# *The Solution of S exp(S) = A is Not Always the Lambert W Function of A*

Corless, Robert M. and Ding, Hui and Higham, Nicholas J. and Jeffrey, David J.

2007

MIMS EPrint: **2007.90**



ISSN 1749-9097

# The Solution of $S\exp(S) = A$ is Not Always the Lambert $W$ Function of $A$

Robert M. Corless*
Ontario Research Centre for Computer Algebra, and the Dept. of Applied Mathematics
University of Western Ontario
London, Canada
rcorless@uwo.ca

Hui Ding
Ontario Research Centre for Computer Algebra, and the Dept. of Applied Mathematics
University of Western Ontario
London, Canada
hding5@uwo.ca

Nicholas J. Higham
School of Mathematics
The University of Manchester
Manchester, UK
higham@ma.man.ac.uk

David J. Jeffrey
Ontario Research Centre for Computer Algebra, and the Dept. of Applied Mathematics
University of Western Ontario
London, Canada
djeffrey@uwo.ca

## ABSTRACT

We study the solutions of the matrix equation $S\exp(S) = A$. Our motivation comes from the study of systems of delay differential equations $y'(t) = Ay(t-1)$, which occur in some models of practical interest, especially in mathematical biology. This paper concentrates on the distinction between *evaluating a matrix function* and *solving a matrix equation*. In particular, it shows that the matrix Lambert $W$ function evaluated at the matrix $A$ does not represent all possible solutions of $S\exp(S) = A$. These results can easily be extended to more general matrix equations.

**Categories and Subject Descriptors:**
I.1.4 [**Symbolic Manipulation**] Applications

**General Terms:** Algorithms

**Keywords:** Matrix function; Lambert $W$ function; nonlinear matrix equation

## 1. INTRODUCTION

The authors of [7] examined a number of strategies for using computer algebra to solve scalar linear constant-coefficient delay differential equations with constant delays. In this paper, we look at some problems arising in nonlinear matrix equations and matrix functions that are motivated by the extension of the work of [7] to the matrix case. We begin with a simple problem: consider trying to find a differentiable function $y(t)$ such that

$$y'(t) = Ay(t-1)$$

where $A$ is an $n$-by-$n$ matrix of complex numbers, and $y(t)$ is specified on an initial vector history

$$y(t) = f(t) = [f_1(t),\, f_2(t),\, \ldots,\, f_n(t)]^{\mathrm{T}}$$

on the interval $-1 \le t \le 0$.

As in [7], this is a special problem, useful for some models in mathematical biology and elsewhere. Powerful numerical techniques exist for solving general delay differential equations; see for example [14]. The ultimate aim of the present work, in contrast, is to look for special-purpose techniques that may be more efficient for these 'niche' problems, or give greater insight. The approach used here is to note that the ansatz

$$y(t) = \exp(tS)c$$

for some constant $n$-vector $c$ (later to be used as one term in a Fourier-like series solution of the delay differential equation) leads to some interesting matrix computations, such as the computation of any and all $S$ such that[1]

$$S\exp(S) = A. \tag{1.1}$$

We shall consider this equation for $A, S \in \mathbb{C}^{n\times n}$.

Equation (1.1) is a matrix analogue of the scalar equation $se^s = a \in \mathbb{C}$, whose solutions are $s = W_k(a)$, where $W_k$ is the $k$th branch of the Lambert $W$ function [3]. Matrix functions can exhibit much more complicated behaviour than their scalar counterparts. For example, the number of square roots of an $n \times n$ matrix ($n \ge 2$) can vary from none to finitely many to infinitely many, depending on the matrix [8], [9]. Even the matrix exponential presents difficulties, both in computation and in characterizing ill conditioning [13]. The Lambert $W$ function is more akin, however, to the logarithm, and we may therefore expect that some of

*Corresponding author.

[1]Notation: We shall write $e^s$ for a scalar exponential, and $\exp(S)$ for a matrix exponential.


*ISSAC'07,* July 29–August 1, 2007, Waterloo, Ontario, Canada.

the lessons learned in computing the matrix logarithm will be of use in solving (1.1) [2]. Here, we will barely scratch the surface of numerical computation for this problem, concentrating mainly on the theoretical aspects, and we will return to the problem of numerical computation in a future paper.

## 2. MATRIX FUNCTIONS AND EQUATIONS

We first recall some definitions. For a more comprehensive discussion, see [9].

*Definition 1.* If

$$F(z) = \sum_{k\geq 0} a_k z^k$$

is a convergent power series in $|z| < r$, then the *matrix function* $F : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is defined to be

$$F(A) = \sum_{k\geq 0} a_k A^k \;, \tag{2.1}$$

which converges for $\rho(A) < r$, where $\rho$ is the spectral radius.

Recalling that any square matrix $A$ has the Schur decomposition $A = QTQ^*$, where $Q$ is unitary and the diagonal entries of $T$ are the eigenvalues of $A$, we may without loss of generality assume $A$ to be upper triangular, since $F(A) = QF(T)Q^*$. For theoretical work, it can be convenient to transform $A$ to Jordan canonical form, though this is not generally advisable for numerical computation. In upper triangular form, the diagonal entries of $F(T)$ are simply $F(t_{ii})$, and the upper triangle of $F(T)$ can be computed by the block recurrences of Parlett [5].

A more general definition is based on *polynomial interpolation.*

*Definition 2.* Let $A$ have distinct eigenvalues $\lambda_1, \ldots, \lambda_s$, with $\lambda_i$ having index $n_i$, where the index of an eigenvalue is the dimension of the largest Jordan block in which it appears. Let $r$ be the (unique) Hermite interpolating polynomial that satisfies

$$r^{(\ell)}(\lambda_i) = F^{(\ell)}(\lambda_i), \quad \ell = 0, 1, \ldots, n_i - 1, \quad i = 1, \ldots, s,$$

where $F^{(\ell)}$ and $r^{(\ell)}$ denote derivatives of order $\ell$. Then

$$F(A) = r(A). \tag{2.2}$$

It is this definition that is used in Maple's `MatrixFunction` command [11]. Specifically, it can be used with $F = W_k$ to define a matrix $W$ function.

*Remark 1.* If a matrix $A \in \mathbb{C}^{n\times n}$ happens to be a Jordan block, say $A = J_n(\mu)$ where

$$J_n(\mu) = \begin{bmatrix} \mu & 1 & & & \\ & \mu & 1 & & \\ & & \ddots & \ddots & \\ & & & \mu & 1 \\ & & & & \mu \end{bmatrix},$$

then because

$$F(z) = \sum_{\ell\geq 0} \frac{F^{(\ell)}(\mu)}{\ell!}(z-\mu)^\ell$$

and $(J_n(\mu) - \mu I)^\ell = 0$ for $\ell \geq n$ we see that $F(J_n(\mu)) =$

$$\begin{bmatrix} F(\mu) & F'(\mu) & \frac{1}{2}F''(\mu) & \cdots & \frac{1}{(n-1)!}F^{(n-1)}(\mu) \\ & F(\mu) & F'(\mu) & \ddots & \vdots \\ & & \ddots & \ddots & \\ & & & F(\mu) & F'(\mu) \\ & & & & F(\mu) \end{bmatrix}$$

is particularly simple. Note that if $F'(\mu) \neq 0$ this matrix has the same Jordan structure as $J_n(\mu)$ does:

$$F(J_n(\mu)) = ZJ_n(F(\mu))Z^{-1}$$

for some matrix $Z$. We will make use of this observation later.

We define the *primary matrix function* $W_k(A)$ to be the result of this interpolation definition with the singly branched scalar function $W_k$. This function is well-defined, for eigenvalues neither 0 nor $-e^{-1}$. If an eigenvalue is zero and the branch $k \neq 0$, the function is not finite. If an eigenvalue is $-e^{-1}$ and the branch $k \in \{-1, 0\}$, then because $W_k'(-e^{-1}) = \infty$ for these branches (and only these branches), the matrix function is not finite if $n \geq 2$.

The main aims of this work are to show that not all solutions to (1.1) are obtainable as $W_k(A)$ for some $k$ and then to characterize and classify all solutions of (1.1).

It is known in the context of other nonlinear matrix equations, such as $S^2 = A$ and $\exp(S) = A$, that not all solutions are obtainable as the appropriate inverse function of $A$ ($S = \sqrt{A}$ or $S = \log(A)$) [6]. In these two examples and in ours, the relevant inverse function is multibranched and we can mix branches for a particular eigenvalue on the diagonal of the triangular matrix in the Schur or Jordan form—something not allowed by Definition 2.

The organization of this paper is as follows. We begin in the next section with some comments on the scalar case. Then we derive some general results by elementary arguments based on properties of commuting matrices. The $2\times 2$ case is analyzed in Section 5 in order to get more insight. Finally, in Section 6 we present a complete characterization and classification of solutions of (1.1) for general $n$ when $-e^{-1}$ is not an eigenvalue of $A$. Some concluding remarks are given in Section 7.

## 3. THE SCALAR CASE

We first consider (1.1) for the case $n = 1$. We have $se^s = a$, or $s = W_k(a)$ where $W_k(z)$ is the $k$th branch of the Lambert $W$ function [3]. If $a = -e^{-1}$, then there is a double root $W_0(a) = W_{-1}(a) = -1$, together with a countable infinity of simple complex roots $W_k(a)$ for $k \neq 0, -1$; if $a \neq -e^{-1}$ then there is a countable infinity of simple roots only. The structure and asymptotic behaviour of these roots (equivalently, eigenvalues, or values of the Lambert $W$ function) are by now very well known: see e.g. [1]. In the paper [7] we see these values used to solve a scalar delay differential equation. We thus regard this problem as being solved in the $n = 1$ case: the solution to $s \exp s = a$ is exactly $s = W_k(a)$.

*Remark 2.* If $a = 0$, then $W_k(a) = -\infty$ for all branches except for the branch $k = 0$, where $W_0(0) = 0$; in this case the delay differential equation reduces to the ordinary differential equation $y'(t) = 0$ and is of little further interest. Some interest remains in this case when $n > 1$, however, as we shall see.

## 4. SOME GENERAL RESULTS

We are concerned from now on to discover the relations between the matrix function $W_k(A)$ defined above, and the solutions of equation (1.1). As with the functional definition (2.1), we can limit our discussion to upper triangular matrices $A$.

LEMMA 1. *Without loss of generality, we may take $A$ to be upper triangular in equation (1.1).*

**Proof**. If $S\exp(S) = A$, and $A = ZUZ^{-1}$ where $U$ is upper triangular, then $Z^{-1}SZZ^{-1}\exp(S)Z = U$ and if we put $X = Z^{-1}SZ$ we see that $X\exp(X) = U$, and thus any solution $S$ of the original equation is similar to a solution of the same equation where the input $U$ is upper triangular. Conversely, any solution $X$ of the upper triangular equation gives a solution $S = ZXZ^{-1}$ of the original equation. □

Henceforth we assume that $A$ is upper triangular. One question that we can then ask is: if $S\exp(S) = A$ and $A$ is upper triangular must $S$ also be upper triangular? We first establish another lemma.

LEMMA 2. *If $S\exp(S) = A$, then $S$ commutes with $A$.*

**Proof**. $S\exp(S) = A$ and hence

$$AS = S\exp(S)S = S^2\exp(S) = SA\,.$$

In fact, any function of $S$ commutes with $S$, as is well-known. □

The following lemma is from [12].

LEMMA 3. *Every matrix that commutes with $A \in \mathbb{C}^{n\times n}$ is a polynomial in $A$ if and only if no eigenvalue appears in more than one Jordan block in the Jordan canonical form of $A$ (that is, $A$ is nonderogatory).*

THEOREM 1. *If $A = S\exp(S)$ is upper triangular and nonderogatory then $S$ is upper triangular.*

**Proof**. Since $S$ commutes with $A$ by Lemma 2 and $A$ is nonderogatory, the result follows immediately from Lemma 3, since a polynomial in an upper triangular matrix is upper triangular. □

COROLLARY 1. *If $A$ is a single Jordan block $J(\lambda)$ then any solution of $S\exp(S) = A$ is upper triangular with constant diagonal $W_k(\lambda)$, for some $k$.*

**Proof**. $A$ is nonderogatory, so $S$ is upper triangular by Theorem 1. Moreover, $S$ is a polynomial in $A$, as the proof of Theorem 1 shows, so $s_{ii} = p(\lambda)$ is constant for all $i$ and necessarily equal to $W_k(\lambda)$ for some $k$. □

If $A$ is triangular and derogatory then solutions to (1.1) need not be triangular, as the next result shows with $A$ an extreme example of a derogatory matrix.

THEOREM 2. *If $A = \lambda I \in \mathbb{C}^{n\times n}$ there is a continuum of solutions to $S\exp(S) = A$ of the form*

$$S = P\,\mathrm{diag}(W_{k_1}(\lambda), W_{k_2}(\lambda), \ldots, W_{k_n}(\lambda))P^{-1},$$

*where $P$ is an arbitrary nonsingular matrix.*

**Proof.** By direct calculation, we have

$$S\exp(S) = P\,\mathrm{diag}\big(W_k(\lambda)e^{W_k(\lambda)}\big)P^{-1} = P\lambda IP^{-1} = \lambda I. \quad □$$

The only solutions in Theorem 2 that are obtainable from our definition of $W_k(A)$ are $W_k(A) = W_k(\lambda)I$, $k \in \mathbb{Z}$. Theorem 2 shows that further solutions can be obtained by taking a different branch on at least two copies of $\lambda$, but these solutions are not polynomials in $A$. As an example, consider the equation

$$S\exp(S) = \begin{bmatrix} -1/5 & 0 \\ 0 & -1/5 \end{bmatrix} = A.$$

One solution of this equation is

$$X = \begin{bmatrix} W_0(-1/5) & 0 \\ 0 & W_{-1}(-1/5) \end{bmatrix},$$

which, like $A$, is diagonal. Now take the unimodular matrix

$$P = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}.$$

The product $PXP^{-1}$ is

$$S = \begin{bmatrix} W_0(-1/5) & \alpha(W_{-1}(-1/5) - W_0(-1/5)) \\ 0 & W_{-1}(-1/5) \end{bmatrix},$$

which is a non-diagonal solution for all $\alpha \neq 0$. More generally, almost any full, nonsingular $P$ generates a non-triangular solution.

As an indication of the behaviour for more general derogatory upper triangular matrices, consider the matrix

$$A = \left[\begin{array}{cc|ccc} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ \hline 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{array}\right],$$

which is in Jordan form with a single distinct eigenvalue appearing in one $2 \times 2$ block and one $3 \times 3$ block. The same argument as in the proof of Theorem 1 says that $S$ commutes with $A$, but since $A$ is derogatory we can no longer invoke Lemma 3. Instead we can appeal to a standard result giving the general form of a matrix that commutes with a given matrix [12]. This result tells us that all matrices that commute with $A$ have the form

$$\left[\begin{array}{cc|ccc} \alpha_1 & \alpha_2 & 0 & \beta_1 & \beta_2 \\ 0 & \alpha_1 & 0 & 0 & \beta_1 \\ \hline \delta_1 & \delta_1 & \gamma_1 & \gamma_2 & \gamma_3 \\ 0 & \delta_1 & 0 & \gamma_1 & \gamma_2 \\ 0 & 0 & 0 & 0 & \gamma_1 \end{array}\right],$$

where the $\alpha_i$, $\beta_i$, $\delta_i$ and $\gamma_i$ are arbitrary parameters. The equation $S\exp(S) = A$ constrains these parameters in complicated way.

## 5. THE TWO-BY-TWO CASE

We now specialize to the case $n = 2$ in order to glean some more insight.

We consider first the case where $A$ is derogatory, which for $n = 2$ implies $A = \lambda I$ for some $\lambda \in \mathbb{C}$. Theorem 2 shows that there are infinitely many diagonalizable solutions to (1.1) for $\lambda \neq 0$. But these are not the only solutions. For example, if $A = \mathrm{diag}(-e^{-1}, -e^{-1})$ then an easy computation shows that the non-diagonalizable matrix

$$J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

satisfies $J\exp(J) = A$. For any nonsingular $Z$, defining $X = ZJZ^{-1}$ we have $X\exp(X) = ZJZ^{-1} \cdot Z\exp(J)Z^{-1} = ZAZ^{-1} = A = (-e^{-1})I$, so from $J$ we can obtain infinite families of non-diagonalizable solutions. For example, taking the unimodular matrix $P$,

$$P = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}.$$

we obtain

$$S = ZJZ^{-1} = \begin{bmatrix} -3 & 4 \\ -1 & 1 \end{bmatrix},$$

and it is easy to check that $S\exp(S) = (-e^{-1})I$.

THEOREM 3. *If $A \in \mathbb{C}^{2\times 2}$ is upper triangular and has distinct eigenvalues then there are a countably infinite number, and only a countably infinite number, of upper triangular matrices $S\exp(S) = A$.*

**Proof.** Since the eigenvalues of $A$ are distinct the following computation suffices. We know from Theorem 1 that any $S$ satisfying $S\exp(S) = A$ is upper triangular. So it suffices to construct all such $S$. We have

$$S = \begin{bmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix}$$

and, by direct computation,

$$\exp(S) = \begin{bmatrix} e^{s_{11}} & \frac{s_{12}(-e^{s_{22}}+e^{s_{11}})}{-s_{22}+s_{11}} \\ 0 & e^{s_{22}} \end{bmatrix}.$$

Therefore $s_{ii}e^{s_{ii}} = a_{ii}$, $i = 1, 2$, which means that $s_{11} = W_k(a_{11})$ and $s_{22} = W_\ell(a_{22})$ for some $k$ and $\ell$. The denominator of the $\{1, 2\}$ entry of $\exp(S)$ is not zero because $s_{11} = s_{22}$ would imply $a_{11} = a_{22}$, which is a contradiction. The $\{1, 2\}$ entry of $S\exp(S) = A$ can be manipulated to yield

$$s_{12} = \frac{a_{12}\,(s_{11} - s_{22})}{a_{11} - a_{22}}$$

uniquely, given any choice of values for $s_{11} = W_k(a_{11})$ and $s_{22} = W_\ell(a_{22})$. (This equation also follows, indeed more easily, from the commutativity of $A$ and $S$.) Therefore there is a bi-infinite family of matrices $S$ such that $S\exp(S) = A$. These are all the solutions. □

The question now arises whether there may be no solution to $S\exp(S) = A$. Unlike in the matrix square root equation $S^2 = A$, or the matrix logarithm equation $\exp(S) = A$, the answer is positive: there is always a solution for any $A$.

THEOREM 4. *In the $n = 2$ case, there is always a solution to $S\exp(S) = A$.*

**Proof**. By Theorems 2 and 3, the only remaining case is where the Jordan canonical form is nontrivial:

$$A = Z\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}Z^{-1}. \tag{5.1}$$

Choose $\mu = W_k(\lambda)$ for some $k \in \mathbb{Z}$. If $\lambda = 0$, then only $k = 0$ works. If $\lambda = -e^{-1}$, choose $k \notin \{0, -1\}$. Then put

$$S = Z\begin{bmatrix} \mu & \exp(-\mu)/\,(1+\mu) \\ 0 & \mu \end{bmatrix}Z^{-1}$$

and a short computation gives $F(S) = A$. Note $W_k'(\lambda) = \exp(-\mu)/(1+\mu)$. □

*Remark 3.* The case $\lambda = -e^{-1}$ in (5.1) is very special: even though there are an infinity of solutions, there do not appear to be any solutions corresponding to $k = 0$ or $k = -1$, as there are for matrices $A$ that have distinct eigenvalues, or nontrivial JCF with eigenvalues different from $-e^{-1}$. This means that for real matrices $A$ in this case there are no real solutions. This difficult case extends (in a nontrivial way, as we shall see) to larger Jordan blocks with the same eigenvalue, $-e^{-1}$.

*Remark 4.* For our original motivating example, $y'(t) = Ay(t-1)$, the degenerate case $A = J_n(0)$ (and so $S = W_0(J_n(0)) = J_n(0) = A$) represents another difficulty for the ansatz $y = \exp(St)C$. It turns out that the behaviour of $\exp(St) = \exp(At) = I + At$ is adequate dynamically, but there is not enough freedom available in the choice of constants $C$ to match the initial history data—because all the other components $W_k(0) = -\infty$ and thus the constants for those branches are not relevant. This means that this ansatz is unlikely to prove useful numerically in the case when the eigenvalues of $A$ are both very small.

*Remark 5.* Consider the numerical computation of Lambert $W$ of

$$A = \begin{bmatrix} -e^{-1}(1+i\varepsilon) & 1 \\ 0 & -e^{-1}(1-i\varepsilon) \end{bmatrix},$$

where $\varepsilon > 0$. Then $W_{-1}(A)$, $W_0(A)$ and $W_1(A)$ all exist, but they may be difficult to compute. We have that, for example, $W_0(A) =$

$$\begin{bmatrix} -1 + O(\sqrt{\varepsilon}) & e/\sqrt{\varepsilon} - 2e/3 + O(\sqrt{\varepsilon}) \\ 0 & -1 + O(\sqrt{\varepsilon}) \end{bmatrix}$$

to $O(\sqrt{\varepsilon})$ and similarly for the others. One should compare the well-known difficulties with the computation of the matrix square root for matrices near to one that has no square root [11]. Here, since the $\{1, 2\}$ entry is $O(1/\sqrt{\varepsilon})$ as $\varepsilon \to 0$, and we get closer to the branch point where $W_0(x)$ and $W_{-1}(x)$ fail to have derivatives, we anticipate the same kind of numerical problems as occur in the computation of the matrix square root.

*Remark 6.* The secular case of $y'(t) = Ay(t-1)$, that is when two eigenvalues of $A$ coalesce at $-e^{-1}$ giving rise to a solution of the form $y(t) = t\exp(-t)$, may occur in the solution of systems of DDE as it does for ordinary differential equations and for the scalar case of DDE [7]. As in the scalar case, we must add a secular term to the solution; as in the scalar case, this suffices to get an accurate solution.

*Remark 7.* Finally, we note that if $S$ is not constrained to be upper triangular, then the problem of taking the matrix exponential (when the entries of $S$ are symbols) as part of any procedure for solving $S\exp(S) = A$ becomes rather complicated. Consider the $5 \times 5$ example at the end of Section 4; even with such a modest example, the nonlinear equations that arise in $S\exp(S) = A$ are daunting for hand calculation, and the complexity grows rapidly enough with dimension that symbolic methods are in all likelihood not going to be useful for dimensions much larger than 5. Special-purpose numerical schemes for taking the matrix exponential may need to be devised in order to carry out any numerical scheme for solving $S\exp(S) = A$ efficiently and stably.

## 6. THE GENERAL CASE

The following results give a complete characterization of the solutions of the matrix equation $S\exp(S) = A$ when $A$ has no eigenvalue equal to $-e^{-1}$.

THEOREM 5. *Let $A \in \mathbb{C}^{n\times n}$ with no eigenvalue equal to $-e^{-1}$ have the Jordan canonical form $A = ZJZ^{-1}$, where $J = \operatorname{diag}(J_k(\lambda_k))$ with $p$ Jordan blocks. Further, let $L_k^{(j)} = W_j(J_k(\lambda_k))$. All solutions to $S\exp(S) = A$ are given by*

$$S = ZU\operatorname{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)})U^{-1}Z^{-1},$$

*where $j_k \in \mathbb{Z}$ is an arbitrary integer and $U$ is an arbitrary nonsingular matrix that commutes with $J$.*

**Proof.** Let $F(z) = ze^z$ and let $S$ be any solution of $F(S) = A$. The eigenvalues of $F(S)$ are $F(\mu)$, where $\mu$ is an eigenvalue of $S$, and so no eigenvalue of $S$ can be $-1$ by the assumption on $A$. Since $F'(\mu) = (1+\mu)e^{\mu} \neq 0$ for any eigenvalue $\mu$ of $S$, for every Jordan block $J(\mu)$ in $S$ there is a Jordan block of the same size in $F(S)$ associated with $F(\mu)$ (that is, no Jordan block of $S$ splits or merges when $F(S)$ is formed) [10], [12]; this is the key fact used in this proof, and it was illustrated at the end of Section 2. Hence $S$ has Jordan canonical form $J_S = \operatorname{diag}(J_k(\mu_k))$, where $F(\mu_k) = \lambda_k$ and hence $\mu_k = W_{j_k}(\lambda_k)$ for some $j_k \in \mathbb{Z}$.

Now consider $L = \operatorname{diag}(L_k)$, where $L_k = W_{j_k}(J_k(\lambda_k))$. We have $F(L) = J$. In other words, our definition of $S = W_{j_k}(A)$ via Definition 2 ensures that $S$ satisfies $S\exp(S) = A$. This is a special case of the more general result that the composition of a matrix function and its inverse is the identity: $F(F^{-1}(A)) = A$, assuming $F$ is single-valued [12]. So by the same argument as above, $L$ has Jordan form $J_S$, and so $S = TLT^{-1}$ for some nonsingular $T$.

But $F(S) = A$ implies $TJT^{-1} = TF(L)T^{-1} = ZJZ^{-1}$, or $(Z^{-1}T)J = J(Z^{-1}T)$. The result now follows on setting $U = Z^{-1}T$. □

COROLLARY 2. *Let $J(\lambda) \in \mathbb{C}^{m\times m}$ be a Jordan block with $\lambda \neq -e^{-1}$. Then for each $j \in \mathbb{Z}$ the equation $S\exp(S) = J$ has exactly one solution, $W_j(J)$, having eigenvalue $W_j(\lambda)$ (on the $j$th branch of $W$).*

The final result classifies the solutions into those that are primary matrix functions of $A$ and those that are not.

THEOREM 6. *Let $A \in \mathbb{C}^{n\times n}$ with no eigenvalue $-e^{-1}$ have the Jordan canonical form $A = ZJZ^{-1}$, where $J = \operatorname{diag}(J_k(\lambda_k))$, with $p$ Jordan blocks, and let $s \leq p$ be the number of distinct eigenvalues of $A$.*

*If $s = p$ then $S\exp(S) = A$ has a countable infinity of solutions that are primary matrix functions of $A$, given by*

$$S_j = Z\operatorname{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)})Z^{-1},$$

*corresponding to all possible choices of $j_1, \dots, j_p \in \mathbb{Z}$, subject to the constraint that $j_i = j_k$ whenever $\lambda_i = \lambda_k$.*

*If $s < p$ then $S\exp(S) = A$ has non-primary solutions. They form parametrized families*

$$S_j(U) = ZU\operatorname{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)})U^{-1}Z^{-1},$$

*where $j_k \in \mathbb{Z}$, $U$ is an arbitrary nonsingular matrix that commutes with $J$, and for each $j$ there exist $i$ and $k$, depending on $j$, such that $\lambda_i = \lambda_k$ while $j_i \neq j_k$.*

**Proof.** The proof consists of showing that for the solutions in Theorem 5 for which $j_i = j_k$ whenever $\lambda_i = \lambda_k$,

$$\begin{aligned} &U\operatorname{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)})U^{-1} \\ &\quad = \operatorname{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)}) \ , \end{aligned}$$

that is, $U$ commutes with the block diagonal matrix in the middle. This commutativity follows from the explicit form for $U$ provided by [12] and the fact that upper triangular Toeplitz matrices commute. □

## 7. CONCLUDING REMARKS

We have shown that $S\exp(S) = A$ may have solutions not obtainable as primary matrix functions $W_k(A)$ for any branch $k$, specifically whenever $A$ is nonzero and derogatory. In particular, there exist diagonal (or triangular) $A$ for which non-diagonal (or non-triangular) solutions exist. For our original motivating example, these complications may or may not play a direct role, because $y'(t) = Ay(t-1)$ may decouple into smaller systems that may be solved (with $W$) independently. However, in the case when $A$ is known only approximately, it seems very likely that the presence of these 'degenerate' conditions will make numerical solution of $S\exp(S) = A$ difficult, and possibly obviate any advantage of this approach.

The standard definition of a function of a matrix allows MAPLE to compute `LambertW`$(k, A)$. The theme of this paper has been that this is insufficient for solving the associated matrix equation. This suggests that a separate MAPLE routine is required to solve matrix equations.

The results of this paper are quite general, and not restricted to the particular equation $S\exp(S) = A$. In future work, we expect to explicitly extend these results to matrix equations of particular interest, such as $S + \log(S) = A$ where the Wright $\omega$ function and the matrix unwinding number will play a role [4].

We have concentrated on the complex case, because in the solution of delay differential equations all complex nonlinear eigenvalues $W_k(\lambda)$ are needed to represent the initial history. One may extract results about the real case, taking only $k \in \{0, -1\}$ for $-e^{-1} \leq \lambda < 0$ and only $k = 0$ for $\lambda \geq 0$, but we do not do this here (note, however, the real nonexistence result for $\lambda = -e^{-1}$ in Remark 3).

Finally, numerical work in solving nonlinear matrix equations such as $S\exp(S) = A$ has several potential difficulties: near-nonexistence of certain solutions near double branch points; continua of solutions in the derogatory case; and vanishingly small impact of the roots of $\mu \exp \mu = \lambda$ when the eigenvalues $\lambda$ of $A$ are small.[2]

These difficulties should be borne in mind when investigating the solution of more complicated nonlinear matrix equations such as $S = A\exp(-S)+B$ which may arise in the solution of (not much) more complicated delay differential equations.


## Acknowledgements

This work of Corless, Ding and Jeffrey was carried out with the support of the Natural Sciences and Engineering Research Council of Canada, and of the Mathematics of Information Technology and Complex Systems. Higham's work was supported by a Royal Society-Wolfson Research Merit Award.


## 8. REFERENCES


[1] R. Bellman and K. L. Cooke. *Differential-Difference Equations.* Academic Press, New York, 1963.

[2] S. H. Cheng, N. J. Higham, C. S. Kenney, and A. J. Laub. Approximating the logarithm of a matrix to specified accuracy. *SIAM J. Matrix Anal. Appl.*, 22(4):1112–1125, 2001.

[3] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert $W$ function. *Advances in Computational Mathematics*, 5:329–359, 1996.

[4] R. M. Corless and D. J. Jeffrey. *On the Wright $\omega$ function*, volume 2385 of *LNAI*, pages 76–89. Springer, Marseille, 2002.

[5] P. I. Davies and N. J. Higham. A Schur–Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[6] F. R. Gantmacher. *The Theory of Matrices*, volume one. Chelsea, New York, 1959.

[7] J. M. Heffernan and R. M. Corless. Solving some delay differential equations with computer algebra. *Mathematical Scientist*, 31(1):21–34, June 2006.

[8] N. J. Higham. Stable iterations for the matrix square root. *Numerical Algorithms 15*, 2:227–242, 1997.

[9] N. J. Higham. Functions of matrices. In L. Hogben, editor, *Handbook of Linear Algebra*, pages 11.1–11.13. Chapman and Hall/CRC, Boca Raton, FL, USA, 2006.

[10] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[11] D. J. Jeffrey and R. M. Corless. Linear algebra in Maple. In L. Hogben, editor, *Handbook of Linear Algebra*, pages 72.1–72.21. Chapman and Hall/CRC, Boca Raton, FL, USA, 2006.

[12] P. Lancaster and M. Tismenetsky. *The Theory of Matrices.* Academic Press, London, second edition, 1985.

[13] C. B. Moler and C. F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.*, 45(1):3–49, 2003.

[14] L. F. Shampine and S. Thompson. Solving DDEs in MATLAB. *Appl. Numer. Math.*, 37:441–458, 2001.


[2]Because of the analogy with the scalar ODE $y'(t) = Ay(t)$ having solutions with components $y(t) = \exp(\lambda t)$, the numbers $\mu$ such that components $y(t) = \exp(\mu t)$ appear in the solution of $y'(t) = Ay(t-1)$ are sometimes called *nonlinear eigenvalues* in the literature.