

Variational Solutions to Simultaneous Collisions Between Multiple Rigid Bodies

Vlad Seghete and Todd Murphey

Abstract—We present a method of resolving simultaneous collisions between multiple rigid bodies based on the least action principle. By using the generalized directional derivative of the action, we use that the solution is related to the outcomes of nearby trajectories that experience consecutive single impacts. We present an algorithm based on this result, and prove its effectiveness by applying it to several low dimensionality examples based on billiard ball interactions, including a simplified version of Newton’s cradle.

I. INTRODUCTION

When trying to simulate dynamical systems that undergo rigid body collisions, the problem of multiple simultaneous impacts is bound to occur. Whether the system starts in a relevant configuration—as is the case of a billiard ball break—or whether the situation arises from allowing plastic impacts—when bouncing a pen off a table, eventually both ends will be in contact at the same time—the problem can always be abstracted as a point in higher dimensional space impacting a corner, generally a point of non-differentiability on the contact manifold (while this might not be immediately obvious, section II-C provides examples that illustrate this fact, and fig. 4 in particular is meant to illustrate our claim). This provides for non-uniqueness of the result in such an interaction, especially when using set-valued force laws to look for a solution [1]. The problem disappears if one is willing to apply finite element methods that take into consideration the deformability of the objects engaged in the impact [2]. However, this approach also introduces large amounts of complexity to the problem, by including wave propagation effects and substantially adding to the dimensionality of the system. Other methods have been proposed, that still take elasticity into account, without resorting to finite element methods [3]. While lowering the complexity, such methods are nevertheless dependent on a careful choice of parameters to describe the bodies involved in the collision. Rephrasing the impact as a linear complementarity problem (LCP) is the best known method for solving contact problems to date [4], [5], [6], [7]. LCP based methods have, however, several major disadvantages, the main ones being the fact that they are not based on an underlying physical principle and that the solutions they provide are not unique. We believe the issue is solvable without having to drop the rigidity assumption—hence avoiding the needless complexity of finite element methods where such an approach is not needed—and that

a solution should ascribe to well established physical principles. Moreover, most of the methods currently used to model impacts—and in particular LCP based formulations—use a *simultaneous* approach, and solve for all potential impacts during a time step at once, whether or not they would occur. This gives great advantages in terms of speed and ease of implementation, but leaves many open questions when it comes to physical accuracy of the simulation. As we are interested foremost in a physically meaningful solution to the problem, we chose to go the alternative route, and use a *propagative* approach for dealing with multiple impacts, which raises more implementation issues, but allows more freedom in dealing with the phenomenon [8].

In the light of the above discussion, we set out to develop a method of dealing with such situations, that should give physically meaningful—and as close to correct as the rigid body assumption allows—results. Starting from a variational principle, we show how one collision is resolved, and point out the ambiguity that arises when two collisions occur at once. However, small perturbations of the initial trajectory of the system allows us to solve for the collisions “one at a time”. We notice that, under such perturbations of the trajectory (and thus of the action integral), some systems evolve in a unique manner after the collision, independent of the direction of the perturbation. This is the case, for example, in Newton’s cradle, and other similar mechanical systems. Unfortunately, this uniqueness does not generalize, as we will later discuss.

To help us deal with this new kind of non-uniqueness, we refer to a problem originally posed (and partly solved) by Truesdell [9]. In this problem a point hits a convex wedge precisely in the corner, as sketched in fig.1.

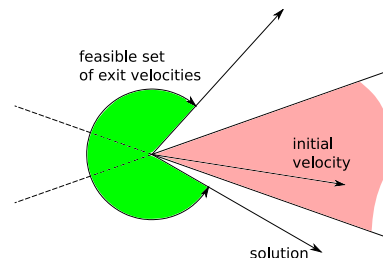


Fig. 1. A point hits a corner. All velocities in the green cone are feasible, as restricted by the way in which the surface interacts with the point. However, arguments exist for choosing the outcome closest to the initial velocity.

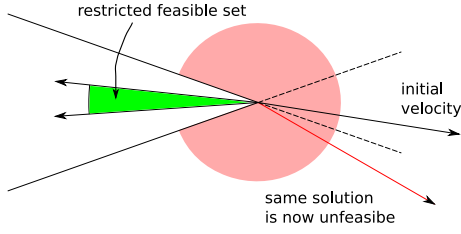


Fig. 2. Modified version of fig.1, in which the corner is concave as opposed to convex. The previous feasible set of exit velocities is too inclusive in this case and a new definition is needed in order to find a correct result.

problem is re-stated and formalized in [10]. The momentum change is restricted to the normal cone to the surface at the point of impact, as defined through non-smooth analysis concepts (see [11]), and a set of feasible exit velocities is created. The predicted velocity is then "projected" by the kinetic energy metric onto this set. This technique is highly attractive, as it also stems from a variational principle and provides an algorithmic way of choosing between multiple valid results.

However, in most of the problems we encounter, as we shall see, the point is hitting the inside corner of a concave boundary, as in fig. 2. The normal cone in the two problems is the same, and thus the interaction with the two surfaces is expected to be similar. However, in this case, the condition that the change of momentum lie in the normal cone is not enough, and a naive application of the algorithmic method will invariably result in an exit velocity that points straight at the walls—while the same solution would have been perfectly feasible in the case of the convex wedge. In order to solve this issue, we take a step back and redefine what the feasible exit velocity set is, by using the perturbation method discussed earlier. To do this, we perturb the system in all possible directions and construct the cone defined by the results of these perturbations. This generalizes the definition of the feasible set onto which we project the predicted velocity, and makes it more restrictive. Notice that the feasible set does not change at all when looking at the system in fig.1. Also, in some systems, the feasible set includes only one element. This is the case for Newton's cradle, and accounts for the uniqueness of the result in this case.

It might not seem physically intuitive that a mechanical system with clear initial conditions can have multiple, equally acceptable outcomes. Clearly, real life does not suffer from such non-uniqueness. However, it is important not to lose track that we are dealing with a model, specifically a model in which all bodies are perfectly rigid. With these assumptions the problem of multiple instantaneous impacts becomes, in a sense, ill-posed [12]. When discussing this issue, Ivanov [13] comes to the conclusion that, even in real life, the outcome of certain multiple impact scenarios is highly dependent on perturbations, either in the system parameters or in the initial conditions. For such cases, he proposes a stochastic model, in which one of the possible outcomes is chosen through a "coin toss." This illustrates the

fact that there are two distinct steps that need to be addressed: first finding the set of possible outcomes and then making a decision based on this set.

The bulk of this paper deals with solving for the feasible set from initial perturbations. Section II describes how to solve for one impact, where the ambiguity comes in when trying to solve for two simultaneous impacts, and briefly discusses two low dimensional canonical examples and their expected outcomes. Section III reformulates initial perturbation directions as impact orderings. This allows us to find the set of all possible outcomes by iteratively solving for impacts with only one surface, until a valid exit velocity is found. For the case of two impact boundaries proof is presented that guarantees a valid solution after at most a given number of iterations, which is inversely proportional to the angle between the two surfaces. Finally, the algorithm is applied to the two examples, showing how the uniqueness of the results is related to the value of the mass matrix. The conclusion summarizes our findings and opens the discussion on the effectiveness of our method when dealing with plastic and inelastic impacts

II. THE PROBLEM

A. A Single Impact

The equations governing a single collision at time t_A are:

$$\frac{\partial L}{\partial \dot{q}} \Big|_{t_A^-} + \lambda_A D\phi_A(q_A) = 0, \quad (1a)$$

$$\left[\frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L \right]_{t_A^-}^{t_A^+} = 0, \quad (1b)$$

where $q_A = q(t_A)$ is the configuration at time of impact, ϕ_A is the function describing the impact surface and λ_A is a Lagrange multiplier. The variables are $\dot{q}(t_A^+)$ and λ_A , and we assume we know $\dot{q}(t_A^-)$ and q_A . From here on, we will work under the assumption that we are dealing with a simple mechanical system: potentials will not depend on the the velocity, \dot{q} . Under this assumptions, we can write the Lagrangian as:

$$L(q, \dot{q}) = \frac{1}{2} \dot{q}^T M(q) \dot{q} - V(q),$$

where we can think of M as being a mass matrix or, alternatively,

$$M(q) = \partial_{\dot{q}\dot{q}} L(q, \dot{q}).$$

It is true, in general, that $M(q)$ will not depend on \dot{q} (hence the notation) and that it is positive definite ($x^T M x > 0$ for all x). We also assume that coordinates were chosen to avoid any degeneracy, and as such $M(q)$ will be invertible. Under these assumptions, (1) becomes

$$\begin{aligned} \dot{q}_{A^+}^T M - \dot{q}_{A^-}^T M + \lambda_A D\phi_A &= 0, \\ \dot{q}_{A^+}^T M \dot{q}_{A^+} - \dot{q}_{A^-}^T M \dot{q}_{A^-} &= 0, \end{aligned}$$

where, for ease of notation, we replaced $M(q_A) = M$ and $\phi_A(q_A) = \phi_A$, as these are constants for any given impact configuration. We can rewrite the first equation as

$$\dot{q}_{A^+} = \dot{q}_{A^-} - \lambda_A M^{-1} D\phi_A^T,$$

which, when plugged into the second equation, gives

$$\lambda_A [\lambda_A (D\phi_A M^{-1} D\phi_A^T) - (D\phi_A \dot{q}_{A-} + \dot{q}_{A-}^T D\phi_A^T)] = 0.$$

The solution $\lambda_A = 0$ is trivial and gives $\dot{q}_{A+} = \dot{q}_{A-}$, which implies no change occurred through the collision. We know this to always be false, hence we will discard this solution. We get, finally

$$\lambda_A = \frac{2D\phi_A \dot{q}_{A-}}{D\phi_A M^{-1} D\phi_A^T},$$

$$\dot{q}_{A+} = \left[I - 2 \frac{M^{-1} D\phi_A^T D\phi_A}{D\phi_A M^{-1} D\phi_A^T} \right] \dot{q}_{A-}.$$

Considering the previous derivation, define a mapping G_u thusly

$$G_u = I - 2 \frac{M^{-1} u^T u}{u M^{-1} u^T}$$

for some covector u . Then, we can write the generalized velocity of a system after a collision as a linear mapping of the velocity before the collision

$$\dot{q}_{A+} = G_{D\phi_A} \dot{q}_{A-}.$$

B. Two Simultaneous Impacts

Suppose now that the impact occurs across two surfaces at the exact same time. Each of the surfaces can act in its normal direction with arbitrary magnitude. As such, (1) becomes

$$\frac{\partial L}{\partial \dot{q}} \Big|_{t_A^-} + \lambda_A D\phi_A(q_A) + \lambda_B D\phi_B(q_A) = 0, \quad (2a)$$

$$\left[\frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L \right]_{t_A^-}^+ = 0, \quad (2b)$$

which has the same number of equations as (1), but one extra variable, λ_B . As a consequence, the system described by (2) will have a continuum of solutions instead of the two we found before. In most cases (see section II-C.2), one cannot solve the problem by using conservation of momentum and energy alone, to get a unique answer. However, we know full well how this system behaves in real life. We expect the rigid body approximation to behave, if not identical, at least qualitatively close to the real system. There are many methods through which the problem is solved to give "real" results. However, most of them start by dropping the rigid body assumption, and thus avoiding, rather than solving the issue. The one method that gives correct results, without dropping the rigid body assumptions, is looking at slightly perturbed versions of the system, which corresponds to introducing infinitesimal gaps between the balls. This allows us to solve for collisions one at a time, by solving (1) repeatedly, as needed, until the collisions are over. This approach, while it gives a unique answer, has problems of its own. The first question is that of termination, and whether we can always find an answer. We show later that this is indeed the case, and that a point bouncing inside a corner will always come out eventually. More important yet is an issue related to the order of impacts. The effect of the initial

perturbations is to give a clear order of the impacts. It follows, then, that one can choose different perturbations to result in a different order of impacts. The question arises whether these different impact orders will ultimately give the same answer or not. It turns out, as we will show, that non-uniqueness does appear, although the maximum number of distinct answers is not only finite, but predictable.

C. Example Systems

In this section we present two low dimensional example systems, their purpose being to illustrate the issue of multiple impacts as seen in configuration space, and to build intuition about higher dimensional systems for which drawings cannot be as readily produced.

1) *Two billiard balls against a wall*: The first system we are going to discuss is comprised of two billiard balls arranged in a line perpendicular to a wall (the edge of the table). In order to simplify the problem, we restrict their movement to the line connecting their centers and we ignore any sort of dissipative phenomena, including friction. We also assume the balls are not spinning, so we can safely ignore issues related to the moment of inertia. The two configuration variables are $q = [x_1, x_2]^T$, as shown in fig.3.

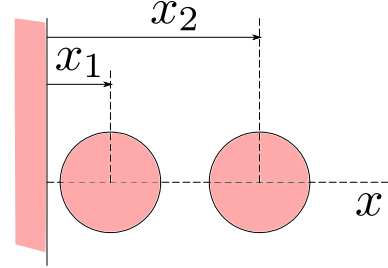


Fig. 3. Two billiard balls sitting in a line, one of them against a wall. The configuration variables are x_1 and x_2 . The radius of the balls (not shown) is r .

Figure 4 shows the configuration space of this system, including the two boundaries (between the first ball and the wall and between the two balls). If the system starts in the green area and the balls are not allowed to overlap at any time, it follows that it will lie in the green area for all time.

A straightforward way of describing the two surfaces of impact is by defining proximity functions thusly

$$\phi_u(q) = x_1 - r,$$

$$\phi_v(q) = x_2 - x_1 - 2r.$$

A configuration q is allowable (does not violate the constraints) if *all* proximity functions at that configuration are positive. Also note that each of these functions defines a surface at $\phi(q) = 0$, and that $\nabla\phi(q)$ is a vector normal to the surface at the point q . In our example, the two surfaces are lines, and the normals defining them are

$$\nabla\phi_u(q) = [1, 0]^T$$

$$\nabla\phi_v(q) = [-1, 1]^T.$$

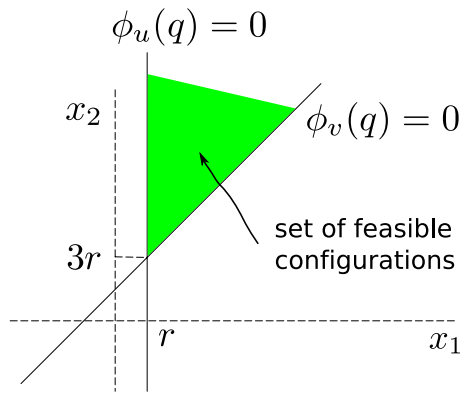


Fig. 4. The configuration space of the system shown in fig. 3. The points in the shaded area are feasible, while all others indicate that an overlap has occurred. The boundary is differentiable everywhere except at the corner, which is where we will focus our attention.

The configuration of the system can be described by a point in the two dimensional space. A collision occurs when the point "impacts" one of the surfaces, described by either $\phi_u(q) = 0$ or by $\phi_v(q) = 0$. The *double* collision occurs when the point hits the corner, making *both* $\phi_u(q) = \phi_v(q) = 0$. In general, an n -dimensional system can be described as a point in the n -dimensional configuration space, while boundaries will be $(n-1)$ -dimensional surfaces in the same space. A multiple collision will occur when the point describing the system hits a corner made by two, three or more surfaces. However, in this present paper, we will focus our discussion mainly on two surface corners, leaving the generalization to three and more surfaces to future work.

2) *Three ball Newton's cradle*: Figure 5 shows a second example, of dimensionality three, but still with only two boundary surfaces. One can think of it as the previous example with the third ball replacing the wall. This example also serves as a simplification of Newton's cradle, which gives us the benefit of knowing many real outcomes of the collision.

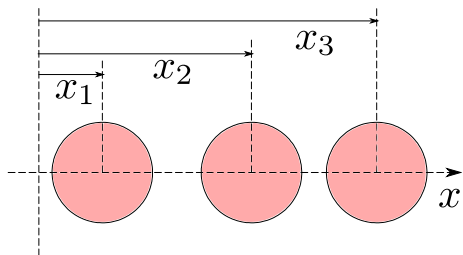


Fig. 5. Three balls restricted to move on a line. The configuration space is three dimensional, with x_1 , x_2 and x_3 as the configuration variables.

As before, we have

$$\begin{aligned}\phi_u(q) &= x_2 - x_1 - 2r, \\ \phi_v(q) &= x_3 - x_2 - 2r, \\ \nabla\phi_u(q) &= [-1, 1, 0]^T, \\ \nabla\phi_v(q) &= [0, -1, 1]^T.\end{aligned}$$

Notice that, in this case, the configuration space is three dimensional and that the boundaries are planes separating the space, just as before, in four parts.

III. THE SOLUTION

Previous works, like [10], made use of nonsmooth analysis concepts and re-posed the problem of corner impacts as a minimization problem over the feasible solutions. The term feasible is meant to indicate that the change in energy is zero, while the change of momentum lies in the normal cone at the point of impact. The technique was used to solve ambiguities in the case of a point hitting the corner of a convex wedge. However, the problem is a little more complex when dealing with concave corners, as is the case in both of our examples. The reason for this is that, solutions that are feasible in the sense just described, are not necessarily valid as discussed in the previous section. In fact, if one were to apply the algorithmic approach described in [10] and project the predicted velocity onto the feasible set, one would *always* get an invalid solution when dealing with a concave boundary. A new definition of the feasible set is in order. Such a definition would have to coincide with the feasible set in the convex wedge case, but differ from it such that it gives correct results in the case of concave corners.

A. A more general definition of the feasible set

In order to derive (1) for the simple impact scenario, we used a variational principle, by which we extremize the action

$$A(q(t)) = \int L(q(t), \dot{q}(t)) dt.$$

If the action is smooth with respect to q (as is the case for one impact), then one can simply take the variation in the action

$$\begin{aligned}\delta A &= \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \int L(q + \varepsilon \delta q, \dot{q} + \varepsilon \delta \dot{q}) dt \\ &= \lim_{\varepsilon \rightarrow 0} \frac{A(q + \varepsilon \delta q) - A(q)}{\varepsilon},\end{aligned}$$

and require that it is zero. However, the above only works in a smooth setting. A more general approach is to look at the *generalized directional derivative* of the action, $A^\circ(q, \delta q)$

$$A^\circ(q, \delta q) = \limsup_{\substack{\tilde{q} \rightarrow q \\ \varepsilon \downarrow 0}} \frac{A(\tilde{q} + \varepsilon \delta \tilde{q}) - A(\tilde{q})}{\varepsilon}, \quad (3)$$

and require that this be zero. Here q is the trajectory in which the simultaneous collisions occur—a point hits the corner—while \tilde{q} are trajectories close to it in which impacts occur one at a time. Thus, in order to evaluate (3) we will have to consider all close-by trajectories \tilde{q} for which the limit exists (i.e. the action is smooth).

We use this observation to define the feasible set of solutions thusly: the convex hull of all possible solutions under infinitesimal perturbations of the impact point that result in a series of strictly single impacts, intersected with the feasible set in the sense that the change in momentum is restricted to the normal cone at the point of impact. As we

shall see in the next section, there is always a finite number of solutions used in construction the first part of the set—in the case of two boundaries there are no less than one and no more than two.

With the above definition, the algorithm we are proposing to solve the impact scenario can be summarized:

- 1) find the trivial solution, with $\lambda_i = 0, \forall i$
- 2) find a minimal set of initial infinitesimal perturbations that give unique solutions
- 3) solve for the exit velocity under each of these perturbations
- 4) construct the feasible solution set through the convex hull of all the solutions
- 5) project the trivial solution onto the set of feasible solutions using the Riemannian metric

As we mentioned before, in the case of two boundaries, there can be at most two distinct solutions, each corresponding to a distinct initial perturbation. Basically, one corresponds to the point hitting surface u first, while the other corresponds to the point hitting surface v first. Moreover, in special cases, the two solutions are identical and this solution is the whole of the feasible set. For instance, in both our examples, if the masses of the billiards are identical, there is only one element in the feasible set, as we shall see shortly.

B. Finding the feasible solutions

The hardest part of the above algorithm is finding the feasible solutions and making sure they are all the possible ones. In order to do this, we consider the case of a system with n boundaries defined by covectors $u_i, i = 1 \dots n$. Let the initial velocity be v_0 . Like the examples show, we can think of the problem of simultaneous impact across all the n boundaries as a point hitting a corner. We consider all perturbation directions that avoid corner impacts. Thus, the first collision occurs across one of n surfaces. This gives us n branches of the algorithm. Each branch gives one exit velocity, call this $v_1 = G_{u_1} v_0$. We then check v_1 for validity against all u_i (by taking the Euclidian dot product and checking its sign). Each surface for which v_1 is invalid adds a branch, and since v_1 *must* be valid as far as the first surface we chose is concerned, there are a maximum of $n - 1$ branches at this point. Proceeding with the algorithm, we solve for $v_2 = G_{u_2} v_1$ and continue inductively, until we find a v_N that is valid with respect to all impact boundaries. At each step we will be able to choose from at most $n - 1$ surfaces, which makes the maximum number of feasible solutions be $n(n - 1)^{N-1}$. A proof that sets an upper bound for the number of iterations N needed to get a valid solution in the case of two boundaries follows in the next section. Before that, however, we will apply this method to our examples, in an effort to make it clearer.

1) *Two balls of equal mass*: The example system was described in sec. II-C.1. Let the covectors corresponding to the surface boundaries be $u = [1, 0]$ and $v = [-1, 1]$. Assume, for now, that the two balls have identical mass. This makes the mass matrix equal to identity. The corresponding

transformations are

$$G_u = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad G_v = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We assume an initial velocity v_0 such that both $u v_0 \leq 0$ and $v v_0 \leq 0$. Since both constraints are violated we have a choice between applying G_u or applying G_v . Suppose we apply G_u , and we get $v_1 = G_u v_0$. Then we will have $v v_1 \leq 0$, but $u v_1 \geq 0$. This leaves us no choice as to which transformation to apply next (G_v), and so, we will find that after four iterations, we get

$$v_4 = G_v G_u G_v G_u v_0 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} v_0 = -v_0,$$

which is clearly a valid solution (one can easily show that the previous three outcomes are not valid). Notice that we only chose which transformation to apply once—literally choosing the first impact surface—and at all the other iterations we had no reasonable choice to make. However, in order to find all possible outcomes we need to cover all possible choices. It turns out, however, that in this case we have

$$G_v G_u G_v G_u = G_u G_v G_u G_v = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

which means that, for this system, all possible outcomes are identical and exactly the opposite of the entering velocity. For example, if the two balls are moving together towards the wall, they will bounce together from the wall. If one is sitting touching the wall and the other hits it, all the momentum will go into the moving ball. These outcomes are exactly what would occur if one were to add small gaps in between the balls before solving the problem with standard techniques.

2) *Newton's cradle with three balls*: The only major difference between this and the previously described system is the dimensionality. We also know, for this example, the expected behavior in several select cases. We will present the results of our algorithm for these cases, rather than focus on the details on how we obtained them, hoping that the interested reader will be able to follow the step described in the previous sections to reproduce our results.

First, let $u = [-1, 1, 0]$ and $v = [0, -1, 1]$. We assume the masses are all equal, which gives

$$G_u = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad G_v = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

It will take three iterations to solve this problem, and we will have that

$$v_f = G_u G_v G_u v_0 = G_v G_u G_v v_0 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} v_0. \quad (4)$$

Just as before, we get the same final result regardless of the order in which we solve for the impacts. Let us apply this result to a few initial velocities. For example, let us take $v_0 = [1, 0, 0]^T$, a case in which the first ball is moving and the last two are stationary. The result we obtain, as by means of (4), is $v_f = [0, 0, 1]^T$, which corresponds to the first

two balls being stationary and the third moving away from the two. Again, let us take $\mathbf{v}_0 = [1, 1, 0]^T$, in which the first two balls are moving together and striking the third. The result is, as expected, $\mathbf{v}_f = [0, 1, 1]^T$, in which the second and third ball are moving away from the first ball, which stays still after the impact. Both these examples represent very good approximations of the real system, such that the difference in behavior between the ideal and the real system can only be observed with the use of high speed camera systems, differences that can be easily attributed to elastic effects neglected in the rigid body approximation we are working under.

3) *Two balls of slightly different mass*: The previous two examples showed cases in which the results are unique regardless of the order in which impacts are considered. However, imagine that in the first example system, one of the balls is slightly bigger than the other, making the mass matrix be different:

$$M = \begin{bmatrix} 1.1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We expect the results to be different, although not drastically so. This is true in some sense, but not in others, as we shall soon see. The new transformation matrices are

$$G_u = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad G_v = \begin{bmatrix} 0.0476 & 0.9524 \\ 1.0476 & -0.0476 \end{bmatrix}.$$

What this tells us is that the interaction between the ball and the wall is going to be unchanged, but the interaction between the two balls is going to be slightly different. The choices are two, just as before, as we can only choose the first surface of impact. We thus have two possible outcomes

$$\begin{aligned} \mathbf{v}_{4a} &= G_v G_u G_v G_u = \begin{bmatrix} -0.9955 & -0.0907 \\ 0.0998 & -0.9955 \end{bmatrix} \mathbf{v}_0, \\ \mathbf{v}_{4b} &= G_u G_v G_u G_v = \begin{bmatrix} -0.9955 & 0.0907 \\ -0.0998 & -0.9955 \end{bmatrix} \mathbf{v}_0. \end{aligned}$$

It is clear that the two exit velocities are going to be close to $-\mathbf{v}_0$ but will not be identical to each other. As a result, the set of feasible velocities will be

$$\mathbb{V} = \{ \mathbf{v} \in TQ, \mathbf{v} = a\mathbf{v}_{4a} + b\mathbf{v}_{4b}, a, b \in \mathbb{R}_+ \}$$

Projecting \mathbf{v}_0 onto this set with respect to the Riemannian metric is the same as finding $\mathbf{v}_f \in \mathbb{V}$ such that $\|\mathbf{v}_f - \mathbf{v}_0\|_M$ is minimized. Invariably, in our example, this will be either \mathbf{v}_{4a} or \mathbf{v}_{4b} (although this might not be the case in more complicated systems).

C. Proof of termination

Here we show that, for two surfaces, solving successive impacts is guaranteed to give a solution. While we have no proof for a higher number of surfaces, we strongly believe that this result generalizes and that such a proof is close at hand. Also, notice that our arguments are only restricted to two surfaces, but *not* to two dimensional configuration spaces. Indeed, the following proof applies to systems of any dimensionality, as long as the impact occurs only across two surfaces.

The mapping G_u , as defined in II-A, has several properties that will be useful later in this section. Firstly, one can easily show that $G_u^2 = I$. This simply means that resolving an impact across a surface twice in a row gets one back to the original velocity, as expected. Also, notice that G is not dependent on the magnitude of u , only on its direction, as $G_{\alpha u} = G_u$.

It is very easy to show that G_u , as defined above, preserves the Riemannian metric on the tangent space. One can easily check that

$$G_u^T M G_u = M, \quad (5)$$

which means that a G_u transformation will preserve the energy of the system. These mappings are also area preserving. This can be shown by looking at the value of the determinant of the matrix corresponding to G_u :

$$\begin{aligned} \det(G_u) &= \det\left(I - 2\frac{M^{-1}u^T u}{uM^{-1}u^T}\right) \\ &= \frac{1}{\alpha} \det(\alpha I + M^{-1}u^T u) \\ &= \frac{1}{\alpha} \det(M^{-1}) \det(\alpha M + u^T u) \\ &= \frac{1}{\alpha \det(M)} (1 + u(\alpha M)^{-1}u^T) \det(\alpha M) \\ &= \frac{1}{\alpha \det(M)} \left(1 + \frac{uM^{-1}u^T}{\alpha}\right) \alpha \det(M) \\ &= 1 - 2\frac{uM^{-1}u^T}{uM^{-1}u^T} = -1, \end{aligned}$$

where we used the notation $\alpha = -(uM^{-1}u^T)/2$, the fact that

$$\det(A + vw^T) = (1 + w^T A^{-1}v) \det(A),$$

and several other well known determinant properties. Thus we've shown that $\det(G_u) = -1$, which makes it an area preserving mapping from and onto a sphere associated with the Riemannian metric. This also confirms our interpretation of G_u as a reflection transformation, under the Riemannian metric.

Keeping the previous facts in mind, we want to show that, given any velocity $\dot{q} = \mathbf{v}_0$ lying on a Riemannian sphere, and two boundaries locally described by u and v , a finite number of successive applications of G_u and G_v will transform \mathbf{v}_0 into a valid velocity \mathbf{v}_f , so that

$$u^T \cdot \mathbf{v}_f \geq 0, \quad (6a)$$

$$v^T \cdot \mathbf{v}_f \geq 0, \quad (6b)$$

where the dot represent the Euclidian dot product. The interpretation of 6 is that the velocity is pointing away from both the boundaries described by u and v , and thus the system is moving out of the corner. For ease, let the set of velocities that lie on the Riemannian sphere and that satisfy (6) be called \mathbb{S}_0 :

$$\mathbb{S}_0 = \{ \mathbf{v}_f \in TQ, \|\mathbf{v}_f\|_M = \|\mathbf{v}_0\|_M, u^T \cdot \mathbf{v}_f \geq 0, v^T \cdot \mathbf{v}_f \geq 0 \}.$$

To show this, we make use of the fact that reflections are invertible and that their inverse is the reflection itself. Thus, if \mathbf{v}_0 can be obtained from \mathbf{v}_f through a series of transforms, then \mathbf{v}_f can be obtained from \mathbf{v}_0 through the same series of transforms applied in the opposite order. Thus, if we can show that through of successive transformations on \mathbb{S}_0 we can cover the whole sphere, then we will have also shown

that any point on the sphere can be transformed to some point in \mathbb{S}_0 . In order to do this, we will use an argument

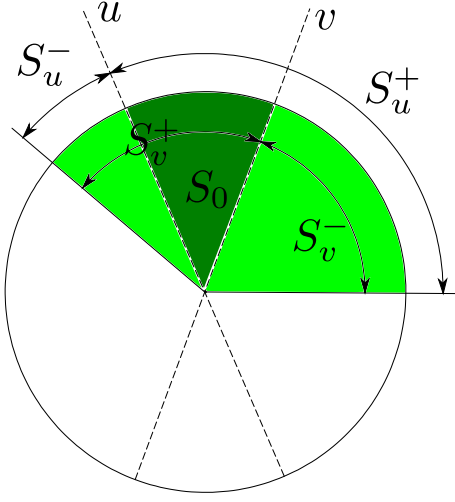


Fig. 6. Sketch of an area \mathbb{S} on the Riemannian metric sphere (shaded area) and several subsets of it, as used in the proof in this section. Here \mathbb{S} is depicted after a reflection across v , hence $S_v^+ = S_v^-$, but $S_u^+ \geq S_u^-$. S_0 is the dark shaded area between the planes defined by u and v .

based on area. The notation S will stand for "the area of the set \mathbb{S} ." Furthermore, let \mathbb{S}_{total} be the set of all velocities on the sphere, and let $\mathbb{S} \subset \mathbb{S}_{total}$ be a subset of the sphere. We can always partition \mathbb{S} into two subsets, \mathbb{S}_u^+ and \mathbb{S}_u^- , such that

$$\begin{aligned} \mathbb{S}_u^+ &= \{v \in \mathbb{S}, u^T \cdot v \geq 0\}, \\ \mathbb{S}_u^- &= \{v \in \mathbb{S}, u^T \cdot v \leq 0\}. \end{aligned}$$

Now, assume that $v \in \mathbb{S}_{total}$ and $u^T \cdot v = \alpha$. Applying G_u to v and calculating the Euclidian cross product we get

$$u^T \cdot (G_u v) = u^T \cdot v - 2 \frac{u M^{-1} u^T u}{u M^{-1} u^T} v = -\alpha.$$

Unexpectedly, the transform reflects points across the boundary. Now, let \mathbb{S} be a connected subset of \mathbb{S}_{total} . We apply the G_u transformation to \mathbb{S} and join the resulting set with \mathbb{S} , we get

$$\begin{aligned} \mathbb{S}_u &= \mathbb{S} \cup (G_u \mathbb{S}), \\ S_u &= 2 \max\{S_u^+, S_u^-\}. \end{aligned}$$

Moreover, \mathbb{S}_u becomes symmetric about u , such that if $v \in \mathbb{S}_u$ then $G_u v \in \mathbb{S}_u$ as well, and subsequent applications of G_u will not change this set. In order to cover the whole sphere, we start with \mathbb{S}_0 and apply G_u and G_v sequentially, joining the resulting sets into a "reachable" set. Let us assume we have done this several times, and that G_v was the last transformation we applied. The reachable set at this point is, as we have just shown, symmetric with respect to v . This also means that $S_v^+ = S_v^- = S_v = \frac{1}{2} S$. However, we can also write S in terms of its u -defined subsets, $S = S_u^+ + S_u^-$. If either of the two parts is larger than half the sphere, then applying G_u will end up covering the whole sphere. If not, we have

$$\begin{aligned} S_u^+ &= S_v + S_0, \\ S_u^- &= S_v - S_0, \end{aligned}$$

where, without loss of generality, we assumed $S_u^+ \geq S_u^-$. Figure 6 should help with seeing these relations.

Applying G_u , as before, we get that the new \mathbb{S} is such that

$$S_{new} = 2 \max\{S_u^+, S_u^-\} = 2S_v + 2S_0 = S + 2S_0,$$

which shows an increase of the reachable area by $2S_0$ on every iteration. This, of course, does not apply to the last iteration which adds just enough area to finish covering the sphere, and to the first iteration: no iteration can add more than the area it starts with, and since the first iterations starts with S_0 it can only add S_0 ; after this, the reachable area is at least as big as $2S_0$. This implies that it will take at most

$$N = \left\lceil \frac{S_{total}}{2S_0} \right\rceil$$

iterations to guarantee that the sphere is covered.

Going back to our original argument that transformations move points both ways, it will also take at most N iterations to transform any initial velocity v_0 into a velocity v_f that point outside of the corner. In other words, a point moving into a corner will bounce out after at most N interactions with the walls.

IV. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

We have presented a method of solving multiple simultaneous impacts, through investigating nearby trajectories that undergo successive single impacts with the surfaces in question. For some well known mechanical systems based on billiard ball interactions, we showed that our method gives a unique solution which coincides with the expected result in a hardware implementation of such systems. We also showed that uniqueness is not to be expected always, and we presented an argument to choose between the possible outcomes, an argument we based on applying non-smooth analysis concepts to a variational principle. We further showed that our algorithm is guaranteed to give a result when there are only two boundaries, and we found an upper bound for the number of steps required to get such a result.

B. Future Works

Clearly, a general termination proof is needed in order to show that our algorithm will still finish in more complicated scenarios. Also, an extension of this theory to inelastic and perfectly plastic impacts is needed in order to fully implement our methods toward simulating useful scenarios. In fact, the main problem we are trying to solve occurs very rarely when perfectly plastic impacts are not considered, but highly prevalent when one does allow them (see the example of a pen bouncing on a table in the introduction). If we consider perfectly plastic impacts, the problem becomes trivial. This is because the outcome of a plastic impact is characterized by a velocity tangent to every contact manifold and conservation of momentum in that direction. Clearly, this implies that the exit velocity will be along the intersection of the contact manifolds. For example, in the case of Newton's cradle, the consequence of a plastic impact will be all three

balls in contact with each other, moving in the same direction with the same velocity. The most useful and realistic scenario is when the energy lost in an impact follows a coefficient of restitution, unless it is under a certain threshold — in which case the impact is modeled as perfectly plastic. Clearly our termination result does not apply in this case, and a more thorough analysis would be required. However, we argue that the only way in which applying our algorithm to such a system would result in non-termination is if the system evolved in such a way that each successive collision happened at a smaller incident angle than the previous one — otherwise the coefficient of restitution would guarantee that the system eventually lost enough energy to go below the plastic threshold, and the solution would be trivial. We also believe that such a decrease in the angle of incidence would imply a finite number of surface reflections before a feasible exit velocity is found. Both of these assertions require rigorous mathematical proofs, which should be the topic of future works.

V. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under award CMMI 0951688. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] C. Glocker and U. Aeberhard, “The geometry of newton’s cradle,” *Advances In Mechanics And Mathematics*, vol. 12, p. 185, 2006.
- [2] C. Kane, E. Repetto, M. Ortiz, and J. Marsden, “Finite element analysis of nonsmooth contact,” *Computer methods in applied mechanics and engineering*, no. 180, pp. 1–26, Jan 1999.
- [3] Y.-B. Jia, M. T. Mason, and M. Erdmann, “A state transition diagram for simultaneous collisions with application in billiard shooting,” *Eighth International Workshop on Algorithmic Foundations of Robotics*, 2008.
- [4] D. Stewart and J. Trinkle, “An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and coulomb friction,” *Int. J. Numer. Meth. Engng.*, Jan 1996.
- [5] N. Chakraborty, S. Berard, S. Akella, and J. Trinkle, “An implicit time-stepping method for multibody systems with intermittent contact,” *Robotics Science and Systems*, Jan 2007.
- [6] M. Anitescu and F. Potra, “Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems,” *Nonlinear Dynamics*, Jan 1997.
- [7] J. Trinkle, J. Pang, S. Sudarsky, and G. Lo, “On dynamic multi-rigid-body contact problems with coulomb friction,” *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift fur Angewandte Mathematik und Mechanik*, vol. 77, no. 4, 1997.
- [8] D. Baraff, “Analytical methods for dynamic simulation of non-penetrating rigid bodies,” *Computer Graphics*, vol. 23, no. 3, pp. 223–232, July 1989.
- [9] C. Truesdell, “A simple example of an initial-value problem with any desired number of solutions,” *Istituto Lombardo (Rend. Sc.) A*, vol. 108, pp. 301–304, 1974.
- [10] A. Pandolfi, C. Kane, J. E. Marsden, and M. Ortiz, “Time-discretized variational formulation of non-smooth frictional contact,” *Int. J. Numer. Meth. Engng.*, vol. 53, no. 8, pp. 1801–1829, Mar 2002.
- [11] F. H. Clarke, *Optimization and Nonsmooth Analysis*. New York: John Wiley and Sons, 1983.
- [12] A. Chatterjee and A. Ruina, “A new algebraic rigid-body collision law based on impulse space considerations,” *Journal of Applied Mechanics*, vol. 65, no. 4, pp. 939–951, Dec 1998.
- [13] A. Ivanov, “On multiple impact,” *Journal of Applied Mathematics and Mechanics*, vol. 59, no. 6, pp. 887–902, Jan 1995.