INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng 2006; **68**:267–297
Published online 30 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nme.1716

Mitigating the integration error in numerical simulations of Newtonian systems

Pini Gurfil*,† and Itzik Klein‡

Faculty of Aerospace Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel

SUMMARY

We introduce a method for mitigating the numerical integration errors of linear, second-order initial value problems. We propose a methodology for constructing an optimal state-space representation that gives minimum numerical truncation error, and in this sense, is the optimal state-space representation for modelling given phase-space dynamics. To that end, we utilize a simple transformation of the state-space equations into their variational form. This process introduces an inherent freedom, similar to the gauge freedom in electromagnetism. We then utilize the gauge function to reduce the numerical integration error. We show that by choosing an appropriate gauge function the numerical integration error dramatically decreases and one can achieve much better accuracy compared to the standard state variables for a given time-step. Moreover, we derive general expressions yielding the optimal gauge functions given a Newtonian one degree-of-freedom ODE. For the n degrees-of-freedom case we describe MATLAB® code capable of finding the optimal gauge functions and integrating the given system using the gauge-optimized integration algorithm. In all of our illustrating examples, the gauge-optimized integration outperforms the integration using standard state variables by a few orders of magnitude. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: initial value problems; variation of parameters; linear ordinary differential equations; gauge theory

1. INTRODUCTION

There are many important problems in engineering and science that require the solution of ordinary differential equations (ODEs). The study of differential equations dates back to the dawning of calculus [1–3] in the seventeenth century. Newton classified the first-order differential equations, while Leibniz discovered procedures for solving them. Following these pioneers,

*Correspondence to: Pini Gurfil, Faculty of Aerospace Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel.
†E-mail: pgurfil@technion.ac.il
‡E-mail: itzikml@012.net.il

Contract/grant sponsor: Asher Space Research Institute

Received 16 October 2005
Revised 15 February 2006
Accepted 15 February 2006

many analytical techniques for solving differential equations have been developed; however, there are many practical problems on which these analytical methods cannot be applied. It was Euler [4] who first attempted to overcome this situation by solving initial value problems (IVPs) by discretization.

Euler's idea was to propagate the solution of an IVP forward by a sequence of small time-steps. In each step, the rate of change of the solution is treated as constant and is found from a formula for the derivative evaluated at the beginning of the step. However, the accuracy of Euler's method is very limited and the method is usually stable only for small time-steps. Although one can improve the accuracy and stability of Euler's method by taking smaller time-steps, this approach requires an excessive number of steps to calculate a solution at a given interval. The integration scheme most commonly used to circumvent this problem usually relates to the family of methods known as the Runge–Kutta (RK) methods, named after two the renowned German mathematicians.

Carl David Runge worked for many years in the field of spectroscopy. The analysis of data led him to consider problems in numerical computation. The RK method originated in his 1895 paper on the numerical solution of differential equations [5]. Runge's idea was to derive the approximate solution based on improved formulas, such as the midpoint and trapezoidal rules. Evaluation of the derivative at the midpoint or endpoint of a step was performed by carrying out an Euler-type calculation to obtain a preliminary approximation to the solution at one of these points. The method was extended in 1901 to a system of differential equations by Wilhelm Kutta [6]. Additional high-order methods were subsequently developed by Nyström [7], Huťa [8], Butcher [9], Curtis [10, 11], Cooper and Verner [12] and Hairer [13].

Another method that was invented by Euler, seemingly unrelated to his work on IVP propagation, was variation of parameters (VOP) [14, 15], which is an analytical formalism for solving inhomogeneous (forced) differential equations. Euler invented this method for treating highly non-linear problems emerging in celestial mechanics. It was Joseph Louis Lagrange who perfected this method for deriving his system of equations describing the evolution of the orbital elements [16–18], known as Lagrange's Planetary Equations.

Although the VOP method can be used to solve inhomogeneous linear differential equations, the main feature of VOP is its generality. It can be applied to the solution of *any* differential equation, requiring no preliminary assumptions regarding the topology of the solution. According to the VOP method, the integration constants of the homogeneous solution are endowed with a time variation due to the presence of an external force. However, the transformation from the state variables of the original problem's phase space to the new state variables defined as the time-varying constants involves an inherent freedom, which, in practical calculations, can be removed by means of a user-defined constraint. The constraint may be essentially arbitrary insofar as it does not come into contradiction with the equations of motion written for the variable 'constants'.

The internal freedom emerges under the following circumstances: First, one should perturb some N-dimensional differential equation, and solve it by the VOP method (i.e. using the unperturbed generic solution $\mathbf{x}(t, C_1, \ldots, C_N)$ as an *ansatz*, and making its constants C_i time varying); second, the number of 'constants' promoted to variables must exceed N. Thus, when the said equations are written as equations for the new state variables C_i, the number of these variables will exceed that of equations; hence, the internal freedom. Mathematically, this freedom is analogous to the gauge freedom in electrodynamics, while the removal of this

freedom by imposing an arbitrary constraint is analogous to fixing of a gauge in the Maxwell theory.

The existence of the internal freedom in ODEs has long gone unnoticed due to the following long-established tradition. Whenever one begins with the unperturbed solution for the position and velocity as functions of time, $\mathbf{x}=\mathbf{x}(t, C_1, \ldots, C_N)$, $\mathbf{v}=\mathrm{d}\mathbf{x}(t, C_1, \ldots, C_N)/\mathrm{d}t$, he always assumes by default that the VOP solution of the perturbed equation will read identically: $\mathbf{x}=\mathbf{x}[t, C_1(t), \ldots, C_N(t)]$, $\mathbf{v}=\partial\mathbf{x}[t, C_1(t), \ldots, C_N(t)]/\partial t$. As noticed by Efroimsky [19], however, the VOP solution may look more general: $\mathbf{x}=\mathbf{x}[t, C_1(t), \ldots, C_N(t)]$, $\mathbf{v}=\partial\mathbf{x}[t, C_1(t), \ldots, C_N(t)]/\partial t+\mathbf{\Phi}$, where the convective term $\mathbf{\Phi}\equiv\sum(\partial\mathbf{x}/\partial C_i)\dot{C}_i$ may be chosen to be an *arbitrary* function of the time and of the 'constants' $C_i(t)$. This introduction of the gauge function $\mathbf{\Phi}$ has proven to be very fruitful in the problems of celestial mechanics [19–22]. Our goal here will be to broaden the applications of this concept.

In this paper, we show that there is a remarkable connection between the choice of the gauge function and the integration error produced when numerically solving IVPs. We study linear Newtonian second-order ODEs with n degrees-of-freedom (n-DOF) and transform each second-order ODE into two separate first-order IVPs by following the VOP formalism. A method for minimizing the numerical integration error of the resulting IVPs via the choice of an *optimal gauge function*, minimizing the quadratic norm of the RK truncation error, will be introduced. We refer to the new method as *gauge-optimized integration*.

Implementation of gauge-optimized integration for linear systems reduces the numerical integration error by *several orders of magnitude*, and in many cases nullifies it completely, so that the only remaining source of numerical discrepancies is the computer round-off error. This allows precise integration of large-scale second-order linear systems (encountered in ubiquitous engineering disciplines), for which a closed-form solution via calculation of the state transition matrix is impractical.

Lastly, we provide closed-form expressions for the optimal gauge function minimizing the numerical integration error, and develop a MATLAB® code which automatically recasts any n-DOF system into the VOP-based IVPs and finds the optimal gauge function, which minimizes the numerical integration error. To demonstrate the newly developed methodology we give several practical examples taken from various engineering fields.

2. BACKGROUND

Problems involving high-order ODEs can often be reduced to a system of first-order ODEs by introducing new state variables. Thus, consider the non-linear non-autonomous IVP

$$\begin{aligned}\dot{\mathbf{x}}&=\mathbf{y}(t, \mathbf{x})\\ \mathbf{x}(a)&=\boldsymbol{\alpha}\end{aligned}\tag{1}$$

where $\mathbf{x}=[x_1, x_2, \ldots, x_n]^{\mathrm{T}}\in D\subseteq\mathbb{R}^n$, $\dot{\mathbf{x}}=\mathrm{d}\mathbf{x}/\mathrm{d}t$, $\mathbf{y}:\mathbb{R}_{\geqslant 0}\times D\rightarrow D$ and $\boldsymbol{\alpha}$ is the initial condition vector at $t=a$ in the interval $t_m\leqslant t\leqslant t_{m+1}$. One of the most commonly used methods for numerically integrating (1) is the classical fourth-order Runge–Kutta (RK4) method, given by

$$\mathbf{x}_{m+1}=\mathbf{x}_m+\frac{h}{6}(\mathbf{k}_{m1}+2\mathbf{k}_{m2}+2\mathbf{k}_{m3}+\mathbf{k}_{m4})\tag{2}$$

where h is the time-step and

$$\begin{aligned}\mathbf{k}_{m1} &= \mathbf{y}(t_m, \mathbf{x}_m)\\ \mathbf{k}_{m2} &= \mathbf{y}(t_m + \tfrac{1}{2}h, \mathbf{x}_m + \tfrac{1}{2}h\mathbf{k}_{m1})\\ \mathbf{k}_{m3} &= \mathbf{y}(t_m + \tfrac{1}{2}h, \mathbf{x}_m + \tfrac{1}{2}h\mathbf{k}_{m2})\\ \mathbf{k}_{m4} &= \mathbf{y}(t_m + h, \mathbf{x}_m + h\mathbf{k}_{m3})\end{aligned} \tag{3}$$

If the IVP can be transformed into a linear quadrature problem of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{y}(t)\\ \mathbf{x}(a) &= \boldsymbol{\alpha}\end{aligned} \tag{4}$$

then Equations (3) reduce to

$$\begin{aligned}\mathbf{k}_{m1} &= \mathbf{y}(t_m)\\ \mathbf{k}_{m2} &= \mathbf{k}_{m3} = \mathbf{y}\left(t_m + \frac{h}{2}\right)\\ \mathbf{k}_{m4} &= \mathbf{y}(t_m + h)\end{aligned} \tag{5}$$

and Equation (2) can be re-written as

$$\mathbf{x}_{m+1} = \mathbf{x}_m + \frac{h}{6}[\mathbf{y}(t_m) + 4\mathbf{y}(t_m + h/2) + \mathbf{y}(t_m + h)] \tag{6}$$

leading to the well-known Simpson rule [23, 24].

As any other numerical integration scheme, the Simpson rule comprises two types of errors: The round-off error and the truncation error. The round-off error stems from the finite-precision arithmetic of a digital computer. Normally, the round-off error is not considered in the numerical analysis of the algorithm, since it strictly depends on the computer the algorithm is implemented upon, and is thus somewhat exogenous to the discrete approximation. The truncation error, however, is directly induced by truncating the infinite Taylor series forming the discrete approximation. This error depends on the size of the time-step, the order of the associated Taylor series, and the problem under consideration. The local truncation error of the Simpson rule for a single ODE is given by Gerald and Wheatley and Scarborough [24, 25]

$$E_s = -\frac{1}{90}h^5 y^{(4)}(\xi), \quad \xi \in (t_0, t) \tag{7}$$

where t_0 is the initial time.

In the next section, we shall use the VOP method to transform second-order ODEs into first-order variational equations, and show how this transformation can be used to minimize the numerical integration error.

3. THE EMERGENCE OF GAUGE FREEDOM IN VARIATION OF PARAMETERS

Although the VOP formalism is powerful enough to deal with non-linear ODEs, it is most often introduced as a method for solving inhomogeneous linear differential equations. The

transformation of the phase space from the original state variables to variational state variables introduces an inherent freedom, which can be used to nullify the RK truncation error. We shall dwell upon this issue by first examining the single degree-of-freedom (DOF) case.

3.1. One degree-of-freedom

Let us begin with the single DOF case. We shall then show that our new method can be straightforwardly implemented on multiple DOFs as well. To that end, consider the one-dimensional, second-order forced ODE

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = F(t), \quad x(t_0) = x_0, \quad \dot{x}(t_0) = \dot{x}_0 \tag{8}$$

where ω_n is the natural frequency, ξ is the damping coefficient and the forcing term $F(t)$ is piecewise continuous. Equation (8) was chosen as benchmark due to its ubiquity; it appears in diverse fields such as control theory [26], dynamics [27] and aeroelasticity [28]. The corresponding homogeneous equation is

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = 0 \tag{9}$$

Assuming an underdamped case ($\xi<1$), the homogeneous solution is given by

$$x_h = C_1 \cdot x_1(t) + C_2 \cdot x_2(t) \tag{10}$$

where $x_1(t)$ and $x_2(t)$ are the fundamental solutions

$$x_1(t) = \mathrm{e}^{-\xi\omega_n t}\cos(\omega_d t) \tag{11a}$$

$$x_2(t) = \mathrm{e}^{-\xi\omega_n t}\sin(\omega_d t) \tag{11b}$$

and $\omega_d = \omega_n\sqrt{1-\xi^2}$.

The main idea of the VOP method is to replace the constants C_1 and C_2 in Equation (10) with the time-dependent functions $C_1(t)$ and $C_2(t)$, respectively, in order to solve the inhomogenous Equation (8). Therefore, Equation (10) becomes

$$x(t) = C_1(t) \cdot x_1(t) + C_2(t) \cdot x_2(t) \tag{12}$$

Equation (12) constitutes a general solution candidate to the inhomogeneous Equation (8); i.e. adding time dependence to the coefficients of the homogenous solution may render it a general solution. Differentiating Equation (12) yields

$$\dot{x}(t) = \frac{\partial x}{\partial t} + \frac{\partial x}{\partial \mathbf{C}} \cdot \dot{\mathbf{C}} = \dot{x}_1(t) \cdot C_1(t) + \dot{x}_2(t) \cdot C_2(t) + \dot{C}_1(t) \cdot x_1(t) + \dot{C}_2(t) \cdot x_2(t) \tag{13}$$

where $\mathbf{C} = [C_1, C_2]^{\mathrm{T}}$. It is common to use the constraint [23]

$$\dot{C}_1(t) \cdot x_1(t) + \dot{C}_2(t) \cdot x_2(t) = 0 \tag{14}$$

i.e. to nullify the convective term $(\partial x/\partial \mathbf{C})\dot{\mathbf{C}}$, in order to facilitate the mathematical derivation and to solve for the excess freedom introduced by the mapping $(x_1, x_2) \mapsto (C_1, C_2, \dot{C}_1, \dot{C}_2)$.

However, by applying constraint (14) an inherent freedom of the VOP method vanishes. Instead, one may use the general form of this constraint, which can be written as

$$\Phi(\mathbf{C},t) \stackrel{\Delta}{=} \frac{\partial x}{\partial \mathbf{C}} \cdot \dot{\mathbf{C}} = \dot{C}_1(t) \cdot x_1(t) + \dot{C}_2(t) \cdot x_2(t) \tag{15}$$

where Φ is an *arbitrary*, *user-defined* smooth ($\mathscr{C}^\infty$) function referred to as the *gauge function*. We shall show in the sequel how the choice of the gauge function, Φ, affects the numerical integration error and how Φ can be used to considerably reduce and even completely eliminate this error.

To continue, we substitute Equation (15) into Equation (13), yielding

$$\dot{x}(t) = \Phi(t) + \dot{x}_1(t) \cdot C_1(t) + \dot{x}_2(t) \cdot C_2(t) \tag{16}$$

and

$$\ddot{x}(t) = \dot{\Phi}(t) + \dot{x}_1(t) \cdot \dot{C}_1(t) + \dot{x}_2(t) \cdot \dot{C}_2(t) + \ddot{x}_1(t) \cdot C_1(t) + \ddot{x}_2(t) \cdot C_2(t) \tag{17}$$

Substituting (12), (16) and (17) into (9) and rearranging entails

$$\begin{aligned} &\dot{\Phi}(t) + 2\xi\omega_n\Phi(t) + C_1(t)[\ddot{x}_1(t) + 2\xi\omega_n\dot{x}_1(t) + \omega_n^2 x_1(t)] \\ &\quad + C_2(t)[\ddot{x}_2(t) + 2\xi\omega_n\dot{x}_2(t) + \omega_n^2 x_2(t)] + \dot{x}_1(t)\dot{C}_1(t) + \dot{x}_2(t)\dot{C}_2(t) = F(t) \end{aligned} \tag{18}$$

Each of the expressions in square brackets in Equation (18) equals zero because both x_1 and x_2 are solutions of the homogeneous equation (9). Therefore, Equation (18) reduces to

$$\dot{\Phi}(t) + 2\xi\omega_n\Phi(t) + \dot{x}_1(t)\dot{C}_1(t) + \dot{x}_2(t)\dot{C}_2(t) = F(t) \tag{19}$$

Equations (15) and (19) form a system of two linear algebraic equations for the derivatives $\dot{C}_1(t)$ and $\dot{C}_2(t)$. By solving this system we obtain

$$\dot{C}_1(t) = \frac{\Phi(t)\dot{x}_2(t) - x_2(t)[F(t) - \dot{\Phi}(t) - 2\xi\omega_n\Phi(t)]}{w[x_1(t), x_2(t)]} \tag{20a}$$

$$\dot{C}_2(t) = \frac{x_1(t)[F(t) - \dot{\Phi}(t) - 2\xi\omega_n\Phi(t)] - \Phi(t)\dot{x}_1(t)}{w[x_1(t), x_2(t)]} \tag{20b}$$

where $x_1(t)$ and $x_2(t)$ are given in Equation (11) and $w[x_1(t), x_2(t)]$ is the Wronskian determinant,

$$w[x_1(t), x_2(t)] = \begin{vmatrix} x_1(t) & x_2(t) \\ \dot{x}_1(t) & \dot{x}_2(t) \end{vmatrix} \tag{21}$$

Division by $w[x_1(t), x_2(t)]$ is permissible since $x_1(t)$ and $x_2(t)$ constitute a fundamental set of solutions, and therefore their Wronskain is always non-zero. This result is a generalization of Newman's example [29].

Thus, we have transformed the second-order ODE (8) into the two first-order ODEs (20). The initial conditions for system (20) are found from Equations (12) and (16):

$$x(t_0) = C_1(t_0) \cdot x_1(t_0) + C_2(t_0) \cdot x_2(t_0) \tag{22a}$$

$$\dot{x}(t_0) = \Phi(t_0) + \dot{x}_1(t_0) \cdot C_1(t_0) + \dot{x}_2(t_0) \cdot C_2(t_0) \tag{22b}$$

Solving (22) for $C_1(t_0)$ and $C_2(t_0)$ yields

$$\begin{aligned} C_1(t_0) &= \frac{-x_0\dot{x}_2(t_0) + x_2(t_0)\dot{x}_0 - x_2(t_0)\Phi(t_0)}{\dot{x}_1(t_0)x_2(t_0) - \dot{x}_2(t_0)x_1(t_0)} \\ C_2(t_0) &= -\frac{-x_0\dot{x}_1(t_0) + x_1(t_0)\dot{x}_0 - x_1(t_0)\Phi(t_0)}{\dot{x}_1(t_0)x_2(t_0) - \dot{x}_2(t_0)x_1(t_0)} \end{aligned} \tag{23}$$

Integrating system (20) with these initial conditions will yield the solution $x(t)$, given by Equation (12). It was shown that Equation (12) is the solution of the inhomogeneous system (8), comprising the time-dependent functions $C_1(t)$ and $C_2(t)$, which depend in turn on the user-defined gauge function. Will the selection of the gauge function affect the solution $x(t)$? From Cauchy's theory of existence and uniqueness, we know that there is only a single solution for $x(t)$ for any given initial conditions; thus, $x(t)$ must remain invariant to the selection of the gauge function.

Theorem 1 (*Boyce and Diprima* [*23*])
Consider the IVP

$$\ddot{x} + p(t)\dot{x} + q(t)x = F(t), \quad x(t_0) = x_0, \quad \dot{x}(t_0) = \dot{x}_0 \tag{24}$$

where p, q and F are continuous on an open interval I. Then there exists exactly one solution $x = \Psi(t)$ to this problem.

Corollary 1
The solution $x(t)$ is invariant to a selection of Φ.

Although the uniqueness of $x(t)$ emerges immediately from Theorem 1, it is interesting to show how the gauge function cancels out from the expression for $x(t)$. We perform this exercise in the Appendix.

3.2. n degrees-of-freedom

Now, consider the following n-DOF system:

$$\begin{aligned} &\ddot{x}_1 + c_{11}\dot{x}_1 + c_{12}\dot{x}_2 + \cdots + c_{1n}\dot{x}_n + k_{11}x_1 + \cdots + k_{1n}x_n = f_1(t) \\ &\ddot{x}_2 + c_{21}\dot{x}_1 + c_{22}\dot{x}_2 + \cdots + c_{2n}\dot{x}_n + k_{21}x_1 + \cdots + k_{2n}x_n = f_2(t) \\ &\qquad\qquad\qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ &\ddot{x}_n + c_{n1}\dot{x}_1 + c_{n2}\dot{x}_2 + \cdots + c_{nn}\dot{x}_n + k_{n1}x_1 + \cdots + k_{nn}x_n = f_n(t) \end{aligned} \tag{25}$$

or in matrix form

$$\ddot{\mathbf{x}} + C\dot{\mathbf{x}} + K\mathbf{x} = \mathbf{F}(t), \quad \dot{\mathbf{x}}(t_0) = \dot{\mathbf{x}}_0, \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{26}$$

The solution of the corresponding homogenous equation is

$$\mathbf{x}_h = C_1\mathbf{x}^{(1)} + \cdots + C_{2n}\mathbf{x}^{(2n)} \tag{27}$$

where $C_1, \ldots, C_{2n}$ are the integration constants and $\mathbf{x}^1, \ldots, \mathbf{x}^{2n}$ are the $2n$ linearly independent solutions of the n-DOF homogenous differential equation. Following the VOP method we replace the constants $C_1, \ldots, C_{2n}$ in Equation (27) by the time-dependent functions $C_1(t), \ldots, C_{2n}(t)$, respectively, in order to solve the inhomogenous Equation (26). Therefore, Equation (27) becomes

$$\mathbf{x} = C_1(t)\mathbf{x}^{(1)} + \cdots + C_{2n}(t)\mathbf{x}^{(2n)} \tag{28}$$

We denote $\mathbf{x}_k = [x_{1k}, x_{2k}, \ldots, x_{nk}]^{\mathrm{T}}$ to represent a solution of the homogenous equation (i.e. $x_{11} = \mathbf{x}_1^1$ refers to the 1st component of the 1st solution $\mathbf{x}^1$). Consider the matrix X_{s} whose columns are the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^{2n}$,

$$X_{\mathrm{s}} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,2n} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,2n} \end{bmatrix}_{[n\times 2n]} \tag{29}$$

We follow the same steps as in the single DOF case and map $(\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_{2n}) \mapsto (C_1, C_2, \ldots, C_{2n}, \dot{C}_1, \dot{C}_2, \ldots, \dot{C}_{2n})$. To perform that mapping let us define $\mathbf{\Phi} = [\phi_1, \ldots, \phi_n]^{\mathrm{T}}$ as the gauge function vector (each DOF has a gauge function defined in Equation (15)) and

$$W = \begin{bmatrix} \dot{X}_{\mathrm{s}} \\ X_{\mathrm{s}} \end{bmatrix}_{[2n\times 2n]}, \quad \Psi = \begin{bmatrix} \mathbf{F} - (\dot{\mathbf{\Phi}} + C\mathbf{\Phi}) \\ \mathbf{\Phi} \end{bmatrix}_{[2n\times 1]} \tag{30}$$

Thus, following the VOP procedure we obtain

$$\dot{C}(t)_{[2n\times 1]} = W^{-1}\Psi \tag{31}$$

The corresponding initial conditions are found via Equation (27) and its derivative and the original initial conditions given in Equation (26). Substituting Equation (31) into Equation (28) will yield the solution of system (26). Regardless of the choice of $\mathbf{\Phi}$ the solution of the system (26) does not change; to show that, we use a vector version of Theorem 1.

Theorem 2 (*Boyce and Diprima* [*23*])
If each entry of the matrices K and C of the linear system (26) is continuous on an open interval $\alpha < t < \beta$, then there exists a unique solution $\mathbf{x}(t)$ to system (26).

Following Theorem (2) we conclude that the solution exists and is unique, and thus, must be invariant to a selection of the gauge function vector; in other words,

Corollary 2
$\mathbf{x}$ is invariant under the variational symmetry transformation $\dot{\mathbf{x}} \mapsto \dot{\mathbf{x}} + \mathbf{\Phi}$.

Although the choice of a gauge function, $\mathbf{\Phi}$, does not affect the general solution for $\mathbf{x}(t)$, it has a significant effect on the *accuracy* of the numerical integration procedure. For many applications it is crucial to reduce the numerical integration error; this can be done by a judicious selection of $\mathbf{\Phi}$. In the next section we show how to find $\mathbf{\Phi}$ that minimizes the truncation error of the numerical integration procedure.

4. MINIMIZING THE INTEGRATION ERROR USING THE GAUGE FREEDOM

We define the numerical integration error of some state variable $(\cdot)$ as the difference between the true solution and the numerical solution:

$$e_{(\cdot)} = (\cdot)_{\text{true}} - (\cdot)_{\text{numerical}} \tag{32}$$

In this section we will demonstrate how to mitigate the numerical integration error of a (fixed-step) RK45 by several orders of magnitudes. This merit is rendered by applying a gauge-optimized integration based on the generalized VOP method described in the previous section.

4.1. One degree-of-freedom

In the previous section we transformed the original problem (one-dimensional, second-order) given in Equation (8) into the system of first-order IVPs (20). We will now illustrate the procedure for finding the *optimal* Φ that minimizes the numerical integration error. To that end, let us re-write Equations (20) into

$$\dot{C}_1(t) = f(t, \Phi) \tag{33}$$

$$\dot{C}_2(t) = g(t, \Phi) \tag{34}$$

The integration errors resulting from numerically integrating Equations (33)–(34) are calculated based upon Equation (7),

$$\begin{aligned} e_{C_1} &= -\tfrac{1}{90} h^5 f^{(4)}[\xi, \Phi(\xi)] \\ e_{C_2} &= -\tfrac{1}{90} h^5 g^{(4)}[\xi, \Phi(\xi)] \end{aligned} \tag{35}$$

where, as before, $\xi \in (t_0, t)$.

The total integration error of $x(t)$ is calculated as follows:

$$\begin{aligned} e_x &= x_1(t)e_{C_1} + x_2(t)e_{C_2} \\ &= x_1(t)[-\tfrac{1}{90}h^5 f^{(4)}(\xi, \Phi)] + x_2(t)[-\tfrac{1}{90}h^5 g^{(4)}(\xi, \Phi)] \end{aligned} \tag{36}$$

Ideally, if some Φ could be found such that $e_x \equiv 0, \ \forall \xi \in (t_0, t)$, then the only remaining integration error of $x(t)$ would be the numerical round-off error. This observation gives rise to the following optimization problem.

Find $\Phi^\star$ s.t.

$$\Phi^\star = \arg\min_{\Phi} J \tag{37}$$

where[§]

$$J = e_x^2 = \left(\frac{h^5}{90}\right)^2 [x_1(t) f^{(4)}(\xi, \Phi) + x_2(t) g^{(4)}(\xi, \Phi)]^2 \tag{38}$$

This variational problem may be converted into a simple parameter optimization problem by guessing a general parameter-dependent topology for Φ. For example, if the forcing term may be expanded into a Fourier series of the form

$$F(t) = \frac{a_0}{2} + \sum_{k=1}^{K_c} a_k \cos(k\omega_0 t) + \sum_{k=1}^{K_s} b_k \sin(k\omega_0 t) \tag{39}$$

Φ may also be chosen as a Fourier series:[¶]

$$\Phi_N(t) = \frac{A_0}{2} + \sum_{n=1}^{N_c} [A_n \cos(n\omega_0 t)] + \sum_{n=1}^{N_s} [B_n \sin(n\omega_0 t)] \tag{40}$$

Remark 1
The topology of the gauge function minimizing the numerical integration error depends on the topology of the forcing term. In this paper, we shall concentrate on periodic (quasi-periodic) forcing. Similar treatments can be developed for general forcing terms, affecting only the expressions for the gauge function, but not affecting the generality of the approach endorsed herein.

In order to minimize the numerical integration error we need to minimize the cost functional (37) with the coefficients of the proposed Fourier series serving as our optimization parameters. Therefore, Φ_N (given by Equation (40)) is substituted into Equation (37), and differentiated

[§]The 2-norm cost function considerably assibilates the calculation of the global minimum, $J = 0$, as shown in the sequel.

[¶]Under certain smoothness assumptions, one can represent any quasi-periodic function by a Fourier series [30, 31].

with respect to the Fourier series coefficients,

$$\begin{aligned}
\frac{\partial J}{\partial A_0} &= 0 \\
\frac{\partial J}{\partial A_1} &= 0 \\
\frac{\partial J}{\partial B_1} &= 0 \\
&\vdots \\
\frac{\partial J}{\partial A_{N_c}} &= 0 \\
\frac{\partial J}{\partial B_{N_s}} &= 0
\end{aligned} \tag{41}$$

This leads to $(N_c + N_s + 1)$ algebraic equations for $A_0, \ldots, A_{N_c}, B_1, \ldots, B_{N_s}$, which define the parameter optimization problem given by

$$\alpha^\star = \arg\min_{\alpha} J \tag{42}$$

where $\alpha = [A_0, A_1, \ldots, A_{N_c}, B_1, \ldots, B_{N_s}]^{\mathrm{T}}$ and $(\cdot)^\star$ denotes an optimum value.

Remark 2
When solving the algebraic system given in Equation (41), the equations may be linearly dependent, giving rise to an underdetermined system. To circumvent the linear dependence one can substitute $t = T$, since (41) must hold $\forall t$. By choosing different values of T, additional equations are constructed until the system of equations is exactly determined.

The truncation orders of the sine and cosine sub-series in the Fourier series (40) must be equal in order to achieve a local minimum solution. To prove this we shall utilize the following theorem.

Theorem 3 (*Horn and Johnson* [*32*])
A Hermitian matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if and only if all of its eigenvalues are non-negative.

Lemma 1
If $N_c = N_s = N$, then $\alpha^\star$ is a local minimum of J.

Proof
To prove that the solution $\alpha^\star$ is a local minimum point of the cost function J, the Hessian matrix will be constructed. Since the cost function J is a twice continuously differentiable

function, the Hessian matrix has the form [32]

$$H \equiv \frac{\partial^2 J}{\partial \alpha_i \partial \alpha_j}, \quad i, j = 1, 2, \ldots, N \tag{43}$$

One of the Hessian properties is that the mixed partials are equal,

$$\frac{\partial^2 J}{\partial \alpha_i \partial \alpha_j} = \frac{\partial^2 J}{\partial \alpha_j \partial \alpha_i}, \quad i, j = 1, 2, \ldots, N \tag{44}$$

thus H is a symmetric (Hermitian) matrix of the form

$$H = \begin{bmatrix} \dfrac{\partial^2 J}{\partial A_0^2} & \dfrac{\partial^2 J}{\partial A_0 \partial A_1} & \dfrac{\partial^2 J}{\partial A_0 \partial B_1} & \cdots & \dfrac{\partial^2 J}{\partial A_0 \partial B_n} \\ \dfrac{\partial^2 J}{\partial A_1 \partial A_0} & \dfrac{\partial^2 J}{\partial A_1^2} & \dfrac{\partial^2 J}{\partial A_1 \partial B_1} & \cdots & \dfrac{\partial^2 J}{\partial A_1 \partial B_n} \\ \dfrac{\partial^2 J}{\partial B_1 \partial A_0} & \dfrac{\partial^2 J}{\partial B_1 \partial A_1} & \dfrac{\partial^2 J}{\partial B_1^2} & \cdots & \dfrac{\partial^2 J}{\partial B_1 \partial B_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 J}{\partial B_n \partial A_0} & \dfrac{\partial^2 J}{\partial B_n \partial A_1} & \dfrac{\partial^2 J}{\partial B_n \partial B_1} & \cdots & \end{bmatrix} \tag{45}$$

Although the eigenvalues of the Hessian are time dependent, this fact does not have an influence on the following result. Thus, without loss of generality, we take $t = 0$. The diagonal elements of H are given by the following relationships:

$$\begin{aligned} \frac{\partial^2 J}{\partial A_k^2} &= \frac{1}{2}(-40\omega_n \xi k^3 + 80\omega_n^3 \xi^3 k - 40\omega_n^3 \xi k)^2 \\ &\triangleq \frac{1}{2}\eta_A^2(k), \quad k = 1, 2, \ldots, N \\ \frac{\partial^2 J}{\partial B_k^2} &= \frac{1}{2}(10k^4 + 2\omega_n^4 - 80\omega_n^2 \xi^2 k^2 + 20\omega_n^2 k^2 - 24\omega_n^4 \xi^2 + 32\omega_n^4 \xi^4)^2 \\ &\triangleq \frac{1}{2}\eta_B^2(k), \quad k = 1, 2, \ldots, N \\ \frac{\partial^2 J}{\partial A_0^2} &= \frac{1}{2}(16\omega_n^4 \xi^4 - 12\omega_n^4 \xi^2 + \omega_n^4)^2 \\ &\triangleq \frac{1}{2}\eta_0^2 \end{aligned} \tag{46}$$

and the off-diagonal terms are given by

$$\begin{aligned}
\frac{\partial^2 J}{\partial A_i \partial A_j} &= \frac{\partial^2 J}{\partial A_j \partial A_i} = \frac{1}{2}\eta_A(k=i)\eta_A(k=j), \quad i=1,2,\ldots,n, \quad j=1,2,\ldots,n, \quad i \neq j \\
\frac{\partial^2 J}{\partial B_i \partial B_j} &= \frac{\partial^2 J}{\partial B_j \partial B_i} = \frac{1}{2}\eta_B(k=i)\eta_B(k=j), \quad i=1,2,\ldots,n, \quad j=1,2,\ldots,n, \quad i \neq j \\
\frac{\partial^2 J}{\partial A_i \partial B_j} &= \frac{\partial^2 J}{\partial A_j \partial B_i} = \frac{1}{2}\eta_A(k=i)\eta_B(k=j), \quad i=1,2,\ldots,n, \quad j=1,2,\ldots,n, \quad i \neq j
\end{aligned} \tag{47}$$

The first row/column elements are

$$\begin{aligned}
\frac{\partial^2 J}{\partial A_0 \partial A_j} &= \frac{\partial^2 J}{\partial A_j \partial A_0} = \frac{1}{2}\eta_0\eta_A(k=j), \quad j=1,2,\ldots,n \\
\frac{\partial^2 J}{\partial A_0 \partial B_j} &= \frac{\partial^2 J}{\partial B_j \partial A_0} = \frac{1}{2}\eta_0\eta_B(k=j), \quad j=1,2,\ldots,n
\end{aligned} \tag{48}$$

Substitution of Equations (16)–(48) into Equation (45) entails

$$H(J) \stackrel{\Delta}{=} [h_{ij}(J)] = \frac{1}{2}\begin{bmatrix}
\eta_0^2 & \eta_0\eta_A(1) & \eta_0\eta_B(1) & \cdots & \eta_0\eta_B(n) \\
\eta_0\eta_A(1) & \frac{1}{2}\eta_A(1)^2 & \eta_A(1)\eta_B(1) & \cdots & \eta_A(1)\eta_B(n) \\
\eta_0\eta_B(1) & \eta_A(1)\eta_B(1) & \eta_B(1)^2 & \cdots & \eta_B(1)\eta_B(n) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\eta_0\eta_B(n) & \eta_A(1)\eta_B(n) & \eta_B(1)\eta_B(n) & \cdots & \eta_B(n)^2
\end{bmatrix} \tag{49}$$

Thus, the characteristic polynomial of the Hessian is given by

$$\lambda^{(N-1)}(\lambda - \mu) \tag{50}$$

where

$$\mu = \sum_{k=1}^{N} h_{kk}(J) \tag{51}$$

The eigenvalues of the Hessian are therefore

$$\lambda_1 = \mu, \quad \lambda_2 = \lambda_3 = \cdots = \lambda_N = 0 \tag{52}$$

Since h_{kk} is given by Equations (49), μ is a non-negative eigenvalue, therefore all the eigenvalues of H are non-negative, and, based on Theorem 3, H is a positive semidefinite matrix. Since $H = H|_{\alpha^\star}$ (the Hessian of the cost function J does not depend on the Fourier coefficients in $\alpha^\star$, as seen in Equations (46)–(48)), it follows that $H|_{\alpha^\star}$ is positive semidefinite, thus $\alpha^\star$ is the local minimum of the cost function J. □

To upgrade our result to a global minimum, consider the following Lemma.

Lemma 2
If $N = N_c = N_s = \max(K_s, K_c)$, then $\alpha^\star$ is the global minimum of J.

Proof
Consider the one-dimensional second-order forced ODE given by Equation (8), where the forcing function is given by Equation (39). We follow the procedure described in Section 3 to receive system (20). Then, we substitute the result into (37). The resulting cost function of the numerical integration error is

$$\begin{aligned} J = (&-\omega_n^4\Phi + 12\omega_n^4\Phi\xi^2 - 40\omega_n^3\xi^3\Phi^{(1)} + 10\omega_n^2\Phi^{(2)}4bn\ \sin(nt\omega_0)n^3\omega_0^3 + 4\omega_n^2 bn\ \sin(nt\omega_0)n\omega_0 \\ &- 12\omega_n bn\cos(nt\omega_0)n^2\omega_0^2\xi - 12\omega_n an\ \sin(nt\omega_0)n^2\omega_0^2\xi - 16\omega_n^4\xi^4\Phi - 40\omega_n^2\xi^2\Phi^{(2)} \\ &- 16\omega_n^2 bn\ \sin(nt\omega_0)n\omega_0\xi^2 - 4\omega_n^3 an\ \sin(nt\omega_0)\xi + 8\omega_n^3 bn\cos(nt\omega_0)\xi^3 - 20\omega_n\xi\Phi^{(3)} - 5\Phi^{(4)} \\ &- 4\omega_n^2 an\cos(nt\omega_0)n\omega_0 + 20\omega_n^3\xi\Phi^{(1)} - 4\omega_n^3 bn\cos(nt\omega_0)\xi + 8\omega_n^3 an\ \sin(nt\omega_0)\xi^3 \\ &+ 16\omega_n^2 an\cos(nt\omega_0)n\omega_0\xi^2 - 4an\cos(nt\omega_0)n^3\omega_0^3)^2 \end{aligned} \tag{53}$$

where $\Phi^{(n)}$ is the nth-order time derivative of Φ. Due to the non-negativity of J, if $\exists\Phi^\star : J = 0$, then the global minimum occurs when $J = 0$. Therefore, if there exists some gauge function that nullifies J, this gauge function is the global minimizer of J. Choose $N_c = N_s = N$ where $N = \max(K_c, K_s)$ and substitute Equation (88) (this equation is the general form of the optimal gauge function; we shall show how we drive it in the following section) into (53) to get $J = 0$. □

Remark 3
An alternative method for finding a minimizer to the numerical integration error is to solve a differential equation based on the cost function given by Equation (37) directly. As we want the numerical integration error to be zero, we can nullify the cost function and solve for $J(\Phi, \partial\Phi/\partial t, \ldots, \partial^n\Phi/\partial t^n) = 0$ to obtain a general solution for the gauge function. However, in most cases this equation cannot be easily solved to yield closed-form analytic solution.

4.2. n degrees-of-freedom

In the n-DOF case each DOF constitutes a component in the Pareto cost function vector. For each DOF we shall write a single cost function component similarly to the single DOF case:

$$J_i = \left(\frac{h^5}{90}\right)^2 [x_{1_i}(t)\dot{C}_{1_i}^{(4)} + \cdots + x_{2n_i}(t)\dot{C}_{2n_i}^{(4)}]^2 \tag{54}$$

where $i = 1, 2, \ldots, n$. Then, to construct the n-DOF cost function we take the 1-norm of the Pareto cost function vector

$$J = \sum_{i=1}^{n} J_i \tag{55}$$

If each component of the forcing vector can be expanded into a Fourier series,

$$f_i(t) = \frac{A_{0_i}}{2} + \sum_{k=1}^{K_{c_i}} (A_k)_i \cos(k\omega_{0_i} t) + \sum_{k=1}^{K_{s_i}} (B_k)_i \sin(k\omega_{0_i} t) \tag{56}$$

we can write each $\Phi_i(t)$ as a Fourier series of the form

$$\Phi_{N_i}(t) = \frac{A_{0_i}}{2} + \sum_{k=1}^{N} [A_{k_i} \cos(k\omega_{0_i} t)] + \sum_{k=1}^{N} [B_{k_i} \sin(k\omega_{0_i} t)] \tag{57}$$

Following Lemma 2 the truncation order, N, must be $N = \max(K_{s_1}, \ldots, K_{s_n}, K_{c_1}, \ldots, K_{c_n})$ to achieve the global minimum. Substituting all $\Phi_{N_i}(t)$ into the cost function (55), and differentiating with respect to the Fourier series coefficients,

$$\frac{\partial J_k}{\partial A_{0_1}} = 0, \ldots, \frac{\partial J_k}{\partial A_{0_n}} = 0, \quad \frac{\partial J_k}{\partial A_{1_1}} = 0, \ldots, \frac{\partial J_k}{\partial A_{N_n}} = 0, \quad \frac{\partial J_k}{\partial B_{1_1}} = 0, \ldots, \frac{\partial J_k}{\partial B_{N_n}} = 0 \tag{58}$$

will result in $(2N + 1)$ algebraic equations for the Fourier constants. System (58) defines the optimization problem. Thus, by solving system (58) we obtain an extremum of the multi-DOF cost function. If upon substitution into the cost function the result is zero, then the Fourier constants constitute the global minimizer of the cost function.

In the following section we demonstrate a method for finding an optimal gauge function to globally minimize the numerical integration error.

5. FINDING CLOSED-FORM EXPRESSIONS FOR THE OPTIMAL GAUGE FUNCTION

In the previous section we outlined a method for finding Φ to minimize the numerical integration error. Now we will derive a general expression for this optimal Φ, denoted by $\Phi^\star$. Let us consider first the case where $\xi = 0$, so Equation (8) reduces to

$$\ddot{x} + \omega_n^2 x = F(t) \tag{59}$$

Now, assume that

$$F(t) = F_s(t) = \sum_{k=1}^{K_s} a_n \sin(k\omega_0 t) \tag{60}$$

We will begin with $K_s = 1$, i.e. $F(t) = a_1 \sin(\omega_0 t)$ and then generalize the treatment. Substituting Φ_1 into J and differentiating with respect to A_0, A_1, B_1, results in the following set of equations:

$$\begin{aligned} \frac{\partial J}{\partial A_0} &= 0 \\ \frac{\partial J}{\partial A_1} &= 0 \\ \frac{\partial J}{\partial B_1} &= 0 \end{aligned} \tag{61}$$

Solving this set of equations for A_0, A_1, B_1 results in

$$\Phi_1 = -4\frac{(5\omega_0^6 + 11\omega_n^4\omega_0^2 + \omega_n^6 + 15\omega_n^2\omega_0^4)a_1\omega_0\cos(t\omega_0)}{(20\omega_n^6\omega_0^2 + \omega_n^8 + 110\omega_n^4\omega_0^4 + 100\omega_n^2\omega_0^6 + 25\omega_0^8)} \tag{62}$$

Following the same procedure as above for $K_s = 2$, we have

$$\Phi_2 = -8\frac{(320\omega_0^6 + 44\omega_n^4\omega_0^2 + \omega_n^6 + 240\omega_n^2\omega_0^4)a_2\omega_0\cos(2t\omega_0)}{(80\omega_n^6\omega_0^2 + \omega_n^8 + 1760\omega_n^4\omega_0^4 + 6400\omega_n^2\omega_0^6 + 6400\omega_0^8)} \tag{63}$$

and for $K_s = N$

$$\begin{aligned}\Phi_N &= -4N\frac{(5\omega_0^6N^6 + 11N^2\omega_n^4\omega_0^2 + \omega_n^6 + 15N^4\omega_n^2\omega_0^4)a_n\omega_0\cos(Nt\omega_0)}{(20\omega_n^6\omega_0^2N^2 + \omega_n^8 + 110\omega_n^4\omega_0^4N^4 + 100\omega_n^2\omega_0^6N^6 + 25\omega_0^8N^8)} \\ &\triangleq -a_n\psi\cos(Nt\omega_0)\end{aligned} \tag{64}$$

where

$$\psi = 4N\frac{(5\omega_0^6N^6 + 11N^2\omega_n^4\omega_0^2 + \omega_n^6 + 15N^4\omega_n^2\omega_0^4)\omega_0}{(20\omega_n^6\omega_0^2N^2 + \omega_n^8 + 110\omega_n^4\omega_0^4N^4 + 100\omega_n^2\omega_0^6N^6 + 25\omega_0^8N^8)} \tag{65}$$

Therefore, the general expression for Φ in the problem defined by (59)–(60) is given by

$$\Phi_{N_s} = -\sum_{k=1}^{N_s} a_k\psi(k)\cos(kt\omega_0) \tag{66}$$

Example 1
Integrate numerically the following ODE:

$$\ddot{x} + x = \sin(2t), \quad x(0) = 0, \quad \dot{x}(0) = 0 \tag{67}$$

□

Taking $F(t) = \sin(2t)$, $K_c = 0, K_s = 2, a_1 = 0, a_2 = 1$ and $\omega_0 = 1$, the cost function (37) becomes

$$J = \left[5\frac{\mathrm{d}^4\Phi}{\mathrm{d}t^4} - 10\frac{\mathrm{d}^2\Phi}{\mathrm{d}t^2} + 40\cos(2t) + \Phi\right]^2 \tag{68}$$

Following Lemma 2 we choose $N = K_s$. The candidate optimal gauge function is

$$\Phi_2 = \frac{A_0}{2} + A_1\cos(t) + B_1\sin(t) + A_2\sin(2t) + B_2\cos(2t) \tag{69}$$

Substituting $\Phi = \Phi_2$ into J and differentiating with respect to the Fourier series coefficients gives

$$\alpha_2^\star = \{A_0 = A_1 = B_1 = A_2 = 0 \quad B_2 = -\tfrac{40}{121}\} \tag{70}$$

thus,

$$\Phi^\star = -\tfrac{40}{121}\cos(2t) \tag{71}$$

The Hessian of the cost function is

$$H = \begin{bmatrix} 1/2 & 16\sin(t) & 16\cos(t) & 121\sin(2t) & 121\cos(2t) \\ 16\sin(t) & 512(\sin(t))^2 & 512\sin(t)\cos(t) & 3872\sin(2t)\sin(t) & 3872\cos(2t)\sin(t) \\ 16\cos(t) & 512\sin(t)\cos(t) & 512(\cos(t))^2 & 3872\cos(t)\sin(2t) & 3872\cos(t)\cos(2t) \\ 121\sin(2t) & 3872\sin(2t)\sin(t) & 3872\cos(t)\sin(2t) & 29282(\sin(2t))^2 & 29282\sin(2t)\cos(2t) \\ 121\cos(2t) & 3872\cos(2t)\sin(t) & 3872\cos(t)\cos(2t) & 29282\sin(2t)\cos(2t) & 29282(\cos(2t))^2 \end{bmatrix}$$

The eigenvalues of H are $0, 0, 0, 0, 29794.5$, i.e. non-negative. Following Lemma 2, $\alpha^\star$ is the global minimum of J if the cost is nullified. Substitution of Equation (71) into Equation (68) gives

$$J|_{\alpha^\star} = \{5[-\tfrac{640}{121}\cos(2t)] - 10[\tfrac{160}{121}\cos(2t)] + 40\cos(2t) - \tfrac{40}{121}\cos(2t)\}^2 = 0 \tag{72}$$

thus, $\alpha_2^\star$ gives indeed the global minimum of J.

Now, suppose that instead of taking $N = K_s$ we take $N = 1$. The resulting gauge function is

$$\Phi_1 = \frac{A_0}{2} + A_1\cos(t) + B_1\sin(t) \tag{73}$$

following the same procedure as above we find that

$$\alpha_1^\star = \{A_0 = 0, \quad A_1 = -\tfrac{5}{2}, \quad B_1 = \tfrac{5}{2}\} \tag{74}$$

thus,

$$\Phi_1 = -\tfrac{5}{2}\cos(t) + \tfrac{5}{2}\sin(t) \tag{75}$$

Therefore, the cost function in Equation (68) becomes

$$\begin{aligned} J|_{\alpha_1^\star} &= -3200\sin(t)\cos(t) + 6400\sin(t)\cos(t)^2 - 3200\sin(t) + 3200 - 6400\cos(t)^3 \\ &= -6400\cos(t)^2 + 3200\cos(t) + 6400\cos(t)^4 \neq 0 \end{aligned}$$

However, $\alpha_1^\star$ is a local minimum of J since the resulting eigenvalues of the new Hessian are non-negative. Now, suppose that we chose an arbitrary truncation order, $N_s = 0$, $N_c = 1$, the resulting gauge function is

$$\Phi_{1,0} = \frac{A_0}{2} + A_1\cos(t) \tag{76}$$

Repeating our calculations we find that

$$\Phi_{1,0} = 40 - 5\cos(t) \tag{77}$$

The resulting eigenvalues of the new Hessian are $\lambda_1 = 0$, $\lambda_2 = 0.5 + 512\cos(t)^2$ thus, the Hessian is time-dependant, and hence, the critical point is not necessarily a minimum point of the cost function.

Figure 1 shows the effect of the various gauge functions on the numerical integration error, e_x (the graphs in this and the following figures are based on actual numerical simulations). The graph is presented in a log–log scale to better illustrate the different orders of magnitude of the numerical integration error. As can be seen, the numerical integration error for $\Phi = 0$, which is the most common choice for implementing VOP methods, can be greatly improved if we choose an optimal Φ which is not necessarily zero.

More importantly, gauge-optimized integration considerably reduces the integration error obtained when using standard state variables. In Figure 2 we depict a comparison of integration errors between the gauge-optimized integration utilizing the VOP-based equations with the optimal gauge function $\Phi^\star = (40/121)\cos(2t)$ and the standard choice of state variables $x_1 = x$, $x_2 = \dot{x}$. As can be plainly seen, the gauge-optimized integration decreases the integration error by three orders of magnitude in the examined time interval. Moreover, the integration error using the standard state variables is diverging while the error of the gauge-optimized integration is bounded. Therefore, in a larger time interval, the use of gauge-optimized integration becomes increasingly important.

Remark 4

In our examples we chose the time-step to be $\Delta T = 0.05$. When the time-step is decreased, the numerical integration error decreases and the difference between the numerical integration errors of the gauge-optimized integration and the standard method becomes less significant; this implies that the gauge-optimized integration may be used to expedite numerical integration by using larger time intervals.

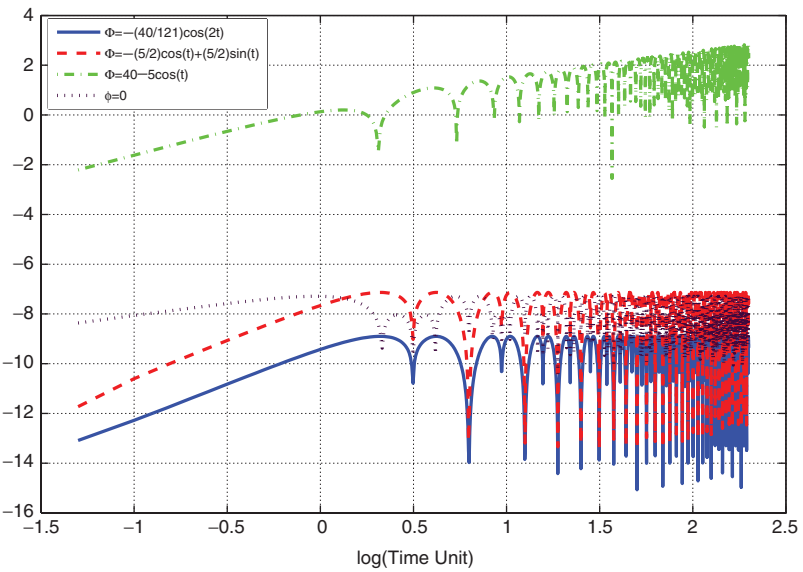

Figure 1. The effect of different gauge functions on the numerical integration error.

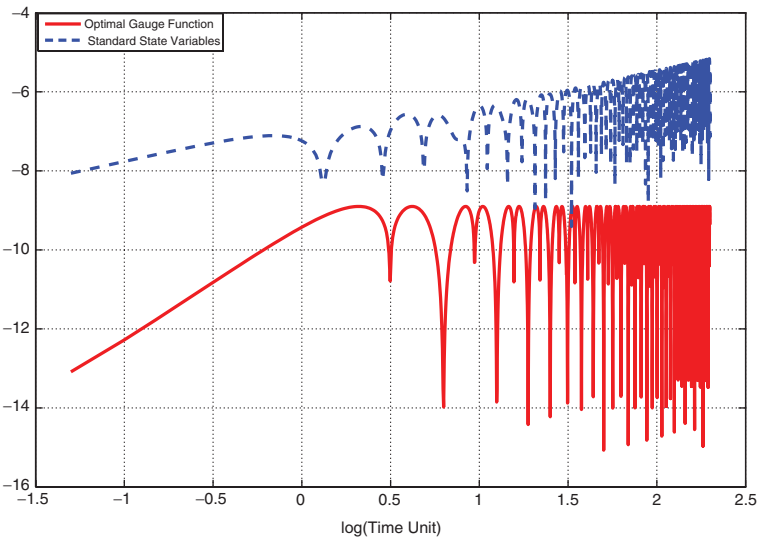

Figure 2. The numerical integration error of the gauge-optimized integration considerably reduces the integration error compared to standard choice of state variables.

To continue our discussion, let us follow the above steps for

$$F_{\mathrm{c}}(t) = \sum_{n=1}^{K_{\mathrm{c}}} b_n \cos(n\omega_0 t) \tag{78}$$

We find that

$$\Phi_{N_{\mathrm{c}}} = \sum_{n=1}^{N_{\mathrm{c}}} b_n \psi(n) \sin(nt\omega_0) \tag{79}$$

We note the following symmetry:

$$\begin{aligned} F_{\mathrm{s}}(t) = \sum_{n=1}^{K_{\mathrm{s}}} a_n \sin(n\omega_0 t) &\Longrightarrow \Phi_{N_{\mathrm{s}}} = -\sum_{n=1}^{K_{\mathrm{s}}} a_n \psi(n) \cos(n\omega_0 t) \\ F_{\mathrm{c}}(t) = \sum_{n=1}^{K_{\mathrm{c}}} b_n \cos(n\omega_0 t) &\Longrightarrow \Phi_{N_{\mathrm{c}}} = \sum_{n=1}^{K_{\mathrm{c}}} b_n \psi(n) \sin(n\omega_0 t) \end{aligned} \tag{80}$$

Thus, if $F(t) = F_{\mathrm{c}}(t) + F_{\mathrm{s}}(t)$, the general solution for Φ_N, minimizing the numerical integration error, is given by superposition of $\Phi_{N_{\mathrm{s}}}$ and $\Phi_{N_{\mathrm{c}}}$:

$$\Phi_N = \sum_{n=1}^{N} b_n \psi(n) \sin(n\omega_0 t) - \sum_{n=1}^{N} a_n \psi(n) \cos(n\omega_0 t) \tag{81}$$

and N is found via Lemma 2.

We now return to the general form of the one-dimensional ODE given by Equation (8):

$$\ddot{x} + 2\xi\omega\dot{x} + \omega_n^2 x = F(t)$$

Assuming that

$$F_{\mathrm{s}}(t)=\sum_{n=1}^{K_{\mathrm{s}}} a_n \sin(n\omega_0 t) \tag{82}$$

we follow the same procedure as in the undamped case, choose $K_{\mathrm{s}}=1,\ldots,N$ and find the corresponding Φ for each K_{s}. It can be shown that, for this case

$$\Phi_{N_{\mathrm{s}}}=\sum_{n=1}^{N_{\mathrm{s}}}\left[\frac{\rho_1(n)}{\rho(n)}\omega_n a_n \xi \sin(nt\omega_0)+\frac{\rho_2(n)}{\rho(n)}\omega_n a_n \cos(nt\omega_0)\right] \tag{83}$$

where

$$\begin{aligned}
\rho_1 &= 4[32\omega_n^6\xi^6+(-40\omega_n^6+32n^2\omega_n^4\omega_0^2)\xi^4+(-24n^2\omega_n^4\omega_0^2+14\omega_n^6+10n^4\omega_n^2\omega_0^4)\xi^2\\
&\quad +5\omega_0^6 n^6+5n^4\omega_n^2\omega_0^4+7n^2\omega_n^4\omega_0^2-\omega_n^6]\\
\rho_2 &= -4n(16\omega_n^6\xi^6+(16n^2\omega_n^4\omega_0^2-16\omega_n^6)\xi^4+(-12n^2\omega_n^4\omega_0^2+4\omega_n^6)\xi^2\\
&\quad +5\omega_0^6 n^6+11n^2\omega_n^4\omega_0^2+\omega_n^6+15n^4\omega_n^2\omega_0^4)\\
\rho &= (256\omega_n^8\xi^8+(320\omega_n^6\omega_0^2n^2-384\omega_n^8)\xi^6+(176\omega_n^8+160\omega_n^4\omega_0^4n^4-320\omega_n^6\omega_0^2n^2)\xi^4+(-24\omega_n^8\\
&\quad -120\omega_n^4\omega_0^4n^4+80\omega_n^6\omega_0^2n^2)\xi^2+20\omega_n^6\omega_0^2n^2+\omega_n^8+110\omega_n^4\omega_0^4n^4+100\omega_n^2\omega_0^6n^6+25\omega_0^8n^8)
\end{aligned} \tag{84}$$

Replacing the sine term with the cosine term in Equation (82) gives

$$F_{\mathrm{c}}(t)=\sum_{n=1}^{K_{\mathrm{c}}} b_n \cos(n\omega_0 t) \tag{85}$$

and

$$\Phi_{N_{\mathrm{c}}}=\sum_{n=1}^{N_{\mathrm{c}}}\left[-\frac{\rho_2(n)}{\rho(n)}\omega_n b_n \sin(nt\omega_0)+\frac{\rho_1(n)}{\rho(n)}\omega_n b_n \xi \cos(nt\omega_0)\right] \tag{86}$$

Note the symmetry in $\Phi_{N_{\mathrm{c}}}$ and $\Phi_{N_{\mathrm{s}}}$ as was pointed out in Equation (80).

Combining $F_{\mathrm{s}}(t)$ and $F_{\mathrm{c}}(t)$ yields

$$F(t)=\sum_{n=1}^{K_{\mathrm{s}}} a_n \sin(n\omega_0 t)+\sum_{n=1}^{K_{\mathrm{c}}} b_n \cos(n\omega_0 t) \tag{87}$$

Consequently, for this general case we have

$$\begin{aligned}
\Phi_N=\sum_{n=1}^{N}\Bigg\{&\frac{\rho_2(n)}{\rho(n)}\omega_n[a_n\cos(nt\omega_0)-b_n\sin(nt\omega_0)]\\
&+\frac{\rho_1(n)}{\rho(n)}\omega_n\xi[a_n\sin(nt\omega_0)+b_n\cos(nt\omega_0)]\Bigg\}
\end{aligned} \tag{88}$$

Thus, we constructed a general expression for the required optimal Φ to produce minimum numerical integration error (it should be kept in mind that the truncation order N is chosen via Lemma 2).

There are two special cases that need to be addressed:

1. $K_c = K_s = a_0 = 0$, hence, $F(t) = 0$, which is the homogenous case.
2. $K_c = K_s = 0$, hence, $F(t) = a_0/2$.

We shall first deal with the homogenous case. To this end, consider the following Lemma.

Lemma 3
If $F(t) = 0$, then $\Phi^\star = 0$.

Proof
The total error as given by Equation (36) depends on $\dot{C}_1$ and $\dot{C}_2$, given by Equation (20). Substitute $F(t) = 0$ into Equation (20) to get

$$\begin{aligned} \dot{C}_1(t) &= \frac{\Phi(t)\dot{x}_2(t) - x_2(t)[-\dot{\Phi}(t) - \omega_n^2\Phi(t)]}{w(x_1, x_2)(t)} \\ \dot{C}_2(t) &= \frac{x_1(t)[-\dot{\Phi}(t) - \omega_n^2\Phi(t)] - \Phi(t)\dot{x}_1(t)}{w(x_1, x_2)(t)} \end{aligned} \tag{89}$$

Zero numerical integration error occurs when $\dot{C}_1(t) = 0$ and $\dot{C}_2(t) = 0$, i.e. C_1, C_2 are constants. Choosing $\Phi = 0$ indeed reduces Equation (89) to

$$\begin{aligned} \dot{C}_1(t) &= 0 \\ \dot{C}_2(t) &= 0 \end{aligned} \tag{90}$$

Therefore, no numerical integration error is present when representing the system in the form (90). □

The last case to be examined is $F(t) = a_0/2$.

Lemma 4
If $F(t) = a_0/2$, then $\Phi^\star = a_0/(2\omega_n^2)$.

Proof
Let $F(t) = a_0/2$, then Equation (8) becomes

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = a_0/2 \tag{91}$$

Utilizing the transformation

$$\eta = x - \frac{a_0}{2\omega_n^2} \tag{92}$$

Equation (91) reduces to

$$\ddot{\eta} + 2\xi\omega_n\dot{\eta} + \omega_n^2\eta = 0 \tag{93}$$

Following Lemma 3, the required Φ_N to produce minimum numerical integration error is $\Phi_N = 0$ hence,

$$\Phi^\star = \frac{a_0}{2\omega_n^2} \tag{94}$$

□

We shall now consider the following example, summarizing the 1-DOF case.

Example 2
Integrate the following ODE:

$$\ddot{x} + \dot{x} + x = \cos(t), \quad x(0) = 1/4, \quad \dot{x}(0) = 0 \tag{95}$$

□

To find the gauge function minimizing the numerical integration error we shall utilize Equations (87)–(88). The initial conditions for the variational equations are found by Equation (23): $C_1(0) = -0.0090$, $C_2(0) = 0.25$. Also, $F(t) = \cos(t)$, $K_s = 0$, $K_c = 1$, $b_1 = 1$ and $\omega_0 = 1$. Thus, the resulting optimal gauge function is

$$\Phi^\star = \tfrac{121}{241}\sin(t) + \tfrac{32}{241}\cos(t) \tag{96}$$

We compare the numerical integration error obtained for this optimal gauge function to the case $\Phi = 0$ as well as to an integration using the standard choice of state variables $x_1 = x$, $x_2 = \dot{x}$. The results are depicted by Figure 3. Obviously, the gauge-optimized integration considerably reduces the integration error. Initially, the difference between the methods is five orders of magnitude. At the end of the time interval we can see that the difference is about one order of magnitude due to accumulation of round-off errors.

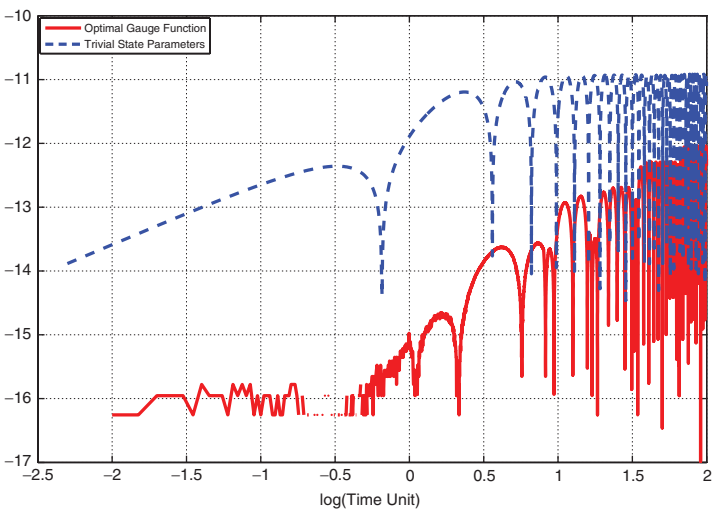

Figure 3. In the damped case as well, the numerical integration error of the gauge-optimized integration considerably reduces the integration error compared to standard choice of state variables.

6. THE n DEGREES-OF-FREEDOM MATLAB®-BASED GAUGE-OPTIMIZED INTEGRATOR: ILLUSTRATIVE EXAMPLES

Instead of finding closed-form expression for the gauge functions in the n-DOF system, we developed a MATLAB® program capable of deriving the optimal gauge functions for any n-DOF second-order system. The program performs the gauge-optimized integration as follows.

- VOP steps
  - Find the corresponding homogenous system.
  - Determine the number of degrees-of-freedom.
  - Solve the homogenous system and add time dependence to the integration constants ($C \rightarrow C(t)$).
  - Calculate the first derivative of the homogenous system.
  - Define the gauge functions.
  - Calculate the second derivative of the homogenous system.
  - Construct the system as a function of the gauge functions and $\dot{C}(t)$'s.
- Construct the cost function
  - Solve for the $\dot{C}(t)$'s.
  - Construct a Fourier series for the gauge functions via Equation (57).
  - Substitute the gauge functions into the $\dot{C}(t)$-equations and differentiate those equations four times.
  - Construct the cost function using Equation (55).
- Evaluating the gauge function coefficients
  - Differentiate the cost function with respect to the gauge function Fourier series coefficients.
  - Solve for the Fourier functions coefficients.
  - If the linear system is underdetermined, randomly choose a time period, substitute it into the system and create new equations until the problem is exactly determined.
  - Solve the above system and find the values of the coefficients.
- Last steps in the mapping of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2n}) \mapsto (C_1, C_2, \ldots, C_{2n}, \dot{C}_1, \dot{C}_2, \ldots, \dot{C}_{2n})$
  - Substitute the coefficient values into the $\dot{C}(t)$-equations and into the gauge functions.
  - Transform the initial conditions.
- Minimum check and integration
  - Build the Hessian of each cost function and find its eigenvalues. If all of the eigenvalues are non-negative, then the coefficients values represent a local minimum of each cost function.
  - Substitute the gauge functions into the cost functional and check whether the overall cost functional is nullified.
  - Integrate the $\dot{C}(t)$-equations to find $C(t)$.
  - Substitute the $C(t)$'s into the homogenous solution to construct the general solution of the inhomogeneous system.
- Optional steps
  - Evaluate the analytical solution of the system as entered by the user.
  - Calculate the error between the VOP solution and the analytical solution.

We shall now study a few examples taken from various fields of engineering. The first example is taken from the field of astrodynamics. We address the Clohessy–Wiltshire (CW)

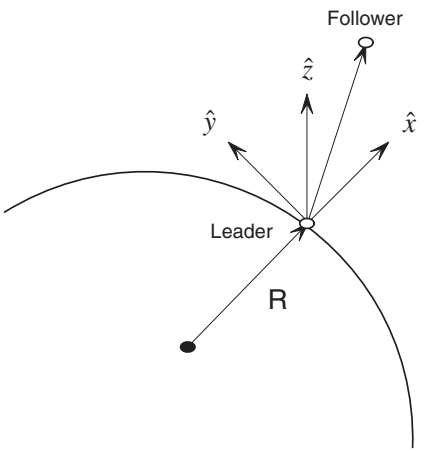

Figure 4. Relative motion of spacecraft in a leader-fixed rotating co-ordinate system.

equations [33], which model the 3-DOF linearized dynamics of a follower spacecraft relative to a leader spacecraft in a leader-fixed rotating reference frame, as shown in Figure 4. The CW equations for the relative position vector $[x, y, z]^{\mathrm{T}}$ are derived assuming that the leader is flying on a circular references orbit in a central Newtonian gravitational field about a spherical primary gravitational body. The precise integration of these equations is important for designing future spacecraft formation flying missions.

Example 3 (*Clohessy–Wiltshire equations*)
Integrate the following normalized CW equations:

$$\begin{aligned} \ddot{x} - 2\dot{y} - 3x &= \sin(t), \quad \dot{x}(0) = 0.1, \quad x(0) = 0.75 \\ \ddot{y} + 2\dot{x} &= \cos(t), \quad \dot{y}(0) = 0.1, \quad y(0) = 0.5 \\ \ddot{z} + z &= \tfrac{1}{2}\sin(t), \quad \dot{z}(0) = 0, \quad z(0) = 0 \end{aligned} \tag{97}$$

Since the highest-order harmonics of the forcing input is one, we choose $N = 1$, so the gauge functions are found via Equation (57):

$$\begin{aligned} \phi_1 &= \frac{A_{0_1}}{2} + A_{1_1}\cos(t) + B_{1_1}\sin(t) \\ \phi_2 &= \frac{A_{0_2}}{2} + A_{1_2}\cos(t) + B_{1_2}\sin(t) \\ \phi_3 &= \frac{A_{0_3}}{2} + A_{1_3}\cos(t) + B_{1_3}\sin(t) \end{aligned} \tag{98}$$

Utilizing the MATLAB® code we obtain the following cost function vector:

$$\begin{aligned} J_1 &= [-6\cos(t) - 30\sin(t)A_{2_1} + 30\cos(t)B_{2_1} + 16\cos(t)A_{1_1} + 16\sin(t)B_{1_1} + A_{0_1}/2]^2 \\ J_2 &= [30\sin(t)A_{1_1} - 30\cos(t)B_{1_1} + 49\sin(t)B_{2_1} + 49\cos(t)A_{2_1} - 6\sin(t) + 2A_{0_2}]^2 \\ J_3 &= [16\cos(t)A_{3_1} + 16\sin(t)B_{3_1} + A_{0_3}/2 + 4\cos(t)]^2 \end{aligned} \tag{99}$$

and construct the scalar cost function by taking

$$J = J_1 + J_2 + J_3 \tag{100}$$

Then, by evaluating the gauge functions coefficient we obtain the optimal gauge functions

$$\begin{aligned} \phi_1^\star &= -\tfrac{57}{58}\cos(t) \\ \phi_2^\star &= \tfrac{21}{29}\sin(t) \\ \phi_3^\star &= -\tfrac{1}{4}\cos(t) \end{aligned} \tag{101}$$

After the optimal gauge functions have been found, we integrate the system using the gauge-optimized VOP method. For comparison, we will also integrate the system by using standard state variables. In both cases we will use a fixed-step RK45 routine with the same time-step, $\Delta T = 0.05$. The integration errors of both methods are plotted in Figure 5.

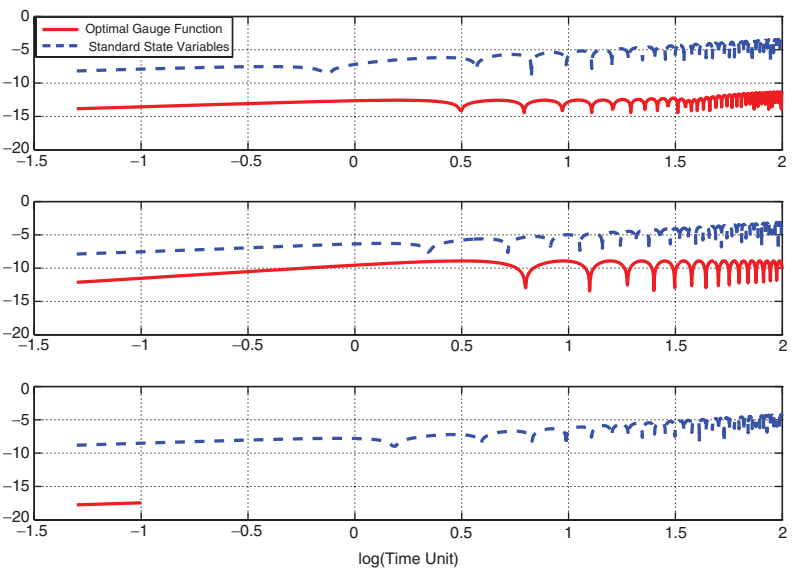

Figure 5. The gauge-optimized integration reduces the integration error of the Clohessy–Wiltshire equations by seven orders of magnitude compared to integration using standard state variables.

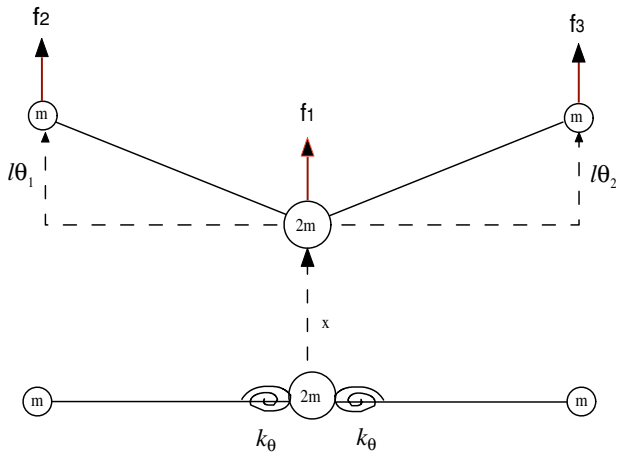

Figure 6. A 3-DOF lumped-mass model of an elastic aircraft.

In the x and y error curves there is a difference of seven magnitudes between the methods. Moreover, the z error curve stops before the time interval is completed, because the error is completely nullified, i.e. the analytic solution is exactly equal to the solution obtained by the gauge-optimized integration.

Our second example is taken from the field of aeroelasticity. Consider a 3-DOF lumped-mass model of an aircraft, where the fuselage is model as the point mass $2m$ and the elastic wings of length l are modelled by torsion springs with spring constant k_θ and tip masses m for each wing, as shown in Figure 6.

Assuming that a periodic force vector $[f_1(t), f_2(t), f_3(t)]^{\mathrm{T}}$ is exerted on the fuselage and both wing tips, respectively, the equations of motion become

$$m\begin{bmatrix} 4 & l & l \\ l & l^2 & 0 \\ l & 0 & l^2 \end{bmatrix}\begin{pmatrix} \ddot{x} \\ \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{pmatrix} + k_\theta \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{pmatrix} x \\ \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{pmatrix} \tag{102}$$

We are now ready to study a quantitative example.

Example 4 (*Elastic wing*)
Integrate the following system:

$$\begin{aligned} &\ddot{x} - \tfrac{1}{2}\theta_1 - \tfrac{1}{2}\theta_2 = 0, \quad \dot{x}(0) = 0.2, \quad x(0) = 0.4 \\ &\ddot{\theta}_1 + \tfrac{3}{2}\theta_1 + \tfrac{1}{2}\theta_2 = \sin(t) + \cos(t), \quad \dot{y}(0) = 0.3, \quad y(0) = 0.0 \\ &\ddot{\theta}_2 + \tfrac{1}{2}\theta_1 + \tfrac{3}{2}\theta_2 = \sin(2t), \quad \dot{z}(0) = 1, \quad z(0) = 0.0 \end{aligned} \tag{103}$$

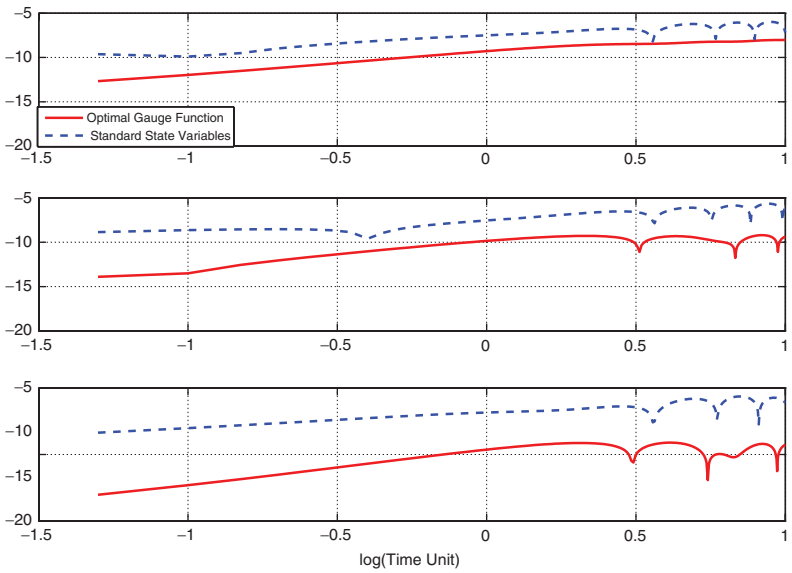

Figure 7. The gauge-optimized integration reduces the integration error of an elastic wing model by three orders of magnitude.

Utilizing the MATLAB® code we obtain the optimal gauge functions

$$\begin{aligned}\phi_1^\star &= -\tfrac{14}{145}\cos(t) + \tfrac{14}{145}\sin(t) - \tfrac{11}{410}\cos(2t)\\ \phi_2^\star &= -\tfrac{53}{116}\cos(t) + \tfrac{53}{116}\sin(t) + \tfrac{194}{4961}\cos(2t)\\ \phi_3^\star &= \tfrac{5}{116}\cos(t) - \tfrac{5}{116}\sin(t) - \tfrac{1546}{4961}\cos(2t)\end{aligned} \tag{104}$$

Figure 7 shows the comparison between an integration using standard state variables and the gauge-optimized integration. As can be seen, the gauge-optimized integration outperforms the standard method by three orders of magnitude in all state variables.

7. CONCLUSIONS

We have successfully developed a method for mitigating the numerical truncation error produced while integrating a system of linear time-invariant second-order Newtonian ordinary differential equations (ODEs).

This achievement has been based upon the observation that the standard, straightforward state-space variables used to model the phase-space dynamics are not necessarily the best state variables to be used for numerically integrating the underlying ODES. In fact, in all of the examples explored in this paper, numerically integrating the standard state variables gave a poor integration performance.

Instead, we proposed a methodology for constructing an optimal state-space representation that gave minimal numerical integration error, and in this sense, it is the optimal state-space representation for modelling the phase-space dynamics. Thus, the main contribution of this work is not in the development of a new integration routine, but rather development of a structured methodology for constructing the state-space model yielding a minimum numerical truncation error of a given integration routine. One could employ the developed formalism on any existing integrator, not necessarily the standard Runge–Kutta algorithm, used in this work as benchmark.

This process entails a transformation of the state-space equations into their variational form while taking into account the known homogenous solution, and using the inherent freedom of this transformation as a tuning mechanism for mitigating the integration error.

Thus, we have shown that there is a direct connection between the numerical integration error and the gauge function arising from the variation of parameters method. By choosing an appropriate gauge function the numerical integration error dramatically decreases and one can achieve much better accuracy compared to the standard state variables solution using a given time-step.

Moreover, by utilizing the new methodology we derived general expressions yielding the optimal gauge functions for a given Newtonian one degree-of-freedom ODE. For the n degrees-of-freedom system we composed a MATLAB® code capable of finding the optimal gauge functions and integrating the given system using the gauge-optimized integration algorithm. In all of our illustrating examples, the gauge-based integration outperformed the integration using standard state variables by at least three orders of magnitude.

In our future research, we show how to extend the gauge-optimized integration to nonlinear non-autonomous systems whose homogenous solution is known.

APPENDIX A: INVARIANCE OF x TO A SELECTION OF Φ

Substituting the solutions of the homogenous equation (11) into Equation (20) gives

$$\dot{C}_1(t) = \frac{\Phi(t)(\mathrm{e}^{(-\xi-\sqrt{\xi^2-1})\omega_n t})' - \mathrm{e}^{(-\xi-\sqrt{\xi^2-1})\omega_n t}[F(t) - \dot{\Phi}(t) - 2\xi\omega_n\Phi(t)]}{w[x_1(t), x_2(t)]} \tag{A1a}$$

$$\dot{C}_2(t) = \frac{\mathrm{e}^{(-\xi+\sqrt{\xi^2-1})\omega_n t}[F(t) - \dot{\Phi}(t) - 2\xi\omega_n\Phi(t)] - \Phi(t)(\mathrm{e}^{(-\xi+\sqrt{\xi^2-1})\omega_n t})'}{w[x_1(t), x_2(t)]} \tag{A1b}$$

Denoting

$$\Delta = \frac{F(t)}{w[x_1(t), x_2(t)]} \tag{A2}$$

and rearranging Equation (A1) entails

$$\dot{C}_1(t) = -\mathrm{e}^{(-\xi-\sqrt{\xi^2-1})\omega_n t}\Delta + \frac{(\Phi(t)\mathrm{e}^{(-\xi-\sqrt{\xi^2-1})\omega_n t})'}{w[x_1(t), x_2(t)]} + \frac{2\xi\omega_n\Phi(t)\mathrm{e}^{(-\xi-\sqrt{\xi^2-1})\omega_n t}}{w[x_1(t), x_2(t)]} \tag{A3a}$$

$$\dot{C}_2(t) = \mathrm{e}^{(-\xi+\sqrt{\xi^2-1})\omega_n t}\Delta - \frac{(\Phi(t)\mathrm{e}^{(-\xi+\sqrt{\xi^2-1})\omega_n t})'}{w[x_1(t), x_2(t)]} - \frac{2\xi\omega_n\Phi(t)\mathrm{e}^{(-\xi+\sqrt{\xi^2-1})\omega_n t}}{w[x_1(t), x_2(t)]} \tag{A3b}$$

$$\begin{aligned}
\dot{C}_1(t) &= \frac{1}{w[x_1(t), x_2(t)]}\left[\frac{\mathrm{d}}{\mathrm{d}t}(x_2(t)\Phi(t)) + x_2(t)\omega_n^2\Phi(t)\right] - \frac{x_2(t)F(t)}{w[x_1(t), x_2(t)]} \\
\dot{C}_2(t) &= -\frac{1}{w[x_1(t), x_2(t)]}\left[\frac{\mathrm{d}}{\mathrm{d}t}(x_1(t)\Phi(t)) + x_1(t)\omega_n^2\Phi(t)\right] + \frac{x_1(t)F(t)}{w[x_1(t), x_2(t)]}
\end{aligned} \tag{A4}$$

Integration of Equation (A4) yields

$$\begin{aligned}
C_1(t) &= -\int^t \frac{x_2(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\mathrm{d}\tau + \int^t \frac{1}{w[x_1(\tau), x_2(\tau)]}\frac{\mathrm{d}}{\mathrm{d}\tau}(x_2(\tau)\Phi(\tau))(\tau)\,\mathrm{d}\tau \\
&\quad + \omega_n^2\int^t \frac{x_2(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau + c_3 \\
C_2(t) &= +\int^t \frac{x_1(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau - \int^t \frac{1}{w[x_1(\tau), x_2(\tau)]}\frac{\mathrm{d}}{\mathrm{d}\tau}(x_1(\tau)\Phi(\tau))(\tau)\,\mathrm{d}\tau \\
&\quad - \omega_n^2\int^t \frac{x_1(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau + c_4
\end{aligned} \tag{A5}$$

Cancelling the $\mathrm{d}\tau$ terms in middle expression entails

$$C_1(t) = -\int^t \frac{x_2(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau + \frac{x_2(\tau)\Phi(\tau)(\tau)}{w[x_1(\tau), x_2(\tau)]} + \omega_n^2\int^t \frac{x_2(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau + c_3 \tag{A6a}$$

$$C_2(t) = +\int^t \frac{x_1(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau - \frac{x_1(\tau)\Phi(\tau)(\tau)}{w[x_1(\tau), x_2(\tau)]} - \omega_n^2\int^t \frac{x_1(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau + c_4 \tag{A6b}$$

Substitution of Equation (A6) into the general solution given by Equation (12) and rearranging results in

$$\begin{aligned}
x(t) &= -x_1(t)\int^t \frac{x_2(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\mathrm{d}\tau + x_2(t)\int^t \frac{x_1(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\mathrm{d}\tau \\
&\quad + x_1(t)\omega_n^2\int^t \frac{x_2(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\,\mathrm{d}\tau - x_2(t)\omega_n^2\int^t \frac{x_1(\tau)\Phi(\tau)}{w[x_1(\tau), x_2(\tau)]}\mathrm{d}\tau + x_1c_3(t) + x_2c_4(t)
\end{aligned} \tag{A7}$$

Since

$$x_1(t)\omega_n^2\int^t \frac{x_2(\tau)\Phi(\tau)}{W(x_1, x_2)(\tau)}\mathrm{d}\tau - x_2(t)\omega_n^2\int^t \frac{x_1(\tau)\Phi(\tau)}{W(x_1, x_2)(\tau)}\mathrm{d}\tau = 0 \quad \forall t \tag{A8}$$

the general solution of $x(t)$ is given by

$$x(t) = -x_1(t)\int^t \frac{x_2(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)]}\mathrm{d}\tau + x_2(t)\int^t \frac{x_1(\tau)F(\tau)}{w[x_1(\tau), x_2(\tau)])}\mathrm{d}\tau + x_1c_3(t) + x_2c_4(t) \quad \text{(A9)}$$

Equation (A9) shows that the general solution has no dependence on Φ.

ACKNOWLEDGEMENTS

This work was supported by the Asher Space Research Institute at the Technion, Israel Institute of Technology. The authors would like to express their gratitude to Dr Michael Efroimsky from USNO for his inspirational research and to Dr Daniella Raveh from the Technion for providing the Elastic Wing Example.

REFERENCES

1. Newton I. Philosophiae naturalis principia mathematica. *Samuel Pepys*, 1687. In *The Principia*: *Mathematical Principles of Natural Philosophy*, Newton I, Bernard Cohen I (Translator), Anne Whitman (Translator) (later edition). University of California Press: Berkeley, CA, 1999.
2. Grant H. Leibniz—beyond the calculus. *Math. Semesterber* 1992; **39**:3–11.
3. Gottfried W. Leibniz. *Acta Eruditorum* 1686.
4. Ball WW. *Short Account of the History of Mathematics* (4th edn). Dover Publications: New York, 1960.
5. Runge C. Üeber die numerische auflosung von differentialgleichungen. *Mathematische Annalen* 1895; **46**:167–178.
6. Kutta W. Beitrag zur näherungsweisen integration totaler differentialgleichungen. *Zeitschr. fur Math. u. Phys.* 1901; **46**:435–453.
7. Nyström EJ. Über die numerisehe integration yon differentialgleiehungen. *Acta Soc. Sci. Fennica* 1925; **50**:5.
8. Huťa A. Une amëtlioration de ia méthode de Runge–Kutta–Nyström pour la résolution nurnérique des équations différentielles du premier ordre. *Acta Mathematica Universitatis Comenianae* 1965; **1**:21–24.
9. Butcher JC. On Runge–Kutta methods of high order. *Journal of the Australian Mathematical Society* 1964; **4**:79–194.
10. Curtis AR. An eighth order Runge–Kutta process with eleven function evaluations per step. *Numerische Mathematik* 1970; **16**:268–277.
11. Curtis AR. High order explicit Runge–Kutta formulae their uses and limitations. *Journal of the Institute of Mathematics and its Applications* 1975; **16**:35–55.
12. Cooper GJ, Verner JH. Some explicit Runge–Kutta methods of high order. *SIAM Journal on Numerical Analysis* 1972; **9**:389–405.
13. Hairer E. A Runge–Kutta method of order 10. *Journal of the Institute of Mathematics and its Applications* 1978; **21**:47–59.
14. Euler L. Recherches sur la question des inegalites du mouvement de Saturne et de Jupiter, sujet propose pour le prix de l'annee. *Piece qui a remporte le prix de l'academie royale des sciences*, 1748. For modern edition see: Euler L. Opera mechanica et astronomica, Birkhauser-Verlag: Switzerland, 1999.
15. Euler L. Theoria motus Lunae exhibens omnes ejus inaequalitates etc. *Impensis Academiae Imperialis Scientarum Petropolitanae*. St. Petersburg: Russia, 1753. Euler L. *Opera mechanica et astronomica*. (modern edition). Birkhauser-Verlag: Switzerland, 1999.
16. Lagrange JL. *Sur le Problème de la détermination des orbites des comètes d'après trois observations, 1-er et 2-ième mémoires*. Nouveaux Mémoires de l'Académie de Berlin, 1778. In *Œuvres de Lagrange*, vol. IV (later edition). Gauthier-Villars: Paris, 1869.
17. Lagrange JL. *Sur la théorie des variations des éléments des planètes et en particulier des variations des grands axes de leurs orbites*. Lu, le 22 août 1808 à l'Institut de France, 1808. In *Œuvres de Lagrange*, vol. VI (later edition). Gauthier-Villars: Paris, 1877; 713–768.
18. Lagrange JL. *Second mémoire sur la théorie générale de la variation des constantes arbitraires dans tous les problèmes de la mécanique*. Lu, le 19 février 1810 à l'Institut de France, 1810. In *Œuvres de Lagrange*, vol. VI (later edition). Gauthier-Villars: Paris, 1877; 809–816.

19. Efroimsky M. Implicit gauge symmetry emerging in the n-body problem of celestial mechanics. Astro-ph/0212245, 2002.
20. Efroimsky M, Goldreich P. Gauge symmetry of the n-body problem in the Hamilton–Jacobi approach. *Journal of Mathematical Physics* 2003; **44**:5958–5977.
21. Efroimsky M, Goldreich P. Gauge symmetry of the n-body problem of celestial mechanics. *Astronomy and Astrophysics* 2004; **415**:1187–1199.
22. Gurfil P. Analysis of j_2-perturbed motion using mean non-osculating orbital elements. *Celestial Mechanics and Dynamical Astronomy* 2004; **90**:289–306.
23. Boyce WE, Diprima RC. *Elementary Differential Equations* (6th edn). Wiley: New York, 1996.
24. Gerald CF, Wheatley PO. *Applied Numerical Analysis* (5th edn). Addison-Wesley Publishing Company: Reading, MA, 1994.
25. Scarborough JB. Formulas for the error in Simpson's rule. *The American Mathematical Monthly* 1926; **33**(2):76–83.
26. Ogata K. *Modern Control Engineering* (4th edn). Pearson Education International: New York, 2002.
27. Cannon RH. *Dynamics of Physical Systems*. McGraw-Hill Book Company: New York, 1967.
28. Bisplinghoff RL, Ashley H. *Principles of Aeroelasticity*. Dover Publications, Inc.: New York, 1962.
29. Newman WI, Efroimsky M. The method of variation of parameters and multiple time scales in orbital mechanics. *Chaos* 2003; **13**:476–485.
30. Fourier J. *Theorie analytique de la chaleur*. Chez Firmin Didot Pre et Fils, 1822.
31. Churchill RV. *Fourier Series and Boundary Value Problems*. McGraw-Hill: Kogakusha, 1963.
32. Horn RA, Johnson CR. *Matrix Analysis*. Cambridge University Press: Cambridge, MA, 1985.
33. Clohessy WH, Wiltshire RS. Terminal guidance system for satellite rendezvous. *Journal of the Aerospace Sciences* 1960; **27**:653–658.

