reference [11]. Section V summarizes our conclusions.

II. BACKGROUND AND NOTATION

The WKB method¹ (named after the three physicists G. Wentzel, H. A. Kramers and L. Brillouin) is a systematic asymptotic approximation method that was first formally developed in the context of quantum mechanics [10], [5]. Earlier versions of WKB can be traced to Liouville, Airy, Rayleigh, and other nineteenth century mathematical physicists. In optics and continuum mechanics it is known as the eikonal approximation, and is useful in bridging the gap between the geometrical (ray) and wave descriptions. In Quantum Mechanics the WKB approximation (also known as the *adiabatic* or *semiclassical approximation*) plays a similar role, with the ray approximation to the Schroedinger equation being Newton's equations of classical mechanics. In WKB, one expands all quantities in a formal power series in a parameter ($\epsilon \geq 0$) that quantifies a ratio of two scales of time and/or space. When $\epsilon \ll 1$, the first few terms in the expansion often suffice to yield an adequate approximation. Although, the power series is an asymptotic one² taking into account more terms will improve the accuracy for small enough ϵ . The physics application of WKB cited above may involve an ODE in space *or* time or a PDE in both. The time-scale separation parameter is typically the ratio of a wave oscillation period to some larger time-scale. The spatial scale-separation ratio is usually the ratio of a wave length to a typical spatial dimension of the structure through which the wave propagates. Here we are concerned with modeling the time evolution of a finite number of degrees of freedom; such dynamics are encoded in systems of coupled ODE's (in our case, (1)). A hierarchy of time scales can arise in control theory when, for instance, the characteristic frequencies of physical vibrations, oscillations, etc, are much larger than learning rates, velocities of desired trajectories, etc.

The original WKB method was applicable only to scalar ODEs or PDEs. In this note we have formulated a program which systematically extends the scalar WKB method to the case of a general system of coupled, time-dependent linear ODEs. We formally introduce the time-scale separation parameter, ϵ , by means of the following 'time-stretching' transformation, $t \rightarrow \epsilon t =: \tau$. Equation (2) now becomes $z'(t) \equiv \dot{z} = A(\epsilon t)z(t) \equiv -\mu x(\epsilon t)x^T(\epsilon t)z(t)$. We also define $z_\tau := \frac{d}{d\tau}z = \frac{1}{\epsilon}\dot{z}$, and similarly for τ derivatives of other quantities. In general, a dot or a prime superscript shall denote differentiation with respect to t . The advantage of this transformation is that the WKB-expansion order of any quantity can now be easily ascertained by counting ϵ powers. Thus, for instance: x , A , x_τ are all $O(\epsilon^0)$ as ϵ tends to zero; $\dot{x} = \epsilon x_\tau$, and $\dot{A} = \epsilon A_\tau$ are $O(\epsilon)$; $(x^T \dot{x})^2$ and \ddot{x} are

¹non-quantum physicists and applied mathematicians often use the acronym WKBJ, with the 'J' referring to Sir Harold Jeffreys.

²An asymptotic series is not guaranteed to have a non-zero radius of convergence.

$O(\epsilon^2)$. We always have the option of setting ϵ to unity which would bring us back to the form of equation (2). However, unless the elements of the matrix, $\dot{A}(\epsilon t)$ are in some sense small relative to some of the eigenvalues of A 'most of the time', none of our extended-WKB techniques (whether Matrix WKB or ITD) will furnish good *approximations* to the dynamics. However, we emphasize that even if neither A_τ nor ϵ are very small in any sense, the ITD formalism introduced in Section III below may still be used, in some cases, to prove stability theorems and establish rigorous bounds. We note that the system (1), investigated in this note, has the special feature that for $n > 1$, all but one of the eigenvalues of the matrix $A(\tau)$ *vanish* for all times, and thus cannot be larger in magnitude than any element of \dot{A} . As we shall see below this is not an obstruction to the applicability to the extended-WKB methods, provided that the single non-vanishing eigenvalue of $A(\tau)$ (which equals $-\mu \|x\|^2$) is $O(\epsilon^0)$.

III. ITERATIVE TRACKING DIAGONALIZATION (ITD) AND MAIN STABILITY RESULTS

Our ITD procedure consists of stages. The 0th stage refers to the original system given in (1). At each subsequent stage, we perform a transformation (detailed below) on the system. We will denote this by the tilde symbol. That is, \tilde{z} , \tilde{A} , etc. refer to the system at the first stage of the ITD. The symbols, $\tilde{\tilde{z}}$, $\tilde{\tilde{A}}$, etc. refer to the system at the second stage of the ITD.

The eigenvalues of the matrix A appear in (1) are given by (we assume that $\mu > 0$)

$$\lambda^{(j)} = \begin{cases} -\mu \|x\|^2 & \text{if } j = 1 \\ 0 & \text{if } j \in [2, n] \end{cases} \quad (2)$$

We impose the following convention on the eigenvalues of the time-varying system matrix for ITD stages one and higher (for \tilde{A} , $\tilde{\tilde{A}}$, etc.)

Definition 1: Eigenvalue Ordering Rule: Let $\lambda^{(j)}$ be the j th eigenvalue of the system matrix at ITD stage one or higher. Then, we order the eigenvalues such that $\lambda^{(n)} < \lambda^{(n-1)} < \dots < \lambda^{(1)}$.

Note that we cannot use the above definition (nor is it necessary) on the eigenvalues at the 0th level of the ITD.

Given this ordering of eigenvalues, we now have an ordered set of corresponding eigenvectors. Denote by $v^{(j)}$ the (right) eigenvector corresponding to $\lambda^{(j)}$. Note that each eigenvector can be multiplied by an arbitrary scalar function, $\phi^{(j)}(t) \in \mathbb{C}^\infty(\mathbb{R}^+, \mathbb{R})$.

Definition 2: Eigenvector scaling rule (only for iterates, \tilde{A} and higher): Let $\phi^{(j)}$ be chosen such that there exists an ϵ_1 , independent of time, such that for all $\epsilon \in [0, \epsilon_1)$, the dominant component of $v^{(j)}$ is the component $i = j$ and this component is unity: $v_i^{(j)} = 1$.

For the matrix A itself, the eigenvector corresponding to the non-zero eigenvalue is always in the direction of x . Take, $v^{(1)} = \frac{x}{\|x\|}$, and normalize the remaining $v^{(j)}$ for $j \in [2, n]$ to have unit norms. Define the similarity transformation matrix: $R := [v^{(1)} \ v^{(2)} \ \dots \ v^{(n)}] \in \mathbb{C}^{n \times n}$. Since A is real and symmetric, R is real and orthogonal. However, the higher

iterates $\tilde{A}, \tilde{\tilde{A}}, \dots$, are not necessarily real nor symmetric, and thus the iterates $\tilde{R}, \tilde{\tilde{R}}, \dots$, are not necessarily real or orthogonal.

Iterative Tracking Diagonalization (ITD) Procedure

We introduce a similarity transformation by the change of variables: $z =: R\tilde{z}$. Substituting into (1) we have

$$\dot{\tilde{z}} = \tilde{A}\tilde{z} \quad (3)$$

where $\tilde{A} = R^{-1}AR - R^{-1}\dot{R}$. We also define: $B = -R^{-1}\dot{R}$. By the product differentiation rule and eigenvector orthonormality, we note the following:

$$v^{(j)T} \dot{v}^{(k)} = \begin{cases} -\dot{v}^{(j)} v^{(k)} & \text{if } k \neq j \\ 0 & \text{if } k = j \end{cases}$$

From $A = -\mu xx^T$ (equation (1)), and using the above eigenvalue and eigenvector ordering and scaling rules, we compute \tilde{A} to be

$$\tilde{A} = \begin{bmatrix} -\mu\|x\|^2 & v^{(2)T} \dot{v}^{(1)} & \dots & v^{(n)T} \dot{v}^{(1)} \\ -v^{(2)T} \dot{v}^{(1)} & 0 & & \vdots \\ & & \ddots & v^{(n)T} \dot{v}^{(n-1)} \\ -v^{(n)T} \dot{v}^{(1)} & & -v^{(n)T} \dot{v}^{(n-1)} & 0 \end{bmatrix} \quad (4)$$

At this point, we can execute a second ITD iteration by diagonalizing \tilde{A} and computing $\tilde{\lambda}^{(j)}$ and $\tilde{v}^{(j)}$ for $j = 1 \dots n$. Then we define: $\tilde{R} = [\tilde{v}^{(1)} \dots \tilde{v}^{(n)}]$, $\tilde{\tilde{A}} = \tilde{R}^{-1}\tilde{A}\tilde{R} - \tilde{R}^{-1}\dot{\tilde{R}}$, $\tilde{\tilde{B}} = -\tilde{R}^{-1}\dot{\tilde{R}}$. We now specialize to the case $n = 2$.

A. The Case $n = 2$

We now consider $x \in \mathbb{R}^{2 \times 1}$. We proceed with the ITD procedure. As before, define the (orthogonal) matrix

$$R = [n \ y]$$

where $n = \frac{x}{\|x\|}$, and y is orthogonal to n with, $\|y\| = 1$. Then, (4) becomes:

$$\tilde{A} = \begin{bmatrix} -\mu\|x\|^2 & \beta \\ -\beta & 0 \end{bmatrix} \quad (5)$$

$$\beta := y^T \dot{n} \quad (6)$$

Note that, B is equal to $-R^{-1}\dot{R} = \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}$ and $\|B\| = O(\epsilon)$, since $\beta = O(\epsilon)$. The eigenvalues of the matrix are:

$$\tilde{\lambda}^{(j)} = \frac{-\mu\|x\|^2}{2} \pm \frac{\sqrt{\mu^2\|x\|^4 - 4(y^T \dot{n})^2}}{2}$$

Notice that $-\mu\|x\|^2 + \sqrt{\mu^2\|x\|^4 - 4\beta^2} = O(\epsilon^2)$. This follows from by applying the binomial expansion to the square root term. Additionally, $-\mu\|x\|^2 - \sqrt{\mu^2\|x\|^4 - 4\beta^2} = O(1)$. Hence, we take

$$\tilde{\lambda}^{(1)} = \frac{-\mu\|x\|^2}{2} - \frac{\sqrt{\mu^2\|x\|^4 - 4(y^T \dot{n})^2}}{2} = O(1) \quad (7)$$

and

$$\tilde{\lambda}^{(2)} = \frac{-\mu\|x\|^2}{2} + \frac{\sqrt{\mu^2\|x\|^4 - 4(y^T \dot{n})^2}}{2} = O(\epsilon^2) \quad (8)$$

Next, we determine the corresponding eigenvectors. We have

$$\tilde{v}^{(1)} = \tilde{\phi}^{(1)} \left[\begin{array}{c} 1 \\ \left(-\frac{\mu\|x\|^2}{2} - \frac{\sqrt{\mu^2\|x\|^4 - 4(\beta)^2}}{2} \right) \frac{1}{\beta} \end{array} \right]$$

$$\tilde{v}^{(2)} = \tilde{\phi}^{(2)} \left[\begin{array}{c} 1 \\ \left(-\frac{\mu\|x\|^2}{2} + \frac{\sqrt{\mu^2\|x\|^4 - 4(\beta)^2}}{2} \right) \frac{1}{\beta} \end{array} \right] \quad (9)$$

where $\tilde{\phi}^{(j)}$, are scaling factors.

The second component of $\tilde{v}^{(1)}$ is of $O(\epsilon)$, and hence no rescaling is necessary. That is, we take $\tilde{\phi}^{(1)} = 1$. Note, however, the division by β in the second component of $\tilde{v}^{(1)}$ is not numerically robust for small ϵ . Hence, multiplying the second component by $\frac{\gamma}{\gamma}$, where $\gamma := \mu\|x\|^2 + \sqrt{\mu^2\|x\|^4 - 4\beta^2}$, we get

$$\tilde{v}^{(1)} = \begin{bmatrix} 1 \\ \frac{2\beta}{\gamma} \end{bmatrix} \quad (10)$$

The second component of $\tilde{v}^{(2)}$ is of $O(\epsilon^{-1})$, and hence by the eigenvector scaling rule, we scale $\tilde{v}^{(2)}$ by $\tilde{\phi}^{(2)} = \frac{2\beta}{\gamma}$ to get:

$$\tilde{v}^{(2)} = \begin{bmatrix} \frac{2\beta}{\gamma} \\ 1 \end{bmatrix} \quad (11)$$

Thus, we have

$$\tilde{R} = \begin{bmatrix} 1 & \frac{2\beta}{\gamma} \\ \frac{2\beta}{\gamma} & 1 \end{bmatrix} \quad (12)$$

and

$$\tilde{R}^{-1} = \frac{\gamma^2}{\gamma^2 - 4\beta^2} \begin{bmatrix} 1 & -\frac{2\beta}{\gamma} \\ -\frac{2\beta}{\gamma} & 1 \end{bmatrix} \quad (13)$$

Additionally, we have

$$\dot{\tilde{R}} = \begin{bmatrix} 0 & \left(\frac{2\beta}{\gamma} \right)' \\ \left(\frac{2\beta}{\gamma} \right)' & 0 \end{bmatrix}$$

recall (Section II) that a prime denotes differentiation with respect to the time variable, t . We have:

$$-\tilde{\tilde{B}} = \tilde{R}^{-1}\dot{\tilde{R}} = \frac{2\gamma^2}{\gamma^2 - 4\beta^2} \left(\frac{\beta}{\gamma} \right)' \begin{bmatrix} -\frac{2\beta}{\gamma} & 1 \\ 1 & -\frac{2\beta}{\gamma} \end{bmatrix}$$

It is easily verified that $\|\tilde{\tilde{B}}\| = O(\epsilon^2)$. We next iterate the ITD procedure by transforming $\tilde{z} =: \tilde{\tilde{R}}\tilde{\tilde{z}}$. This yields

$$\dot{\tilde{\tilde{z}}} = \tilde{\tilde{A}}\tilde{\tilde{z}} \quad (14)$$

where $\tilde{\tilde{A}} = \tilde{R}^{-1}(\tilde{A}\tilde{R} - \dot{\tilde{R}})$.

By direct calculation, we find

$$\tilde{\tilde{A}} = \frac{\gamma}{4\beta^2 - \gamma^2} \begin{bmatrix} \mu\|x\|^2 - 4\frac{\beta^2}{\gamma} - 4\frac{\beta}{\gamma} \left(\frac{\beta}{\gamma} \right)' & 2 \left(\frac{\beta}{\gamma} \right)' \\ 2 \left(\frac{\beta}{\gamma} \right)' & 4\frac{\beta^2}{\gamma} \left(1 - \mu\frac{\|x\|^2}{\gamma} \right) - 4\frac{\beta}{\gamma} \left(\frac{\beta}{\gamma} \right)' \end{bmatrix}$$

We next compute the eigenvalues of \tilde{A} . We find that

$$\tilde{\lambda}^{(j)} = S \cdot \left[\frac{d+a}{2} \pm \frac{1}{2} \sqrt{(d+a)^2 + 4(b^2 - 4ad)} \right]$$

where

$$\begin{aligned} a &= \mu \|x\|^2 - 4\frac{\beta^2}{\gamma} - 4\frac{\beta}{\gamma} \left(\frac{\beta}{\gamma}\right)' & b &= c = 2 \left(\frac{\beta}{\gamma}\right)' \\ d &= 4\frac{\beta^2}{\gamma} \left(1 - \mu \frac{\|x\|^2}{\gamma}\right) - 4\frac{\beta}{\gamma} \left(\frac{\beta}{\gamma}\right)' & S &= \frac{\gamma}{4\beta^2 - \gamma^2} \end{aligned}$$

Note that $d+a = \mu \|x\|^2 \left(1 - 4\frac{\beta^2}{\gamma^2}\right) - 8\frac{\beta}{\gamma} \left(\frac{\beta}{\gamma}\right)' = O(1)$. Additionally, $d+a > 0$ for small enough ϵ . Hence, we take

$$\tilde{\lambda}^{(1)} = S \cdot \left[\frac{d+a}{2} + \frac{1}{2} \sqrt{(d+a)^2 + 4(b^2 - 4ad)} \right] = O(1)$$

$$\tilde{\lambda}^{(2)} = S \cdot \left[\frac{d+a}{2} - \frac{1}{2} \sqrt{(d+a)^2 + 4(b^2 - 4ad)} \right] = O(\epsilon^2)$$

As before, the above follows from the binomial expansion of the square root. Next, we determine the corresponding eigenvectors,

$$\tilde{v}^{(1)} = \tilde{\phi}^{(1)} \begin{bmatrix} 1 \\ \frac{1}{2b} \left(d - a + \sqrt{(d-a)^2 + 4b^2} \right) \end{bmatrix}$$

Since $d - a = 8\frac{\beta^2}{\gamma} - \mu \|x\|^2 \left(1 + 4\frac{\beta^2}{\gamma^2}\right) < 0$. Hence, we have $d - a + \sqrt{(d-a)^2 + 4b^2} = O(\epsilon^4)$. Thus, the second component of $\tilde{v}^{(1)}$ is of order $O(\epsilon^2)$. The dominant component is $O(1)$, and a constant. Hence, we do not need to rescale, that is, $\tilde{\phi}^{(1)} = 1$. However, for numerical robustness, we would like to eliminate the division by b . Defining $\tilde{\gamma} := d - a - \sqrt{(d-a)^2 + 4b^2} = O(1)$, we multiply the second component by $\frac{\tilde{\gamma}}{2b}$ to get:

$$\tilde{v}^{(1)} = \begin{bmatrix} 1 \\ -\frac{2b}{\tilde{\gamma}} \end{bmatrix}$$

The second eigenvector is given by:

$$\tilde{v}^{(2)} = \tilde{\phi}^{(2)} \begin{bmatrix} 1 \\ \frac{\tilde{\gamma}}{2b} \end{bmatrix}$$

In this case, the dominant component is $O(\epsilon^{-2})$ and not a constant. Hence, we rescale by choosing $\tilde{\phi}^{(2)} = \frac{2b}{\tilde{\gamma}}$ to get:

$$\tilde{v}^{(2)} = \begin{bmatrix} \frac{2b}{\tilde{\gamma}} \\ 1 \end{bmatrix}$$

Thus,

$$\tilde{R} = \begin{bmatrix} 1 & \frac{2b}{\tilde{\gamma}} \\ -\frac{2b}{\tilde{\gamma}} & 1 \end{bmatrix}$$

and

$$\tilde{R}^{-1} = \frac{\tilde{\gamma}^2}{\tilde{\gamma}^2 + 4b^2} \begin{bmatrix} 1 & -\frac{2b}{\tilde{\gamma}} \\ \frac{2b}{\tilde{\gamma}} & 1 \end{bmatrix}$$

We now can compute $\tilde{B} = -\tilde{R}^{-1}\tilde{R}'$ and check its ϵ order. We find

$$\tilde{B} = -2 \frac{\tilde{\gamma}^2}{\tilde{\gamma}^2 + 4b^2} \begin{pmatrix} b \\ \tilde{\gamma} \end{pmatrix}' \begin{bmatrix} \frac{2b}{\tilde{\gamma}} & 1 \\ -1 & \frac{2b}{\tilde{\gamma}} \end{bmatrix}$$

Hence, $\|\tilde{B}(t)\| = O(\epsilon^3)$.

Theorem 1: Consider the linear time-varying system given in (1). Let $\lambda^{(j)}$ and $v^{(j)}$ denote the j th eigenvalue and eigenvector of $A(t)$, respectively. Let $R := [v^{(1)} \ v^{(2)} \ \dots \ v^{(n)}] \in \mathbb{C}^{n \times n}$ be the invertible similarity transformation matrix, and $B(t) := -R^{-1}\dot{R}(t)$. Suppose that

$$\operatorname{Re}(\lambda^{(j)}(t)) < -c_1 \epsilon^m \quad (15)$$

and

$$\|B(t)\| < c_2 \epsilon^{m+1} \quad (16)$$

hold for all $t \in [0, \infty)$, and $j \in [1, n]$ for some integer $m > 0$ and some time independent constants $c_1(\epsilon), c_2(\epsilon)$ with $c_j(\epsilon)$ continuous in some neighborhood of $\epsilon = 0$ and $c_1(0) > 0, c_2(0) > 0$. Then, there exists an $\epsilon_1 > 0$ such that for all $\epsilon \in (0, \epsilon_1)$, and for all initial conditions $z(0) \in \mathbb{R}^n$,

$$\|z(t; \epsilon)\| < \|z(0; \epsilon)\| e^{-\frac{c_1}{2}t} \quad (17)$$

Due to page limitations, we do not present the full proof.

ITD WKB Bounds Using the ITD-WKB procedure, we may also derive bounds on the components of $z(t)$. We only summarize the results. Apply the ITD procedure twice, $z(t) = R\tilde{R}\tilde{z}(t)$. Defining $\rho_{ij} := [R\tilde{R}]_{ij}$, we have,

$$\begin{aligned} z_1(t) &= \rho_{11}\tilde{z}_1 + \rho_{12}\tilde{z}_2 \\ z_2(t) &= \rho_{21}\tilde{z}_1 + \rho_{22}\tilde{z}_2 \end{aligned} \quad (18)$$

Our strategy is to bound each of the components $\tilde{z}_j(t)$ for $j = 1, 2$:

$$\tilde{z}_j^-(t) \leq \tilde{z}_j(t) \leq \tilde{z}_j^+(t) \quad (19)$$

and then use (18) to bound the components $z_j(t)$, paying attention to the signs of the ρ_{ij} components. It can be shown that the component $\tilde{z}_k^\pm(t)$ evolve as follows:

$$\begin{aligned} \begin{bmatrix} \tilde{z}_1^-(t) \\ \tilde{z}_2^-(t) \end{bmatrix} &= M^-(t) \begin{bmatrix} \tilde{z}_1^-(0) \\ \tilde{z}_2^-(0) \end{bmatrix} \\ \begin{bmatrix} \tilde{z}_1^+(t) \\ \tilde{z}_2^+(t) \end{bmatrix} &= M^+(t) \begin{bmatrix} \tilde{z}_1^+(0) \\ \tilde{z}_2^+(0) \end{bmatrix} \end{aligned} \quad (20)$$

where each of $M^+(0)$ and $M^-(0)$ is the two by two identity matrix. In turn, $M^+(t)$ is evolved via the linear matrix differential equation:

$$\frac{d}{dt} M^+(t) = \begin{bmatrix} -\tilde{c}_1^{(1)+} + \tilde{c}_2 \epsilon^3 & \tilde{c}_2 \epsilon^3 \\ \tilde{c}_2 \epsilon^3 & -\tilde{c}_1^{(2)+} + \epsilon^2 + \tilde{c}_2 \epsilon^3 \end{bmatrix} M^+(t) \quad (21)$$

where \tilde{c}_j are constants. $M^-(t)$ is written in terms of its components

$$M^-(t) := \begin{bmatrix} \sigma_1(t) & \sigma_2(t) \\ \sigma_3(t) & \sigma_4(t) \end{bmatrix}$$

where the two columns are time-evolved in accordance with the following two vector ODE's:

$$\frac{d}{dt} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} -\tilde{c}_1^{(1)-} - \tilde{c}_2 \epsilon^3 & \tilde{c}_2 \epsilon^3 \\ -\tilde{c}_2 \epsilon^3 & -\tilde{c}_1^{(2)+} \epsilon^2 + \tilde{c}_2 \epsilon^3 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix}$$

$$\frac{d}{dt} \begin{bmatrix} \sigma_3 \\ \sigma_4 \end{bmatrix} = \begin{bmatrix} -\tilde{c}_1^{(1)+} + \tilde{c}_2 \epsilon^3 & -\tilde{c}_2 \epsilon^3 \\ \tilde{c}_2 \epsilon^3 & -\tilde{c}_1^{(2)-} \epsilon^2 - \tilde{c}_2 \epsilon^3 \end{bmatrix} \begin{bmatrix} \sigma_3 \\ \sigma_4 \end{bmatrix} \quad (22)$$

with initial conditions $\sigma_1(0) = \sigma_4(0) = 1$, and $\sigma_2(0) = \sigma_3(0) = 0$. The equations (18-22) can be used to derive upper and lower bounds for the individual components $z_j(t)$ or for their (individual or combined) norms.

IV. THE ITD-WKB APPROXIMATION AND SIMULATIONS

In this section, we present a specific example for which we illustrate the power of the ITD method. In the next subsection, we will illustrate the WKB approximation of the system at each level of the ITD. We will observe an increasing level of accuracy in the WKB approximation at each successive stage in the ITD. At the final stage, we will show that a low order Matrix-WKB of \tilde{z} can then be transformed to yield an excellent approximation of $z(t)$. We term this approximation as the ITD-WKB approximation, and denote the approximation as $z_{\text{ITD-WKB}}$. To summarize our approach, we

- 1) Apply the ITD transformation twice
- 2) Perform a WKB approximation
- 3) Invert the ITD transformation twice

In the last subsection, we will compare our bounds with those obtained using the theorems of [11]. There, Sondhi and Mitra derive two different upper bounds and one lower bound valid for all learning rates, μ . The bounding techniques they use critically depend on the PE and boundedness assumptions. In this section we specialize to a particular $n = 2$: the input vector signal is chosen to be

$$x(t) = \begin{bmatrix} \sin(\epsilon t) \\ \bar{a} \cos(\epsilon t) \end{bmatrix} \quad (23)$$

A. ITD-WKB Approximations

In this subsection, we utilize table I in [6] to construct an $N = 1$ term approximation³ for z , \tilde{z} , and $\tilde{\tilde{z}}$. We will denote the approximant to each ITD iteration as z_{WKB} , \tilde{z}_{WKB} , and $\tilde{\tilde{z}}_{\text{WKB}}$, respectively. That is, z_{WKB} , \tilde{z}_{WKB} , and $\tilde{\tilde{z}}_{\text{WKB}}$ are the one term WKB approximations to the solution of equations (1), (3), and (14), respectively. Application of table I of [6] requires a time-stretching transformation of system (1). By applying the stretching transformation, we can read off the N orders as explicit powers of ϵ . Setting $\tau = \epsilon t$ in (1), we have

$$\dot{z} = \epsilon z_\tau = -\mu x(\tau) x^T(\tau) z(\tau) \quad (24)$$

In the following development, we will apply the stretching transformation at each level in the ITD, and then apply the Matrix WKB method to construct an approximant. Due to

³The index N of the Matrix WKB expansion is defined in [6].

page limitations, we do not detail the required steps in the construction of the WKB approximation at each level of the ITD. The parameters of the system were taken as $\epsilon = 0.1$, $\bar{a} = 2$ and $\mu = 1$.

Fig. 1. illustrates the increasing accuracy of the WKB approximation at each stage of the ITD iteration. As can be seen in Fig. 1., $\tilde{\tilde{z}}_{\text{WKB}}$ is essentially indistinguishable from $\tilde{\tilde{z}}$. Since $z = R\tilde{R}\tilde{\tilde{z}}$, we consider the ITD-WKB approximation, $z_{\text{ITD-WKB}} := R\tilde{R}\tilde{\tilde{z}}_{\text{WKB}}$, and compare this to the true solution, z in (1). This comparison is depicted in Fig. 1. Clearly, $z_{\text{ITD-WKB}}$ is an excellent approximation of z .

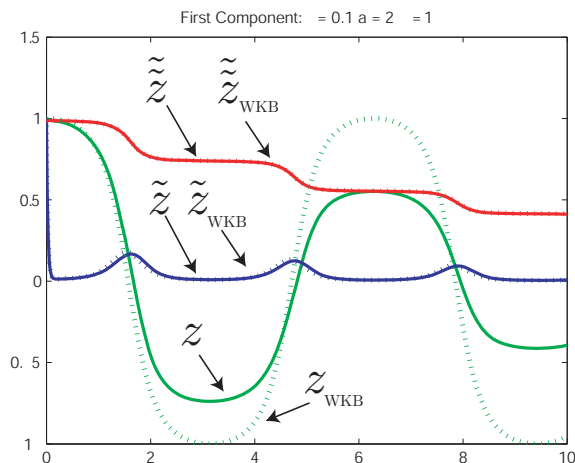


Fig. 1. Comparison of the first component of z at each stage of the ITD. We plot z_1 , \tilde{z}_1 , $\tilde{\tilde{z}}_1$, $z_{1\text{WKB}}$, $\tilde{z}_{1\text{WKB}}$, $\tilde{\tilde{z}}_{1\text{WKB}}$. The numerical solution is in solid while the corresponding WKB approximation is plotted in dotted. We observe that at each stage of the ITD, the WKB approximation becomes more accurate.

B. Comparison of Bounds

In this section, we compare the upper and lower bounds derived using our ITD-WKB bounds equations (section III above), with the upper and lower bounds derived by Sondhi and Mitra [11]. The bounds in [11] are derived via classical persistence of excitation assumptions. To make a fair comparison to Sondhi and Mitra's bounds, we choose the bounding parameters to optimize their bounds while keeping fixed and equal all other common parameters. Due to page limitation we are not able to show the details. They derive [11] two upper and one lower bounds valid for all learning rates and initial conditions. They also derive an additional lower bound, however, that is valid locally, and for a restricted range of learning rates. Since our results hold globally, we restrict the comparison to the three bounds that hold globally.

We now compare the bounds of Sondhi and Mitra with the ITD-WKB bounds given in Section III. This is shown in Fig. 2. We do not show upper bound one of Sondhi and Mitra as this was found to be more conservative than upper bound two. The ITD-WKB-PE bounds are indicated in dotted lines. They were computed by taking the norm of the individual component bounds described in Section III. Bounds on the

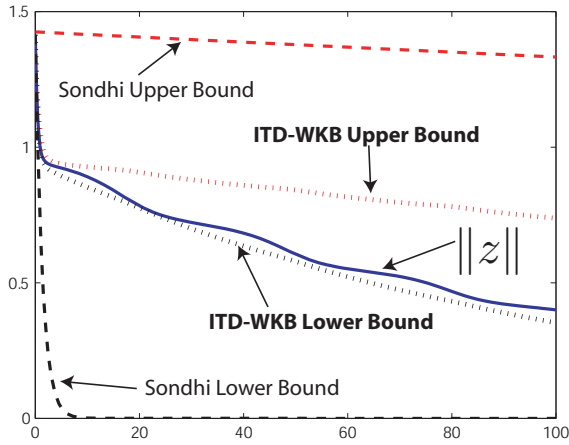


Fig. 2. Comparison of the ITD-WKB bounds with the bounds of Sondhi and Mitra for parameters $\epsilon = 0.1$, $\bar{a} = 1.25$, and $\mu = 1$. Note that our bounds involve *two* time constants, which enables them to bound the numerical more tightly.

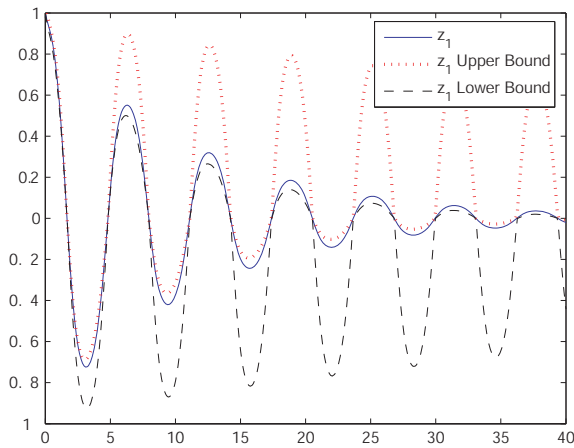


Fig. 3. ITD-WKB upper and lower bounds on the first component of z for parameters $\epsilon = 0.1$, $\bar{a} = 1.25$, and $\mu = 1$. Note that bounding the individual components is not possible using the methods given in [11], and thus illustrates the power of the ITD-WKB technique.

individual components of z is illustrated in Fig. 3. and Fig. 4. Bounding the individual components is not possible using the methods given in [11], and thus illustrate the power of the ITD-WKB technique.

V. CONCLUSION

In this note, we presented a powerful new technique applicable to linear, degenerate gradient flow. This technique required the combination of our Iterative Tracking Diagonalization (ITD) and WKB matrix method. The combined technique, which we name ITD-WKB, yielded excellent upper and lower bounds on the solution norm that captured the-two-time scale behavior of the system. Additionally, we were able to establish good upper and lower bounds on the

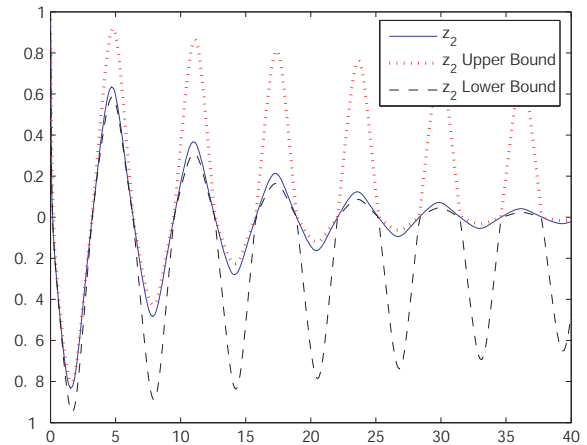


Fig. 4. ITD-WKB upper and lower bounds on the second component of z for parameters $\epsilon = 0.1$, $\bar{a} = 1.25$, and $\mu = 1$. Note that bounding the individual components is not possible using the methods given in [11], and thus illustrates the power of the ITD-WKB technique.

individual components. We illustrated the procedure with a specific example, and found the ITD-WKB bounds to be far superior to the classical results given in Sondhi and Mitra [11].

REFERENCES

- [1] B. O. Anderson. Exponential stability of linear equations arising in adaptive identification. *IEEE Trans. Automat. Contr.*, 22:8388, 1977.
- [2] S.B. Andersson and P.S. Krishnaprasad. Degenerate gradient flows: a comparison study of convergence rate estimates. *Proc. 41st Conf. Dec. Control*, 2002.
- [3] S. Boyd and S. Sastry. On parameter convergence in adaptive control. *Systems and Control Letters*, pages 311–319, 1983.
- [4] R.W. Brockett. The rate of descent for degenerate gradient flows. *Proc. MTNS*, 2000.
- [5] M.H. Holmes. *Introduction to Perturbation Methods*. Springer, 1995.
- [6] A. Ishihara and S. Ben-Menahem. A matrix wkb approach to feedforward and feedback control. *ASME IMECE, Chicago, IL, Nov. 5-10, 2006*.
- [7] A. Loria, A. Chaillet, G. Besancon, and Y. Chitour. On the pe stabilization of time-varying systems: open questions and preliminary answers. *Proc. 44th IEEE Conf. Decision Contr.*, page 68476852, 2005.
- [8] Morgan A. P. and Narendra K. S. On the uniform asymptotic stability of certain linear nonautonomous differential equations. *SIAM J. Control Optim.*, 15, 1977.
- [9] E. Panteley, A. Lora, and A. Teel. Relaxed persistency of excitation for uniform asymptotic stability. *IEEE Trans. on Automat. Contr.*, 46(12):18741886, 2001.
- [10] L. I. Schiff. *Quantum Mechanics*. McGraw-Hill, third edition, 1968.
- [11] M.M. Sondhi and D. Mitra. New results on the performance of a well-known class of adaptive filters. *Proc. of the IEEE*, 64(11):1583–1597, 1976.