

The Method of Variation of Constants and Multiple Time Scales in Orbital Mechanics

William I. Newman*

*Departments of Earth & Space Sciences, Physics & Astronomy,
and Mathematics, University of California, Los Angeles, CA 90095, USA*

Michael Efroimsky†

*Institute for Mathematics and its Applications, University of Minnesota,
207 Church St. SE, Suite 400, Minneapolis, MN 55455, USA*

(Dated: November 10, 2002)

The method of variation of constants is an important tool used to solve systems of ordinary differential equations, and was invented by Euler and Lagrange to solve a problem in orbital mechanics. This methodology assumes that certain “constants” associated with a homogeneous problem will vary in time in response to an external force. It also introduces one or more constraint equations motivated by the nature of the time-dependent driver. We show that these constraints can be generalized, in analogy to gauge theories in physics, and that different constraints can offer conceptual advances and methodological benefits to the solution of the underlying problem. Examples are given from linear ordinary differential equation theory and from orbital mechanics. However, a slow driving force in the presence of multiple time scales contained in the underlying (homogeneous) problem nevertheless requires special care, and this has strong implications to the analytic and numerical solutions of problems ranging from celestial mechanics to molecular dynamics.

PACS numbers: 02.30.Hq, 02.60.Cb, 45.10.-b, 45.50.Pk, 05.45.-a

Keywords: Ordinary differential equations, Numerical simulation and solution of equations, computational methods in classical mechanics, celestial mechanics and classical mechanics of discrete systems, nonlinear dynamics; Multiple time scales.

Running head: Multiple Time Scales in Variation of Constants Method

I. INTRODUCTION

The method of variation of constants provides a means for converting homogeneous solutions for a set of linear ordinary differential equations into a particular solution of an inhomogeneous or driven system. The method was originally conceived by Euler and Lagrange, and is derived assuming that a specified, often seemingly-intuitive constraint is satisfied by the “constants” of the homogeneous problem, which are now allowed to vary with time. Moreover, it is widely assumed that a slowly varying driving force will result in the so-called constants also varying slowly in time. Euler and Lagrange exploited this method for the highly nonlinear N -body problem in celestial mechanics. Were it not for the interaction between the planets—in contrast with their interaction with the sun—the classical action variables (in an action-angle formulation) would be exactly preserved and became the “constants” or parameters to which Euler and Lagrange applied this methodology. For the nonlinear problem, they then constrained the way the constants would vary in analogy to the linear situation. We show that it is possible to generalize the constraints applied, in a manner equivalent to dif-

ferent “gauges” employed in physics, such as in electromagnetic theory, and that it is possible to simplify the equations of motion by utilizing an appropriately selected gauge. However, we also show that the presence of multiple time scales in the original problem is preserved in the variation of constants formulation, and that these will necessarily regulate the computation of numerical solutions for the problem. This, in turn, has profound implications to the solution of systems of ordinary differential equations in many physical environments, ranging from celestial mechanics to molecular dynamics and other disciplines.

The method of variation of constants remains a powerful tool for solving differential equations, highly popular among physicists, mathematicians, astronomers and engineers. The application of this method to linear ordinary differential equations is presented in many classical textbooks^{1,2}, while its applications to nonlinear ordinary differential equations were studied by Alekseev³. Later Brunner⁴ used it to solve integro-differential equations, while Deo and Torres⁵ employed it for solving nonlinear functional differential equations. At present, the book by Lakshmikantham and Deo⁶ provides the most comprehensive reference to applications of the method to the theory of differential equations. No single source covers the vast and still growing number of applications of the

method to problems in many areas of science. Numerous versions of this approach have appeared in the literature, but the underlying idea remains the same. Typically, the user begins from known solutions for a homogeneous linear or nonlinear set of ordinary differential equations, and tries to find a solution to the appropriate inhomogeneous problem—exploiting the homogeneous solutions and associated constants—by permitting these constants to vary according to some specified time or coordinate dependence.

Although the method is particularly simple for linear problems, it first appeared in the highly nonlinear context of celestial mechanics. It was invented, independently, by Euler and Lagrange at the end of eighteenth century as a means of going beyond the two-body problem and obtaining an approximate solution to the N -body one. The basic idea of Euler and Lagrange was to take the solution to two-body problem as a convenient starting point for the N -body problem, and to convert what were constants of the motion or parameters in the homogeneous problem into functions of time—hence the name “variation of constants” or “variation of parameters.” Euler⁷ used this approach between 1748 and 1752 to study the mutual disturbances of Jupiter and Saturn, while Lagrange⁸ developed it to describe a comet’s motion, taking into account planet-caused perturbations.

The underlying idea in Euler’s and Lagrange’s treatment was to express the solution to a problem without disturbances, and then to introduce the disturbances, and identify how they would alter the initial, unperturbed solution. For example, Lagrange’s treatise considered an elliptic (or hyperbolic) trajectory of a comet around the Sun. In the case of an elliptical orbit, the trajectory is given by a Keplerian ellipse which is completely specified by three Euler angles, the eccentricity, the semi-major axis, and the position of the comet on this ellipse at the initial time. These six quantities—the so-called “orbital elements” or simply “elements”—are constants in the two-body problem, where the comet and the Sun are not influenced by the gravitational attraction of the planets. Lagrange realized that the gravitational interaction with the planets could alter the cometary path, but he assumed that at each instant of time the comet still follows some instantaneous ellipse. Thus, instead of a static Keplerian ellipse, he produced a family of ellipses that were instantaneously coincident with the orbit. Mathematically, he assumed that the six orbital elements become functions of time. (Later Lagrange’s method was generalized for the case of disturbances caused by non-sphericity of the primary body⁹.) In § IV, we shall discuss Lagrange’s method in greater detail.

As an illustration of how nonlinear interactions cause what would otherwise be regarded as “constants of the motion” to vary, we illustrate below the behavior observed in the eccentricity in Jupiter’s satellite Io whose mean eccentricity is approximately 0.004, roughly half that of Earth. Its orbit is strongly influenced by resonances with its neighboring large Galilean satellites Eu-

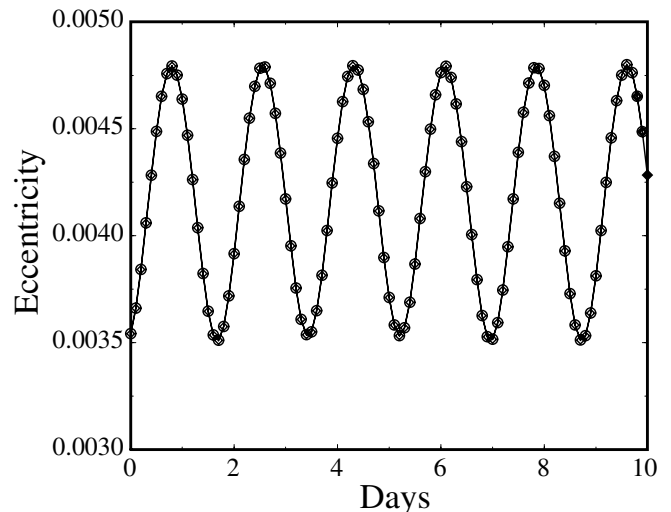


FIG. 1: Io’s eccentricity variation. Jupiter’s oblateness and the neighboring satellites cause variations of about a quarter of Io’s mean eccentricity on the orbital time scale. Furthermore, comparison of Io’s computed orbital eccentricity (circles) with the JPL 166 ephemeris data (diamonds) shows that the ephemeris is reproduced with high accuracy. Adapted from¹⁰.

ropa, Ganymede, and Callisto and, especially, by torques caused by Jupiter’s oblateness. Jupiter is highly oblate—its equatorial radius is approximately 1% larger than its polar radius due to its rapid rotation—and this establishes a quadrupole interaction with Io which, due to Io’s proximity to the planet, is strongly affected by the equatorial bulge. As shown in Fig. 1, the variations in Io’s orbital eccentricity are about 10% of its mean value. The interaction between Io and Jupiter’s equatorial bulge, as well as between Io and its Jovian satellite neighbors, establishes the fast time scales, ranging from Jupiter’s 10-hour rotation period to the periods of revolution of the satellites closest to Io, periods that vary from days to weeks. On the other hand, we have come to expect that successive interactions between Io and its Galilean satellite neighbors would slowly extract tiny amounts of energy and angular momentum from Io with each conjunction. Observed over a long period of time, this seemingly has the effect of an external force acting over very long time scales. Thus, we believe that Io’s orbital elements will change very slowly over time. Contrary to these expectations, Io’s eccentricity is observed to vary on a time scale commensurate with Io’s orbital motion around Jupiter, what one would normally want to regard as a “fast,” and thereby ignorable timescale. Io’s trajectory eloquently shows why it is wrong to assume that the evolution of the orbital elements—the “constants” in the unperturbed problem—will not be influenced by fast time scales contained within the problem.

Many problems in nature blend a variety of time scales, and this is especially evident in celestial mechanics. The undisturbed problem is often characterized by a wide

range of time scales, where the disturbing force is seemingly slowly-varying. It is tempting to think that the slow variation in the perturbation will be mirrored in a slow variation of the unperturbed “constants.” Such an assumption—which is intrinsic to the use of the method of variation of constants by many investigators—is manifestly wrong as is clear from Io’s orbital behavior. A ramification to this observation is that numerical methods employed to simulate the evolution of such constants, thereby capturing the essence of Lagrange’s instantaneous ellipses, will be vulnerable to the fast time scales implicit to the wider problem. Thus, the expectation that we can pursue simulations of the equations governing the orbital elements using much longer time steps is inherently incorrect—the constants do vary, and they vary rapidly.

The goal of this paper is to explore the nature of physical problems whose unperturbed state contains fast time scales and then experiences a slowly-varying perturbation, the typical setting for the method of variation of constants. This is a generic and systemic problem, and is encountered in many disciplines, although we shall focus on a linear Hamiltonian paradigm and on a nonlinear celestial mechanical one. We shall then turn our attention to implications of multiple time scales upon numerical methods—such as orbit integrations that utilize Lagrange’s orbital element description—that employ long time steps motivated by the slowly varying nature of perturbing influences.

In §II, we shall explore some contemporary applications of the method of constants variation, focusing on the issue of multiple time scales and the emergent role of computation. In §III, we shall provide a simple mathematical paradigm based on a (slowly) driven simple harmonic oscillator. Then, §IV will show how this paradigm relates to the problem of orbital motion. This analogy underscores our point that, no matter how slowly the driving force evolves, the changes to the “constants” may be fast. In §V, we shall explore the generic influence of multiple time scales, and shall explain why it cannot be avoided by analytical means such as canonical transformations.

In addition to the problem of multiple time scales and their role in determining the integration step size, we shall address another interesting and overlooked aspect of the method of variation of constants. In both linear and nonlinear problems, when one allows the constants to vary, the problem of establishing their evolution is underdetermined. This necessitates the introduction of additional conditions or constraints “by hand,” in order to make the problem well defined.

The freedom of choice in selecting these constraints is reminiscent of the gauge freedom encountered in electrodynamic theory. This freedom will be explored in §III and §IV. If employed appropriately, this freedom does not influence the final physical solution, but may considerably reduce the complexity encountered in calculating the solution. How to achieve this objective, however, re-

mains a topic of ongoing research.

II. USE AND ABUSE OF THE METHOD OF VARIATION OF CONSTANTS

A common feature of the method of variation of constants is the presence of multiple time scales. The latter embraces situations where it is necessary to deal with frequencies that differ substantially—sometimes by orders of magnitude—from one another. Suppose, for example, that a homogeneous system of (not necessarily linear) ordinary differential equations manifests behavior over some characteristic range of frequencies, while an inhomogeneous source drives the system at some much lower frequency. Numerically, these situations are sometimes called “oscillatorily stiff” since the computational step size must be selected to conform with the highest frequencies in the problem. Thus, even in situations where the high frequency terms are physically insignificant, they control the integration step size. If we were to ignore them, we would encounter nonphysical aliasing effects and, very possibly, artificial computational instabilities¹¹.

A physically transparent example emerges in celestial mechanics where one starts with a homogeneous vector equation that describes the decoupled two-body interactions of each of the planets with the Sun, and then includes, as inhomogeneous terms, the much smaller interactions between the planets. The set of homogeneous two-body problems have as their solution Keplerian ellipses. In order to address the full N -body problem, however, it is useful to represent the orbits of each of the planets by ellipses that undergo time evolution. The motion of each planet along an instantaneous or “osculating” ellipse is called “fast” motion, while the evolution of this ellipse is called “slow motion.” Importantly, as we have seen from Io’s motion, the evolution of the ellipse, through its eccentricity among other orbital elements, does manifest fast time behavior. We have become accustomed to expecting that the presumed slowness of the ellipses’ evolution may be ascribed to weak disturbances, an assumption which is manifestly wrong.

This example illustrates a problem that occurs in many different kinds of applications, where a slow external driver is expected to result only in slow evolution of the system. It is, however, a conceptual trap to expect that the “constants” will present only the low frequency behavior present in the driver, with no discernible affect from the high frequency modes contained within the homogeneous problem. This trap can become especially perilous when the variation of constants must be performed by computational means. Thus, numerical solutions may be invalidated by integration schemes that rely too heavily on a flawed intuition and long integration steps. Here, then there is a typical example¹²: “The method is based essentially on varying the arbitrary constants entering the unperturbed solution and on con-

structing, for the slowly varying functions of the coordinate and time thus created, a system of differential equations....” The authors consider a one-dimensional nonlinear dynamical system perturbed by a force represented as an infinite series over positive powers of a small parameter. They build an asymptotic method based on a known solution for the unperturbed problem, by varying the parameters entering the unperturbed solution, and by providing an approximate solution to the perturbed problem with accuracy up to the terms of a certain power of the parameter. These authors *ab initio* believed that a slowly varying disturbance would result only in a slowly varying evolution of the constants. This line of argument is harmless when all calculations are performed analytically, as in¹², but becomes perilous when numerical integration is required to compute the evolution of the varying constants. The choice of a coarse integration step, based on the slow pace of the disturbance, may result in very considerable and undetected errors, or in the mistaken belief that the system is physically unstable. It is essential that one adheres to the fine step size established by the initial, homogeneous problem, since both homogeneous and inhomogeneous problems have the same spectral content. In some works^{13–15}, this counterintuitive circumstance is taken into account and suitably short time steps are employed. Other authors however continue to claim that the method of variation of constants facilitates the use of a coarse integration step.

Maezawa and Miyagawa¹⁶ consider an underdamped driven harmonic oscillator. The solution with zero driving force is chosen as the homogeneous one, and its constants (the amplitude and the initial phase) are used as variable parameters when the external driver is activated. (This methodology can be employed in principle not only in the case of a time-dependent driving force but also for a force that depends upon the coordinates and velocity.) The authors state that computation based on the parameter-variation approach allows for a larger step size than a direct integration. Interestingly, however, they employed as their “coarse step” time steps much smaller than the periods of the homogeneous oscillators—contrary to their claims—obtaining relative errors of 0.5% and 2.0%.

Dallas and Rinderle¹⁷, in their discussion of numerical-integration methods for celestial mechanics, insist that the integration step size in the method of variation of constants may be selected to be larger than that employed in direct integration schemes. In the same spirit as¹², these authors make a strong opening statement: “The principal advantage of the variation-of-parameters method is that the derivatives are small and change slowly with time....” Based on this argument, the authors insist that larger tabular intervals are allowed when the orbit is parametrized by its instantaneous Kepler elements. Their numerical examples—prepared for a Mars-orbiting spacecraft—conform with their statement. However, examination of their Table 2 reveals that their “coarse” step remains two orders of magnitude shorter than the period

of revolution. Fortunately, they did not abide by their own dictum in performing the orbital calculation.

Tani and Inokuti¹⁸, who developed a variation-of-constants-based method for integration of the Schrödinger equation, state that “the new method achieves high accuracy even when the step size is appreciably large.” Their formulation is a variant of a phase variable approach, and their variation of parameters methodology introduces a Riccati-like equation having a tangent function-like behavior. Thus, while the power-series they introduce has several vanishing derivative terms, the higher order terms will in general have a small radius of convergence, and small step sizes will in general be necessary. See¹⁹ for an alternative approach to the numerical integration of the Schrödinger equation—including error bounds—that exploits the smoothness of the phase-like variable to overcome in many practical situations the small step size problem.

After performing an exhaustive search of the mathematics and physics literature, we have found very few references to the method of variation of constants after the mid 1970’s, while there were literally hundreds prior to that time. We suspect that the advent of the computer brought an effective end to this avenue of analytic investigations. Researchers have largely become content to develop numerical solutions to these kinds of questions, and have all but abandoned analytical methods. Importantly, in the present era, we are beginning to confront problems which challenge our ability to compute accurate solutions to differential equations over very long times, e.g., in celestial mechanics as applied to the long-term dynamics of our solar system. Accordingly, the use of hybrid approaches which blend analytic methods such as the variation of constants with numerical integration will become increasingly important as contemporary problems challenge existing limits to computation. Now, we turn our attention to the simple harmonic oscillator with an external driver as a linear paradigm whose properties can be investigated analytically.

III. DRIVEN SIMPLE HARMONIC OSCILLATOR

We wish to understand why the “fast” variation generally cannot be ignored or eliminated by some averaging procedure over a long interval of time. Let us consider the simple example of a driven harmonic oscillator described by the Hamiltonian

$$H(x, p) = \frac{p^2}{2} + \frac{x^2}{2} + xF(t)$$

which yields the equation of motion

$$\ddot{x} + x = F(t), \quad (1)$$

for specified initial data $x(0)$ and $\dot{x}(0)$. At first glance, this toy model may seem too far removed from real-life

problems, such as those of celestial mechanics. However, we shall show in the following section that this simple paradigm is closely analogous to nonlinear problems encountered in orbital mechanics.

Equation (1) conforms with the standard development of the variation of constants¹. We begin with the solution to the homogeneous equation(s), and then perform the variation of constants, as in

$$x(t) = S(t) \sin(t) + C(t) \cos(t) \quad . \quad (2)$$

It follows that

$$\begin{aligned} \dot{x}(t) = & \left\{ \dot{S}(t) \sin(t) + \dot{C}(t) \cos(t) \right\} \\ & + S(t) \cos(t) - C(t) \sin(t) \quad . \end{aligned} \quad (3)$$

Taking one further time derivative, we obtain a second order ordinary differential equation in *both* $S(t)$ and $C(t)$. The problem is, therefore, underdetermined. Most textbooks, for example¹, eliminate this indeterminacy by introducing a constraint such as $\dot{S}(t) \sin(t) + \dot{C}(t) \cos(t) = 0$. This equation of constraint is motivated conceptually by the notion that the time variability in the “constants” S and C does not contribute substantially to the velocity $\dot{x}(t)$ in the underlying dynamical equation (1).

This specific constraint equation is, however, not required. If we introduce the following expression, where $\Phi(t)$ is to be determined,

$$\dot{S}(t) \sin(t) + \dot{C}(t) \cos(t) = \Phi(t), \quad (4)$$

then we obtain

$$\begin{aligned} \ddot{x}(t) = & \dot{\Phi}(t) + \dot{S}(t) \cos(t) - \dot{C}(t) \sin(t) \\ & - S(t) \sin(t) - C(t) \cos(t), \end{aligned} \quad (5)$$

and

$$\ddot{x}(t) + x(t) = \dot{\Phi}(t) + \dot{S}(t) \cos(t) - \dot{C}(t) \sin(t). \quad (6)$$

The solution that we are seeking should simultaneously satisfy the equation of motion—that follows from (6) and (1)—and the above identity (4), namely,

$$\begin{cases} \dot{\Phi}(t) + \dot{S}(t) \cos(t) - \dot{C}(t) \sin(t) = F(t) \\ \Phi(t) \equiv \dot{S}(t) \sin(t) + \dot{C}(t) \cos(t). \end{cases} \quad (7)$$

The solution becomes

$$\dot{S}(t) = F(t) \cos(t) - \frac{d}{dt} [\Phi(t) \cos(t)] \quad (8)$$

$$\dot{C}(t) = -F(t) \sin(t) + \frac{d}{dt} [\Phi(t) \sin(t)]$$

with the function Φ still being arbitrary. For initial datum specified by $x(0)$ and $\dot{x}(0)$, the system of equations

$C(0) = x(0)$ and $\Phi(0) + S(0) = \dot{x}(0)$ yields $S(0)$ and $C(0)$, for any choice of $\Phi(0)$. We explain in § V that the above solution for $\dot{S}(t)$ and $\dot{C}(t)$ is closely analogous to the Lagrange system of equations for osculating elements in orbital dynamics.

The integration of (8) leads to the first integrals

$$S(t) = \int^t F(t') \cos(t') dt' - \Phi(t) \cos(t) + c_1 \quad (9)$$

$$C(t) = - \int^t F(t') \sin(t') dt' + \Phi(t) \sin(t) + c_2$$

where c_1 and c_2 are constants of integration. When we insert these expressions into (1), we obtain a unique solution for the trajectory, namely

$$\begin{aligned} x(t) = & S(t) \sin(t) + C(t) \cos(t) = \\ & - \cos(t) \int^t F(t') \sin(t') dt' \\ & + \sin(t) \int^t F(t') \cos(t') dt' \\ & + c_1 \sin(t) + c_2 \cos(t) \end{aligned} \quad (10)$$

with the Φ -terms cancelling. The above two integrals can be combined, thereby producing the familiar Green function solution for (1).

In this example the choice of Φ is unimportant because the entire procedure can be carried out completely by analytical means. Were the equation more involved, one would have to numerically integrate (8), and then the Φ -terms will come into play.

When choosing a timestep for the numerical integration of (9), we have to accommodate the existence of two timescales, one defined by the unit frequency established in the homogeneous problem, another by the time dependence of the disturbance $F(t)$. In many practical problems—for example, in many problems of celestial mechanics—the time scale of the disturbance is slower than that associated with the homogeneous problem. While it would be tempting to construct an integration scheme that incorporates a coarse integration step, i.e., one associated with the slowly varying perturbation, we see in our example that this would be impossible, even if one selects the trivial constraint equation $\Phi(t) = 0$. Although it would make the integration of (9) less computationally intensive, the integration step size must be that associated with the fast time scale dictated by the homogeneous solutions $\sin(t)$ and $\cos(t)$. Another choice for $\Phi(t)$ could conceivably increase the time step in one of the equations (9). For example, if we choose $\Phi(t) = F(t)$, then the fast time scale will be completely removed from the second equation of (9) through an integration by parts, and the latter equation will have

the form $C(t) = \int^t \Phi(t') \cos(t') dt'$. However, the first equation of (9) contains the fast time scale and, thus, will require a small integration step. We see from this simple example that the presence of the arbitrary “gauge” $\Phi(t)$ -terms does not facilitate the segregation of the two time scale.

IV. LAGRANGE’S APPLICATION TO ORBITAL MECHANICS

We now wish to establish the generality of the ideas resident in the example above. How is the simple Hamiltonian (1) relevant to more complex, typically nonlinear, problems of science and engineering that become subject to the method of variation of constants? We propose to show that the above example is closely analogous to one of the oldest problems of mathematical physics, the calculation of planetary orbits. To make the analogy transparent, we shall recall the method employed by Lagrange to derive his famous system of equations for the evolution of the instantaneous ellipses along which the planets move.

Lagrange began from the two-body problem which is especially simple because of its mathematical equivalence to the gravitationally bound motion of a (reduced) point mass m_i around a fixed center of mass $m_{\text{planet}} + m_{\text{sun}}$ where $m_i = m_{\text{planet}} m_{\text{sun}} / (m_{\text{planet}} + m_{\text{sun}})$. The position of the planet, $\boldsymbol{\rho}$, relative to the sun obeys the equation

$$\ddot{\boldsymbol{\rho}} = -\frac{\mu}{\rho^2} \frac{\boldsymbol{\rho}}{\rho}, \quad \mu \equiv G(m_{\text{planet}} + m_{\text{sun}}). \quad (11)$$

This yields the celebrated Keplerian solution—the planet describes an ellipse that has one of its foci at the gravitating center of mass $m_{\text{planet}} + m_{\text{sun}}$. The position of the ellipse remains unchanged in time and can be characterized by the three Euler angles: the longitude of the node, Ω , describing how much the ellipse is originally rotated in the reference plane; the inclination, i^* of the orbital plane relative to the reference plane; and the argument of pericenter, ω , a final rotation in the reference plane. [Some authors use the longitude of pericenter, $\tilde{\omega} \equiv \Omega + \omega$ instead of the argument of pericenter ω .] The aforementioned reference plane constitutes the planet’s equatorial plane—when the motion of moons or artificial satellites is discussed—or the invariable plane of the Solar system—when the motion of planets is studied. The ellipse’s shape is completely determined by its semimajor axis, a , and the eccentricity e . The position of a planet on this orbit may be parametrized, for example, by the so-called mean anomaly M given by

$$M = M_0 + n(t - t_0), \quad (12)$$

where the integration constant M_0 represents the value of the mean-anomaly value at an “epoch,” i.e., time t_0 ; and n is the so-called mean motion, $n \equiv \mu^{1/2} a^{-3/2}$, i.e., the frequency of the orbit. (See, for example, Goldstein²⁰.)

The mean anomaly essentially describes the particle’s angular position on the ellipse as a function of time, thereby completing Lagrange’s specification of the reduced two-body system.

In planetary astronomy, the transition from the two-body problem to the N -body problem is traditionally achieved by assuming that the role of planet-planet interactions is weak. (Since the sun is *not* at the center of the reduced two-body problem, an additional correction known as the barycentric term is required.) The emergent perturbations are weak compared with the sun’s gravity, and the mutual interactions between the planets are assumed to be adiabatic, i.e., result only in very slow and gradual variations of the orbits. Accordingly, it is generally assumed that the planets’ trajectories reside on ellipses which are slowly evolving. This picture, as in any adiabatic setting, naturally evokes two different time scales. The interrelation between these scales will be the main concern of this section of our paper. However, before addressing this issue, we should comment on the validity of the preceding presumption of adiabaticity.

While it is true that the disturbances are weak compared to solar attraction, the smallness of the perturbation’s amplitude is irrelevant in the mathematical treatment of the problem—the Lagrange method of variation of constants is not predicated on the smallness of perturbations, just as we saw in the case of the driven simple harmonic oscillator. Another interesting question, which will only be mentioned here but which is developed at length in a separate paper (Efroimsky²¹), is whether the representation of real orbits by instantaneously determined ellipses is unique.

To address these topics mathematically, let us recall how Lagrange in 1808 justified his idea of slowly-evolving instantaneous ellipses. He began with the self-evident observation that a solution to the two-body problem (11) involves three second-order differential equations or, equivalently, six first-order ones. The solution to this system will depend either on both the initial values of the position $\boldsymbol{\rho} \equiv (x, y, z)$ and the velocity $\dot{\boldsymbol{\rho}} \equiv (\dot{x}, \dot{y}, \dot{z})$. Generally, however, the solution can be expressed through some six integration constants C_1, \dots, C_6 which do not necessarily coincide with the initial conditions. What is important here is that these constants fully define the situation of the planet at the initial time—thereafter, the evolution will be fully determined by the equations of motion, namely,

$$\boldsymbol{\rho} = \mathbf{f}(C_1, \dots, C_6, t), \quad \dot{\boldsymbol{\rho}} = \mathbf{g}(C_1, \dots, C_6, t), \quad (13)$$

with $\mathbf{g} \equiv \dot{\mathbf{f}}$. As explained by Brouwer and Clemence²², the six Kepler elements ω , a , e , M_0 , Ω , and i^* fully specify the initial situation, and can therefore act as the integration constants C_1, \dots, C_6 . In terms of the Cartesian coordinates the above solution will read

$$\begin{aligned} x &= f_1(C_1, \dots, C_6, t), & \dot{x} &= g_1(C_1, \dots, C_6, t), \\ y &= f_2(C_1, \dots, C_6, t), & \dot{y} &= g_2(C_1, \dots, C_6, t), \\ z &= f_3(C_1, \dots, C_6, t), & \dot{z} &= g_3(C_1, \dots, C_6, t). \end{aligned} \quad (14)$$

This specifies the solution for the two-body case. As a rule, we prefer to use the Kepler elements in place of Cartesian coordinates and velocities in both the 2-body and N -body descriptions of the orbits, since the Kepler elements provide a physical picture of the trajectory. In particular, the orbital elements provide the only effective way to describe resonant interactions.

Now, we must address the question of how to express the solution for the N -body problem when one has many position vectors ρ_i , as well as inhomogeneous terms, referred to as disturbing potentials R_i . In this setting, Lagrange⁸ introduced his brilliant invention, the method of variation of constants. (This method was discovered independently by Euler who employed it for calculation of the mutual disturbances of Saturn and Jupiter.) Lagrange suggested that, for each ρ_i , the equations of motion (ignoring the subscript i) can now be expected

$$\ddot{\rho} + \frac{\mu}{\rho_i^2} \frac{\rho}{\rho} = \frac{\partial R}{\partial \rho} \quad , \quad (15)$$

in contrast with (11), and can be solved by assuming a solution of the form

$$\rho = \mathbf{f}[C_1(t), \dots, C_6(t), t] \quad , \quad (16)$$

with each Keplerian parameter $C_i(t)$, $i = 1, \dots, 6$, endowed with a to-be-determined time-dependence of its own. This ansatz renders the Kepler ellipses osculatory. Modelling of the real orbit by a series of instantaneous ellipses is a perfectly legitimate procedure, inasmuch as inserting equations (16) into (15) provides a smooth solution for all $C_i(t)$.

A profound question emerges as to whether this description of trajectories in terms of a sequence of Keplerian ellipses is unique. In the two-body case, i.e., where the ellipse describing a Keplerian orbit is assumed to be stationary, the relationship between the positions and velocities, on the one hand, and the orbital elements, on the other, is unique and invertible. However, let us now consider a situation where the governing ellipse is permitted to move at some arbitrary speed. We can then visualize that, at a specified time, a given planet could reside on any member of an infinite *family* of ellipses that share a common focus—each ellipse would have a spatial point in common with each other and with the instantaneous location of the planet—but each of the ellipses would be moving at different speeds. Fig. 2 illustrates this ambiguity. Although the physical trajectory, as a locus of points in the Cartesian frame, is unique, its description through time-evolving Keplerian elements is no longer unique as we see from the intuitive picture provided above. Moreover as explained in²¹, this ambiguity is similar to the gauge freedom encountered in electrodynamics. This intrinsic ambiguity is an unanticipated feature of the problem which has generally been overlooked and may have some unexpected consequences for the theory of orbits.

To understand better the nature of this non-uniqueness, we wish to point out that the assumption

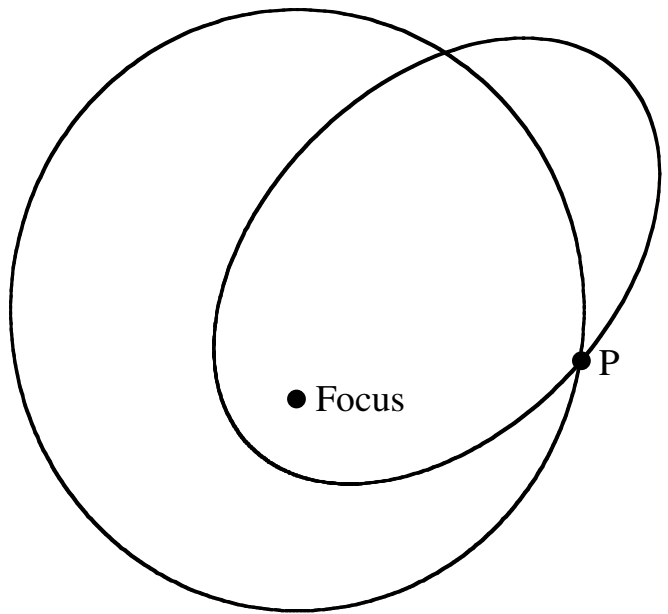


FIG. 2: Two representative coplanar confocal ellipses rotating in the same direction illustrate the ambiguity in the selection of constraints. For example, suppose that the less elongated ellipse rotates and “breathes” slowly, while the more eccentric one rotates and “breathes” swiftly. In one viewpoint, the intersection point P is quickly moving along a slowly-evolving ellipse. In the other, it is slowly moving along a rapidly evolving ellipse. This ambiguity, known by Lagrange, can be removed by the imposition of three additional constraints. Lagrange suggested the trivial constraints (19), but any other choice would lead to a physically identical solution—which will, however, be expressed in a different mathematical form.

implicit to equations (16) completely alters the second column in equations (15). Now $\dot{\rho}$ will also depend also upon the \dot{C}_i which, in the two-body approximation, is assumed to vanish. Therefore, instead of the one-to-one time-dependent mapping $(C_1, \dots, C_6) \leftrightarrow (x, y, z, \dot{x}, \dot{y}, \dot{z})$ one will obtain the mapping $(C_1, \dots, C_6, \dot{C}_1, \dots, \dot{C}_6) \rightarrow (x, y, z, \dot{x}, \dot{y}, \dot{z})$, and that mapping cannot be one-to-one. The three second-order scalar equations contained in (15) will give, after equations (16) are inserted, three second-order equations for the $C_i(t)$. However, in terms of the twelve variables $C_i(t)$ and $V_i(t) \equiv \dot{C}_i(t)$, the three equations (15) will be of the first order only. Together with the six identities $V_i \equiv \dot{C}_i(t)$, we can identify nine first-order differential equations in the twelve variables $C_i(t)$ and $V_i(t)$, which means that the system is underdetermined. To remove these nonphysical degrees of freedom, the system must be augmented by three arbitrary constraints on the constants or parameters $C_i(t)$ and their time-derivatives. We elaborate further on this subtlety after we introduce the equations for $\dot{C}_1, \dots, \dot{C}_6$, namely (22) - (28).

Lagrange evidently appreciated this difficulty, and chose the three additional conditions so that the N -body Cartesian velocities could be expressed through the or-

bital elements in the same manner as they would obtain in the two-body case. Applying the chain rule to equations (16) in that situation, we can write

$$\frac{d\boldsymbol{\rho}}{dt} = \frac{\partial \mathbf{f}}{\partial t} + \boldsymbol{\Phi} = \mathbf{g} + \boldsymbol{\Phi}, \quad (17)$$

where $\boldsymbol{\Phi}$ is a 3-vector

$$\boldsymbol{\Phi}(C_1, \dots, C_6, t) \equiv \sum_i \frac{\partial \mathbf{f}}{\partial C_i} \frac{dC_i}{dt}. \quad (18)$$

Hence, the supplementary condition—preferred by Lagrange—would have the form

$$\boldsymbol{\Phi}(C_1, \dots, C_6, t) = 0, \quad (19)$$

to assure that $d\mathbf{f}/dt$ is identical to $\partial\mathbf{f}/\partial t \equiv \mathbf{g}$. (The functional form of the vector functions \mathbf{f} and \mathbf{g} remains unchanged, independent of their arguments C_i being static or time-dependent.) The next step is to calculate the accelerations.

Using the preceding Lagrange conditions, the second derivatives will satisfy

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = \frac{\partial \mathbf{g}}{\partial t} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} = \frac{\partial^2 \mathbf{f}}{\partial t^2} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt}. \quad (20)$$

After this equation is combined with (16) in the equations of motion (15), we arrive at

$$\frac{\partial^2 \mathbf{f}}{\partial t^2} + \frac{\mu}{r^2} \frac{\mathbf{f}}{r} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} = \frac{\partial R}{\partial \boldsymbol{\rho}}, \quad (21)$$

$$r \equiv \sqrt{f_1^2 + f_2^2 + f_3^2}.$$

Since \mathbf{f} is a solution to the unperturbed equations, and its functional form remains unchanged—even when C_i is permitted to vary with time—the first two terms in (22) cancel just as they did in the two-body case. Therefore, we obtain

$$\sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} = \frac{\partial R}{\partial \boldsymbol{\rho}}. \quad (22)$$

After substantial algebra, the system of two vector equations, (22) and (19), can be expressed with respect to dC_i/dt via the following equations:

$$\frac{da}{dt} = \frac{2}{na} \frac{\partial R}{\partial M_0} \quad (23)$$

$$\frac{de}{dt} = \frac{1-e^2}{na^2e} \frac{\partial R}{\partial M_0} - \frac{(1-e^2)^{1/2}}{na^2e} \frac{\partial R}{\partial a} \quad (24)$$

$$\frac{d\omega}{dt} = \frac{-\cos i^*}{na^2(1-e^2)^{1/2} \sin i^*} \frac{\partial R}{\partial i^*} + \frac{(1-e^2)^{1/2}}{na^2e} \frac{\partial R}{\partial e} \quad (25)$$

$$\frac{di^*}{dt} = \frac{\cos i^*}{na^2(1-e^2)^{1/2} \sin i^*} \frac{\partial R}{\partial \omega} \quad (26)$$

$$- \frac{1}{na^2(1-e^2)^{1/2} \sin i^*} \frac{\partial R}{\partial \Omega}$$

$$\frac{d\Omega}{dt} = \frac{1}{na^2(1-e^2)^{1/2} \sin i^*} \frac{\partial R}{\partial i^*} \quad (27)$$

$$\frac{dM_0}{dt} = -\frac{1-e^2}{na^2e} \frac{\partial R}{\partial e} - \frac{2}{na} \frac{\partial R}{\partial a} \quad (28)$$

This is the celebrated Lagrange system of equations for the osculating Keplerian elements, written under the conditions (19). This system, with some unimportant alterations—such as the choice of a more convenient yet equivalent set of variables—is often used as a basis for long-term numerical integrations.

An orbit belongs to the 12-dimensional space spanned by the Kepler elements $\{C_1, \dots, C_6\} \equiv \{a, e, \omega, i^*, \Omega, M_0\}$ and their derivatives $V_i \equiv \dot{C}_i$. However, due to the conditions (19) imposed by Lagrange, the orbit will be constrained, at each instant of time, to a 9-dimensional submanifold (19) of the said 12-dimensional space. A choice of any other “gauge condition” different from (19) will give different values of the Kepler elements. However, these will be *physically* equivalent to the standard solution (23) – (28), in that they will correspond to the same physical trajectory (i.e., to the same values of $x, y, z, \dot{x}, \dot{y}, \dot{z}$).

Suppose we *did not* impose the trivial constraint conditions (19). Then, instead of equations (20), (22) and (22), we would obtain

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = \frac{\partial \mathbf{g}}{\partial t} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} + \frac{d\boldsymbol{\Phi}}{dt} \quad (29)$$

$$= \frac{\partial^2 \mathbf{f}}{\partial t^2} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} + \frac{d\boldsymbol{\Phi}}{dt},$$

$$\frac{\partial^2 \mathbf{f}}{\partial t^2} + \frac{\mu}{r^2} \frac{\mathbf{f}}{r} + \sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} + \frac{d\boldsymbol{\Phi}}{dt} = \frac{\partial R}{\partial \boldsymbol{\rho}}, \quad (30)$$

$$r \equiv \sqrt{f_1^2 + f_2^2 + f_3^2},$$

and

$$\sum_i \frac{\partial \mathbf{g}}{\partial C_i} \frac{dC_i}{dt} = \frac{\partial R}{\partial \boldsymbol{\rho}} - \frac{d\boldsymbol{\Phi}}{dt}, \quad (31)$$

paralleling the development in § III. The corresponding equations of motion for the osculating elements then reads

$$\frac{da}{dt} = \frac{2}{na} \left[\frac{\partial R}{\partial M_0} - \boldsymbol{\Phi} \frac{\partial \mathbf{g}}{\partial M_0} - \frac{\partial \mathbf{f}}{\partial M_0} \frac{d\boldsymbol{\Phi}}{dt} \right], \quad (32)$$

$$\begin{aligned} \frac{de}{dt} = & \frac{1-e^2}{na^2e} \left[\frac{\partial R}{\partial M_0} - \Phi \frac{\partial \mathbf{g}}{\partial M_0} - \frac{\partial \mathbf{f}}{\partial M_0} \frac{d\Phi}{dt} \right] \\ & - \frac{(1-e^2)^{1/2}}{na^2e} \left[\frac{\partial R}{\partial a} - \Phi \frac{\partial \mathbf{g}}{\partial a} - \frac{\partial \mathbf{f}}{\partial a} \frac{d\Phi}{dt} \right], \end{aligned} \quad (33)$$

$$\begin{aligned} \frac{d\omega}{dt} = & \frac{-\cos i^*}{na^2(1-e^2)^{1/2} \sin i^*} \left[\frac{\partial R}{\partial i^*} - \Phi \frac{\partial \mathbf{g}}{\partial i^*} - \frac{\partial \mathbf{f}}{\partial i^*} \frac{d\Phi}{dt} \right] \\ & + \frac{(1-e^2)^{1/2}}{na^2e} \left[\frac{\partial R}{\partial e} - \Phi \frac{\partial \mathbf{g}}{\partial e} - \frac{\partial \mathbf{f}}{\partial e} \frac{d\Phi}{dt} \right], \end{aligned} \quad (34)$$

$$\begin{aligned} \frac{di^*}{dt} = & \frac{\cos i^*}{na^2(1-e^2)^{1/2} \sin i^*} \left[\frac{\partial R}{\partial \omega} - \Phi \frac{\partial \mathbf{g}}{\partial \omega} - \frac{\partial \mathbf{f}}{\partial \omega} \frac{d\Phi}{dt} \right] \\ & - \frac{1}{na^2(1-e^2)^{1/2} \sin i^*} \left[\frac{\partial R}{\partial \Omega} - \Phi \frac{\partial \mathbf{g}}{\partial \Omega} - \frac{\partial \mathbf{f}}{\partial \Omega} \frac{d\Phi}{dt} \right], \end{aligned} \quad (35)$$

$$\frac{d\Omega}{dt} = \frac{1}{na^2(1-e^2)^{1/2} \sin i^*} \left[\frac{\partial R}{\partial i^*} - \Phi \frac{\partial \mathbf{g}}{\partial i^*} - \frac{\partial \mathbf{f}}{\partial i^*} \frac{d\Phi}{dt} \right], \quad (36)$$

$$\begin{aligned} \frac{dM_0}{dt} = & - \frac{1-e^2}{na^2e} \left[\frac{\partial R}{\partial e} - \Phi \frac{\partial \mathbf{g}}{\partial e} - \frac{\partial \mathbf{f}}{\partial e} \frac{d\Phi}{dt} \right] \\ & - \frac{2}{na} \left[\frac{\partial R}{\partial a} - \Phi \frac{\partial \mathbf{g}}{\partial a} - \frac{\partial \mathbf{f}}{\partial a} \frac{d\Phi}{dt} \right]. \end{aligned} \quad (37)$$

with the three components of the vector function Φ still undetermined. Therefore, we remain at liberty to assign them fixed values, or to impose three arbitrary constraints upon the functions $C_i(t)$ and \dot{C}_i .

As mentioned above, this algebraic freedom does not compromise the physical content of the theory. The Φ -contributions will ultimately disappear if we employ the solutions for \dot{C}_i in order to calculate the orbit in terms of x, y, z and $\dot{x}, \dot{y}, \dot{z}$. This situation is reminiscent of the gradient invariance in classical electrodynamics, where one can describe the field configuration either in terms of the physical entities \mathbf{E} and \mathbf{B} , or in terms of the 4-dimensional potential $A^\mu \equiv (\phi, \mathbf{A})$ from which they can be derived. In the latter case, the description is inherently ambiguous, and gauge fixing is permissible whenever convenient.

The Φ -terms in the Lagrange system may be ignored in analytic calculations, provided that one imposes the Lagrange gauge conditions (19). In the course of numerical integration, however, the accumulation of numerical error will provide a nonzero contribution to Φ . For this reason, the Φ -terms will be in some sense non-vanishing, even in the Lagrange gauge. Whether this will effect the precision of computation remains to be seen. What is important for us here is the unequivocal analogy between the Lagrange system (32) – (37) and equations (8).

In the toy model (1), the variable x is evidently analogous to the Cartesian coordinates emerging in the orbital calculations, i.e., to the x, y, z components of the vector ρ . The variables $S(t)$ and $C(t)$ are analogs of the oscillating elements, while the Φ -term in (8) plays a role identical to that of Φ -terms in (32) – (37). In particular, we see that it is not possible through some choice of Φ , to eliminate the “fast” time scale in (8). For an arbitrary gauge Φ , the integration of (8) must be carried out with small time steps determined by the “fast” functions $\sin(t)$ and $\cos(t)$ and not with the long time steps determined from the “slow” function $F(t)$. Similarly, no choice of the gauge Φ in the Lagrange system (32) – (37) will remove the short-period terms in its right-hand side, although a good choice of gauge may, potentially, simplify the calculations. As is well known, an appropriate choice of gauge in electrodynamics considerably simplifies solution of the equations of motion. In a similar manner, an appropriate choice of the arbitrary vector Φ may simplify the right-hand side in (32) – (37). However, the freedom of choice of Φ is not sufficient to eliminate the short-period terms in all six Lagrange equations. For details, see²¹.

V. CANONICAL TRANSFORMATIONS AND INTRINSIC LIMITS

The elimination of short-period terms in Hamiltonian problems is performed following different theoretical perspectives by various means. One powerful method is to present the orbital-motion problem in canonical form, then to expand the Hamiltonian into a series, utilizing a small parameter, and then to employ the method of canonical transformations in order to get rid of the “fast” terms to a certain order in the parameters¹³. In this section we shall explain why canonical transformations, too, cannot alter in a significant way the presence of the fast time scale.

We begin with the Hamiltonian

$$H(p, x, t) = \frac{p^2}{2} + \frac{x^2}{2} - xF(t) \quad , \quad (38)$$

that generates the dynamical equations

$$\dot{x} = \frac{\partial H}{\partial p} = p \quad (39)$$

and

$$\dot{p} = -\frac{\partial H}{\partial x} = -x + F(t). \quad (40)$$

These can be combined to give

$$\ddot{x} + x = F(t) \quad , \quad (41)$$

as in (1). It is natural to think of the underlying simple harmonic oscillator as the unperturbed part of this problem. Accordingly, we employ the generating function²⁰

to transform from (x, p, t) to (θ, J, t) .

$$F_1(\theta, x, t) = \frac{x^2}{2} \cot \theta \quad (42)$$

from which we obtain

$$p = \frac{\partial F_1}{\partial x} = x \cot \theta \quad (43)$$

and

$$J = -\frac{\partial F_1}{\partial \theta} = \frac{x^2}{2 \sin^2 \theta} \quad (44)$$

Put another way, we can write

$$x = \sqrt{2J} \sin \theta \quad \text{and} \quad p = \sqrt{2J} \cos \theta \quad (45)$$

and

$$\begin{aligned} \tilde{H}(J, \theta, t) &= H(x, p, t) + \frac{\partial F_1}{\partial t} \\ &= J - \sqrt{2J} F(t) \sin \theta \quad . \end{aligned} \quad (46)$$

In this latter expression for the transformed Hamiltonian \tilde{H} , we identify the leading term, namely J , with the action integral and the unperturbed simple Harmonic oscillator, which is cyclic in the angle variable θ . The remaining term, however, corresponds to the ‘‘perturbation.’’

In keeping with the procedures associated with canonical perturbation theory, we seek to eliminate the unperturbed part of the through the use of an appropriate transform of type F_2 wherein we transform from (J, θ, t) to (α, β, t) . In particular, we now define

$$F_2(\theta, \alpha, t) = (\theta - t) \alpha \quad (47)$$

from which we obtain

$$J = \frac{\partial F_2}{\partial \theta} = \alpha \quad \text{and} \quad \beta = \frac{\partial F_2}{\partial \alpha} = \theta - t \quad (48)$$

yielding the new Hamiltonian \hat{H}

$$\begin{aligned} \hat{H}(\alpha, \beta, t) &= \tilde{H}(J, \theta, t) + \frac{\partial F_2}{\partial t} \\ &= -\sqrt{2\alpha} \sin(\beta + t) F(t) \quad . \end{aligned} \quad (49)$$

We observe that the transformed Hamiltonian \hat{H} is composed solely of perturbing terms and, assuming that they are small, we expect both α and β to undergo only small changes and to remain essentially constant. For this reason, canonical perturbation theory is sometimes called the method of parameters. Here, however, the parameters are canonically conjugate variables. With that in mind, we observe that the transformed equations of motion are

$$\dot{\alpha} = -\frac{\partial \hat{H}}{\partial \beta} = -\sqrt{2\alpha} \cos(\beta + t) F(t) \quad (50)$$

and

$$\dot{\beta} = \frac{\partial \hat{H}}{\partial \alpha} = -\frac{1}{\sqrt{2\alpha}} \sin(\beta + t) F(t) \quad . \quad (51)$$

There are several approaches available for solving this pair of coupled ordinary differential equations. The simplest, which is equivalent to the Picard iteration, corresponds to keeping α and β fixed on the right hand side, and integrating these equations to obtain their left hand sides, i.e., their time variability. Significantly, we see that we have to integrate terms such as $\sin(t)$ and $\cos(t)$ multiplied by $F(t)$. This is exactly the same problem we confronted in our simple, yet exact and linear, treatment of the original problem. Thus, we observe that a Hamiltonian treatment does not circumvent the problem that we identified earlier. Even if our ‘‘parameters’’—in this case, the conjugate variables α and β —are expected to be slowly varying, their variation will necessarily contain frequencies commensurate with those of the underlying simple harmonic oscillator. A second approach to this problem—an approach that is closely related to the Floquet theory—corresponds with a Fourier decomposition of $F(t)$, for example, via sine and cosine series in nt , where n is an integer. Importantly, we observe that the fundamental frequency ($n = 1$) describes a resonance, while the other higher frequencies, at least to this order, do not²³.

Accordingly, the presence of a resonant term in $F(t)$ will result in a systematic linear-in-time drift in both α and β . This, in turn, implies a frequency shift in the oscillator since $\beta + t$ will have an effective frequency that is shifted from its original value of unity. This is to be expected, since the modulation of the orbit by the forcing function $F(t)$ can alter both the frequency, phasing, and amplitude of the oscillator. The analysis of fast and slow time variables, higher order resonances, averaging methods, and other approximation schemes is a well-studied topic but remains beyond the scope of this paper. The essential point of our derivation is to show that the computational treatment of an *exact* canonical statement of our problem nevertheless requires that we employ time steps that are small compared with the fundamental frequency of unity, even if $F(t)$ is a very slowly varying function of time. While canonical methodology provides improved opportunities for *approximate* treatments of the problem—e.g., via decomposition into multiple time scales—no significant speed enhancement is achieved using canonical formalism. However, an argument can now be made, since all methodologies require that we employ short time steps, that the application of a symplectic integration scheme with a short time step to the canonically conjugate equations could offer some advantage to the direct numerical integration of the non-canonical ‘‘constants of variation’’ since they preserve the Poincaré invariants of the system²⁴.

VI. DISCUSSION

The method of variation of constants was originally invented by Euler and Lagrange, and it occupies an important role in the repertoire of mathematical tools employed in the solution of ordinary differential equations, among other problems. The usual specification of such techniques often assumes that solutions obtained to homogeneous versions of a problem can be employed in the solution of inhomogeneous problems—both linear and nonlinear—where an external forcing term is present. Moreover, we often find that the “constants” associated with the solution of the homogeneous problem can be utilized in the solution of the inhomogeneous problem, if they are allowed to vary with time. Unfortunately, it is widely assumed that a slowly-varying driver, in the presence of rapidly-varying homogeneous solutions, results in these “constants” varying at a rate determined by the inhomogeneous forcing function.

We introduced a linear paradigm that is closely related to the classical derivation of the method of variation of constants for coupled inhomogeneous linear ordinary differential equations, and we observed that this assumption is incorrect. In addition, we noted that the conventional derivation of the method assumes one or more constraints, often having a trivial form. We showed that these constraints can be generalized, in analogy to gauge theories in physics, and observed that different constraints can offer conceptual advances and methodological benefits to the solution of the underlying problem.

Finally, we turned to the N -body problem of celestial mechanics—the example *par excellence* of nonlinear problems identified by Euler and Lagrange—which con-

stituted the basis of their invention of the method of variation of constants. In the orbital mechanics problem, these constants corresponded to the Kepler or orbital elements associated with the two-body problem. We have shown here that their methodology corresponded to the selection of a specific gauge which is chosen in order to overcome problems of non-uniqueness and the underdetermined character of the problem. We have shown how the selection of the gauge in the N -body problem can be generalized, although we have not identified an optimal or improved choice for it. Nevertheless, the identification of this gauge freedom indicates how readily we can be misled into thinking that mutual interactions between the planets can be treated as though they are slowly varying.

We concluded by turning to canonical perturbation theory, to see if it can provide any additional insights. Importantly, we observed how fast time scales associated with the homogeneous solutions remain embedded in the problem, necessitating the need for considerable caution in any attempt to eliminate by averaging the fast time behavior. Indeed, one outcome of our excursion into canonical perturbation theory was the observation that we may be better served by employing symplectic integration methods—that is, methods based on canonical transformations—with short time steps, in contrast with the method of variation of constants, to preserve better the underlying Hamiltonian character of the problem.

Acknowledgments

The authors wish to thank Susanna Musotto and Ferenc Varadi for providing the data in Figure 1.

* Electronic address: win@ucla.edu

† Electronic address: efroimsk@ima.umn.edu

¹ R. Courant, R. 1938. *Differential and Integral Calculus*. (Nordemann, New York, 1938); H. Jeffreys and B.S. Jeffreys, *Methods of Mathematical Physics*, (Cambridge University Press, Cambridge, 1972).

² R. Courant and F. John, *Introduction to calculus and analysis*. (Springer-Verlag, New York, 1989).

³ V.M. Alekseev, Vestnik Moskovskogo Universiteta, Seria I. Matematika i Mehanika **2**, 28 (1961).

⁴ H. Brunner, Utilitas Mathematica **19**, 255 (1981).

⁵ S.G. Deo and E.F. Torres, Applied Mathematics and Computation **24**, 263 (1987).

⁶ V. Lakshmikantham and S.G. Deo, *Method of Variation of Parameters for Dynamic Systems* (Gordon and Breach, Amsterdam, 1998).

⁷ L. Euler *Recherches sur la question des inegalites du mouvement de Saturne et de Jupiter, sujet propose pour le prix de l'annee*. Berlin, 1748; For modern edition see: L. Euler *Opera mechanica et astronomica*. (Birkhauser-Verlag, Switzerland, 1999).

⁸ J.-L. Lagrange, *Mécanique analytique*. (Académie des Sci-

ences, Paris, 1913).

⁹ W.M. Kaula, *Theory of Satellite Geodesy* (Blaisdell, London, 1966).

¹⁰ Musotto, S., F. Varadi, W. Moore, and G. Schubert, accepted for publication in Icarus.

¹¹ C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, 1971.

¹² M.I. Rabinovich and A.A. Rozenblium, J. Appl. Math. Mech. **36**, 305 (1972).

¹³ A. Milani, A.M. Nobili, and M. Carpino, Astronomy and Astrophysics **172**, 265 (1987).

¹⁴ D.L. Richardson, and J.W. Mitchell, J. Appl. Mech. **66**, 273 (1999).

¹⁵ K.R. Grazier, W.I. Newman, W.M. Kaula, and J.M. Hyman, Icarus **140**, 341 (1999); K.R. Grazier, W.I. Newman, F. Varadi, W.M. Kaula, and J.M. Hyman, Icarus **140**, 353 (1999).

¹⁶ S. Maezawa T. Miyagawa, Bull. JSME **91**, 49 (1973).

¹⁷ S.S. Dallas, and E.A. Rinderle, J. Aeronautical Sciences **21**, 153 (1974).

¹⁸ S. Tani and M. Inokuti, J. Comput. Phys. **11**, 409 (1973).

¹⁹ W.I. Newman and W.R. Thorson, Phys. Rev. Lett. **29**,

- 1350 (1972); W.I. Newman and W.R. Thorson, *Can. J. Phys.* **50**, 2997 (1972).
- ²⁰ H. Goldstein, *Classical Mechanics*, 2nd Ed, (Addison-Wesley, Reading, MA, 1980).
- ²¹ M. Efroimsky, Preprint No 1844, Institute for Mathematics and its Applications, University of Minnesota. <<http://www.ima.umn.edu/preprints/feb02/feb02.html>> (Submitted to " *Celestial Mechanics and Dynamical Astronomy*")
- ²² D. Brouwer and G.M. Clemence, *Methods of Celestial Mechanics* (Academic Press, New York, 1961).
- ²³ A. H. Nayfeh, *Introduction to Perturbation Techniques*, (Wiley, New York, 1993).
- ²⁴ J. M. Sanz-Serna and M. P. Calvo , *Numerical Hamiltonian Problems* (Chapman and Hall, London, 1994).