Newton's Laws of Motion,

Reference Frames

and Inertia

Chris O'Loughlin

2 December 2011

Abstract

The validity of Newton's Laws of Motion depends on the type of reference frame they act in. They are valid in inertial reference frames and not valid in non-inertial reference frames. Reference frames thus play a pivotal role any understanding of the laws themselves. So, what must we know in general about reference frames? And if a reference frame can be inertial or non-inertial, what must we know about inertia? This paper addresses these concepts and clarifies some common misconceptions surrounding them. Additionally, a modified definition of inertia is proposed that allows for a different formulation of Newton's laws making them valid in either type of reference frame.

Introduction

Newton's Laws of Motion (NLM) were originally written in Latin and have various English interpretations. A common form of the laws is:[1]

1. An object at rest will remain at rest, or if in motion will remain in motion in a straight line, unless acted upon by an external force.

2. Force is equal to mass times acceleration: $F = ma$.

3. For every action there is an equal and opposite reaction.

But this is not the whole story. To accurately portray how the laws operate in nature they must be accompanied by a caveat. This caveat introduces subtle complexities that can, and as will be demonstrated below often do, introduce significant misconceptions.

[1] All relativistic effects are assumed to be negligible throughout this paper.

The caveat is that the laws are not always valid. They are valid only under certain conditions. The conditions under which NLM are valid are referred to as Inertial Reference Frames (IRFs). So, what is the definition of an IRF? An IRF is defined to be a system in which NLMs are valid, and that is about all that can be said on the matter if one is to be rigorously precise. While true in the strictest sense, the circularity inherent in this definition, i.e., "NLM are valid only in IRFs, and IRFs are defined as systems in which NLM are valid", does not contribute anything to an understanding of the deeper fundamental principles (and might even raise some vague suspicions about why a more helpful definition is not provided). In lieu of a clear and concise definition for an IRF various authors proceed to provide examples of different types of reference frames to compare and contrast. Unfortunately, the examples provided can be vague, misleading or just plain wrong – as will be demonstrated below. The definitions that are universally true, and will be relied upon in this paper, are "If a system of objects is observed to behave in accordance with NLM then it is said to be in an IRF, otherwise it is an NIRF[2]."

A "quick and dirty" conceptual test to estimate the legitimacy of any example that purports to illustrate whether an IRF or a NIRF is present is based on the first law of motion and will be referred to here as the "Let Go Test (LGT)". The LGT is performed as follows: An observer holds an object in place such that the velocity of the object is zero relative to the observer. The observer here can be thought of as being coincident with the RF relative to which the motion of the object will be measured. There are no external forces on the object and observer lets go of it. If the object remains in the same location relative to the observer then an IRF is present, otherwise a NIRF is present. This test simply makes use of the "An object will remain at rest unless acted upon by an external force" part of the first law. The rest of the first law could be included in the LGT to see whether an object in motion will remain moving at a constant velocity but this is not necessary because when you come right down to it the question is whether the object will be observed to accelerate or not in the absence of an external force – whether the object accelerates from a state of rest or from some established velocity does not matter. The first law is often said to be a special case of the second law and this test is simply a distillation of the combined basic content of the first and second law into an easily performed

[2] This, of course, assumes that there are IRFs in nature, which is, strictly speaking, not true. All space-time curvature is subject to at least some small amount of non-uniformity, which as Einstein pointed out results in non-uniform "gravitational" acceleration, and so the behavior of the objects in any physically real system will be to some extent non-inertial. Even in free fall, two objects in close proximity will gravitationally accelerate towards each other to some extent, however small, in the absence of any external force, thus violating the first law, indicating a NIRF is present. However, at this stage of the game the best approach is probably to proceed as if IRFs do exist while deliberately and clearly noting that in doing so we are only working with conceptual approximations of reality.

experiment to determine whether NLM are adhered to or not. The third law is always assumed to be satisfied.

Flawed Examples of IRFs and NIRFs

As noted above, the definition of an IRF as a RF in which NLM are valid is not helpful, so attempts are often made to differentiate between IFRs and NIRFs through the use of examples. However, as noted above, it is all too common for such examples to be improperly employed. Perhaps one reason for the improper employment is the lack appreciation of the fact that, as Einstein pointed out almost 100 years ago, gravitational acceleration, i.e., free fall motion, is not the result of any force - it is due to the curvature of space-time. While it is true that it is impossible to measure any force causing free fall acceleration, it goes much deeper than that. The salient point is not that there is simply no way to measure such a force, it's that no such force exists at all. This is accepted, "settled science" and has been for some time yet somehow it can slip through the cracks, perhaps because the formidable mathematical and conceptual challenges associated with general relativity make a comprehensive, in-depth explanation outside the scope of most presentations. But it would be a shame, and would show perhaps a certain lack of respect for the audience's intellectual capabilities, to dumb it down for the sake of brevity. A short note acknowledging that gravitational acceleration is not due to any force, that is it is the result of space-time curvature and that additional details are beyond the scope of the current presentation would suffice. ***In this paper the assertion that gravitational acceleration is not caused by any force will be rigorously maintained.*** There is no good reason to do otherwise. Following this approach, the need for the caveat that NLM are only valid in IRFs begins to emerge, that reason being that NLM by themselves do not internally account for the fact that some accelerations, e.g., free fall, occur even in the total absence of any external forces.

With the knowledge that gravitational acceleration is not caused a force firmly established an examination of the difficulties with some examples of the various types of RFs can proceed. One common mistake is to attempt to distinguish between whether IRFs or NIRFs are in operation based on the state of acceleration of the objects and/or observer. Consider the following examples:

1. "If your frame of reference has a non-uniform, or accelerated motion, then Newton's first law of motion, the Law of Inertia, will appear to be wrong, and you must be in a non-inertial frame of reference."

2. "An inertial frame of reference is a frame of reference with constant velocity."

Statements such as these probably arise out of the observation that an accelerating RF will measure an object to be accelerating even when no external force is acting on the object, which is, by definition, inconsistent with Newton's first law. This very real inconsistency is rightly attributed to the accelerative nature of this particular RF, but the conclusion that therefore *all* accelerating RFs must be NIRFs is not justified. However, assertions such as these are very common. To illustrate the problem at hand let's begin with the first assertion above by imagining an RF and an object in free fall above the Earth. Being in free fall this system is accelerating towards the Earth with no forces operating on the object, yet when the LGT is conducted the object remains at rest relative to the accelerating RF. In this case the object behaves exactly as NLM say it will, which means that an IRF is in operation, yet the system is accelerating, which contradicts assertion 1 above. That is, we have shown that in the case of free fall motion a RF with non-uniform, or accelerated, motion can indeed operate as an IRF. As for the second assertion, consider an observer holding an object in a RF on the surface of the Earth. While held by the observer the object is at a constant velocity relative to the RF yet when the object is released it accelerates downwards while no external force acts upon it. In this case, the object accelerates relative to the RF while no external force acts upon it, which violates NLM, and the RF is not accelerating, all of which taken together contradicts the second assertion. That is, we have shown that a reference frames with constant velocity *can* be a NIRF. So, a determination as to whether a system is an IRF or a NIRF cannot be made solely based on the acceleration of the system. The table below lists examples of each type of RF associated with accelerated and non-accelerated RFs. (Again, these examples suffer from the assumption that IRFs exist at all in nature, which they do not, but this will be rectified below.)

	IRF	NIRF
Accelerating	Free Fall	Inside a Rocket Ship in Deep Space with Thrusters On
Non-Accelerating	Inside a Rocket Ship in Deep Space with Thrusters Off	At Rest on the Surface of the Earth

Another common error related to the one addressed above arises when describing the details of the type of RF in operation at the Earth's surface. It is not unusual to find claims along the lines of "We can usually treat reference frames on the surface of the Earth as inertial frames. (Since the Coriolis effect is generally small enough to be ignored.)". Such claims are false. Reference frames fixed on the Earth are clearly not IRFs – and not just because of the Earth's rotation. Even if the Earth did not rotate any released object would not remain at rest relative to an RF affixed to the Earth's surface but would (gravitationally) accelerate downwards in the absence of any force – which by definition means Newton's first law does not hold and a NIRF is present. True, Coriolis effects will contribute to non-inertial observations but that is only part of the story. Again, the fact that gravitational acceleration is not produced by any force is a critical element here.

Another example of a very common, yet inaccurate assertion regards the relationship between RFs. It is not difficult at all to find even in very well respected, pier reviewed publications and texts assertions along the lines of that “Any reference frame that moves with constant velocity relative to an inertial reference frame is also an inertial reference frame.” The literature is saturated with such claims, yet they are simply untrue. To see why, consider some reference frame, *RFa*, in free fall above the Earth. Reference frames in free fall are IRFs, so *RFa* is an IRF. Now consider another reference frame, *RFb*, that is directly above *RFa* and that is also in free fall. *RFb* is also an IRF for the same reasons that *RFa* is an IRF. However, gravitational acceleration varies as a function of height such that *RFa* is accelerating more rapidly than *RFb* is. Therefore, *RFa* and *RFb* are accelerating relative to each other - yet they are both IRFs. The only way to achieve a constant relative velocity between *RFa* and *RFb* is to accelerate (at least) one of them via the application of a force, which then makes that reference frame non-inertial. In fact, since all space-time is to some extent non-uniformly curved, it is generally true that all objects that are not subjected to an external force will (inertiallly) accelerate at different rates relative to each other. So, since at least one force is needed for two RFs to move with a constant velocity relative to each other the quote above would be correct if re-worded to “Any reference frame that moves with constant velocity (say, a car or an airplane) relative to an inertial frame ***cannot*** also be an inertial reference frame.” or similarly, “For any 2 RFs that are not accelerating relative to each other at least one of the RFs must be non-inertial.”

To perhaps shed additional light on the flaw in the example above, and suggest a possible origin of the flaw, consider a mathematical or computer simulation of Newton’s laws in two RFs, *R* and *R’*, moving with a constant velocity relative to each other as in the example. In such a model various objects and their associated masses can be defined. Simulated forces can be applied to the objects which will respond in accordance with Newton’s second law $F = ma$. At some time *To* let all the forces be set to zero. After *To*, any objects that were at rest relative to *R* will remain at rest relative to *R* and any objects that had acquired non-zero velocities relative to *R* will retain those velocities relative to *R*, and the same goes for *R’*. As such, Newton’s first law is dutifully obeyed in both *R* and *R’* and so both *R* and *R’* are IRFs. In particular, consider two objects, *O1* and *O2*, that are at rest relative to each other. If no subsequent external forces are applied to these objects they will continue to remain at rest relative to each other indefinitely as measured in both *R* and *R’*.

While this is a relatively simple simulation to generate, it is entirely impossible to observe any such behavior in the physical universe due to the fact that NIRFs will always be in action to some extent. In an actual physical environment, in the absence of any external forces, the two objects will gravitationally accelerate towards each other. . To illustrate, let’s measure the motion of *O1* in *R1* and the motion of *O2* in *R2*, where both *O1* and *O2* are actual objects in the physical universe and while no external forces are operating on the objects. For Newton's Laws to be valid separately in either RF, that is, for both RFs to be IRFs relative to the objects being observed in them, *O1* must be at rest relative to *R1* and *O2* must be at rest relative to *R2*. This is only possible if *R1* and *R2* are accelerating relative to each other since *O1* and *O2* are accelerating relative to each other. Thus, unlike in mathematical models or computer simulations, in the real physical universe IRFs cannot have constant velocities relative to each other. The effects of the physical existence of the non-uniform curvature of space-time will always make themselves known in any measure of motion in the physical universe, but can be left out when making such measurements in conceptual or simulated RFs. We can conjure up a purely mathematical or simulated environment where "Any reference frame moving with constant velocity relative to an inertial reference frame is also an inertial reference frame." is true, but physical reality cannot be so manipulated.

It is interesting to note that problems are encountered in the example above even though the mathematics used in the example is entirely correct. That is, the problem is not that the calculations were wrong it's that some physics has been left out. In fact, the simplicity and flawlessness of the pure mathematics can give one an enhanced level of confidence which tends to mask the inadequacy of the underlying simulation of physical nature. It is an important reminder of the difference between the mathematician and the physicist and good example of the need to fully appreciate the difference between the theoretical and the real.

At this point, as no helpful definition of an IRF has been found, perhaps it would be instructive to consider a concept that is very closely related to that of a RF: The coordinate system (CS). As used here, a CS is any system that allows an observer to measure location and time. By providing an observer the ability to determine various locations at various times a CS will allow one to compute such derived parameters as velocity and acceleration. One can physically construct a real CS in ones garage out of articles purchased at a home store or a CS can be a purely conceptual or imagined part of a thought experiment. So, is there any difference between a RF and a CS? What if we assume that "RF" and "CS" are really different terms for the same concept? In that case then a CS will be either an IRF or a NIRF. But, as noted above, one can physically construct a CS ones garage. Having done so, what type of CS has been constructed – an IRF or a NIRF? If it is a NIRF, could one somehow purchase different materials in order to make an IRF instead? Or, if it is an IRF, does it become a NIRF if I move it to another location even though the materials used to construct the CS don't change? Or, is it simply a CS and that is all there is to say about it and we leave whether NLMs are satisfied or not to the term RF? Or, can we rely solely on CSs and dispense with the concept of RFs when examining NLMs? More will be said about this below.

<u>Inertia and the Laws of Motion</u>

As discussed above, gravitational acceleration is not the result of a force. A force *can* cause an acceleration, (such accelerations are often referred to as proper accelerations) but the acceleration due to gravity is not caused by a force. To reiterate, it's not just that there is no way to measure any force causing a gravitational acceleration, it's that there simply is no such force present. To see why, consider an observer in a rocket with the motors turned off in a remote region of space. The rocket has no windows and has walls through which no information of any kind can travel. Ignoring any tidal effects, the observer will have absolutely no way of determining whether he is "at rest" or moving with a constant velocity relative to any other objects outside of the rocket. Any objects inside the rocket with him simply float along with the rocket itself. Now consider the same rocket in free fall towards the Earth. As first pointed out by Einstein there is no physical difference of any kind between the conditions inside the rocket whether it is in the remote region of space or in free fall towards the Earth. There were no forces operating on the objects in the rocket while it was in the remote region and there are no forces operating on the objects in the rocket while it is in free fall.

To be thorough, and as general as possible, we must note that the recessional accelerations observed at great distances due to universal expansion as well are similarly not the result of any applied forces. Just as in the case of gravitational acceleration, from inside the rocket there is no way to measure our acceleration relative to the very distant heavenly bodies we are receding away from as the universe expands, and so we must concluded that there is no force causing that relative acceleration either. In general then, there can be attractive or recessional inertial accelerations throughout the universe. The term "ambient acceleration" will be used to cover both the attractive and recessional inertial accelerations. As such, we see that every location in the universe has an associated ambient acceleration.

It should be noted that while the claim that any location in the universe has an associated ambient acceleration has been made, what those accelerations are relative to has not yet been defined. To see why this is a very bad practice, and to proceed towards correcting this error, consider an object in free fall very near the surface of the Earth. We may say that this motion is an acceleration of magnitude g, but what is that acceleration relative to? The acceleration does have a magnitude of g relative to the surface of the Earth, but relative to a rocket ship that has launched and is accelerating upwards the free fall acceleration will be measured to be greater than g. So, as illustrated here, we see that virtually any magnitude what-so-ever can be measured. So what coordinate system should be used? To answer this question, recall that we are trying to understand the real, physical interaction between the two objects – Earth and the object in free fall – that's all. We are not trying to see how many different imaginary reference frames we can measure the acceleration in, nor how many different acceleration magnitudes we can observe. We are trying to describe the physics underlying the way the objects and space-time interact to produce motion. Since we are trying to explain this interaction let us define the ambient accelerations of the Earth and the object to be their accelerations measured ***relative to***

each other. This acceleration is not arbitrary – it is invariant under any coordinate transformation. All RFs will measure this acceleration to be the same magnitude. More generally, for two gravitationally interacting objects that are sufficiently isolated from any other objects, given the masses and distance between them one can calculate the acceleration of each object relative to the other when no external forces are present – and that acceleration is the ambient acceleration as defined here. Any number of additional objects can be added to such a configuration and the ambient acceleration of each individual object can be calculated as the acceleration of that object relative to the center of mass of all the other objects.

All of the above highlights the significance of the caveat that NLM are valid only in IRFs. However, since space-time is in fact nowhere perfectly flat there is no such thing as a true IRF in nature meaning that, strictly speaking, NLM are *never* valid as anything other than conceptual approximations. Thus if NLM can be reformulated to describe motion in NIRFs they will describe motion in all possible physical RFs. In order to restate NLM such that they are valid generally in nature we need to reconsider the concept of inertia.

The concept of inertia is introduced in Newton's First Law of Motion, which is often referred to as the Law of Inertia. This law essentially states that an object will not accelerate unless acted upon by an external force. However, as noted above, every location in the universe has an associated ambient acceleration that material objects will partake of – without the application of any external force. Consider an observer holding an object near the surface of the Earth. If the object is released it will accelerate with no force acting on it. If the object is lifted to a greater height and released, again it will accelerate with no force acting on it. This can be repeated indefinitely, showing that no matter where the object is released from it will not "remain at rest unless an external forces acts upon it". In fact, in direct contradiction to the first law, a force is required to *prevent* the object from accelerating. And note that a force is only measurable when the object is not accelerating per the ambient acceleration. This means that whereas inertia had been seen to be the tendency of all material objects to resist acceleration it is now seen to be the tendency of all material objects to resist deviating from the ambient acceleration at their location. As such, a reformulation of the first law that is true in general - without any additional caveat - is "An object will accelerate in accordance with the ambient acceleration at it's location unless acted upon by an external force." By doing away with the need for any caveat with this definition no distinction between IRFs and NIRFs is needed at all and in fact any reliance on the concept of RFs in relation to NLM can be completely done away with. The challenges associated with defining and considering all the subtleties surrounding RFs are eliminated because RFs are no longer required when describing how the laws of motion operate. After all, there is only one type of RF in nature so why bring other non-existent concepts into the conversation?

This way of viewing inertia impacts the second law of motion as well. As noted above, all objects that are not under the influence of an external force will partake of the ambient acceleration at their location. They will already be accelerating before any force is applied to them so the second law becomes "The force on an object is equal to the mass of that object multiplied by the difference between the measured acceleration and the ambient acceleration at the location of the object.", or $F = m(Am - Aa)$ where Am is the measured acceleration of the objects relative to each other and Aa is ambient acceleration as defined above. Said a different way, a force is required for any deviation from the ambient acceleration. Again, any need to specify whether an IRF or NIRF is in operation has been removed. And why not? There are no IRFs in nature. The only type of RF possible is a NIRF. If there is only one type of RF in the physical universe there is no need, in fact there is no way, to distinguish between physical RF types when describing actual, physical phenomena.

Consider some applications of this form of the second law. First imagine some object in free fall above the Earth. Any attempt to reconcile this behavior with the original second law alone is unsuccessful because the object is accelerating when no force is producing the acceleration. That is, neither the acceleration a nor the mass m in $F = ma$ are zero, but as pointed out by Einstein the force F is zero, so the equation $F=ma$ cannot be true in this case. However, if we apply the reformulated version of the second law we note that both the ambient acceleration Aa and the measured acceleration Am are equal to g, the familiar gravitational acceleration. We then have $F = m(Am - Aa) = m(g - g) = 0$ in agreement with Einstein. Similarly, consider an object at rest upon the surface of the Earth. As any weight scale placed between the object and the surface will demonstrate, a force is clearly present yet the acceleration of the object relative to the Earth (and scale) is zero, and so $F = ma$ by itself cannot be true. The aforementioned caveat is required in order to explain this contradiction away. However, if we apply the reformulated second law we see that the measured acceleration $Am = 0$, and we have $F = m(0 - g) = -mg$ as required, and without any need for an additional caveat. In essence the caveat is required in conjunction with the original formulation of the laws because the laws do not internally take into consideration the inertial acceleration produced by the curvature of space-time. That is, the original formulation of the laws of motion by themselves, i.e., without any caveat, do not take into account that objects will always tend accelerate without any force being present. By including consideration of the ambient acceleration into the first law and by including the ambient acceleration term directly into the equation of the second law, the need for any additional caveat is eliminated. In this manner, the laws stand on their own and the process of analyzing the laws of motion in all scenarios is simplified in that the need for the superfluous concept of IRFs can be done away with entirely.