
Figure 8: Visualising a deformable body $\Omega$ before (left) and during (right) impact into a rigid wall $\Gamma_R$. Here, $\boldsymbol{n}$ denotes the surface normal of the wall and $\Gamma_c$ the set of contact points.

# A Appendix: extended CD-Lagrange networks

The CD-Lagrange scheme, in full generality, is designed for deformable body impacts for finite element calculations (Fekak et al., 2017). In the non-deformable case, Fekak et al. (2017) propose employing Newton's restitution law for rigid body impacts. This variant has been benchmarked by Di Stasio et al. (2019). We extend the framework to support three-dimensional rigid body-body collisions by ensuring that the overall momentum is conserved during the impact. Instead of multi-node bodies—as used, for example, in finite element simulations of deformation during multi-body collision—we restrict ourselves to single-node bodies in order to simplify the problem.

CD-Lagrange is based on an asynchronous version of the central difference method and uses Lagrange multipliers $\lambda^{(k)}$ to enforce the contact constraints. For a given problem, $\boldsymbol{\lambda}$ must fulfill the Karush–Kuhn–Tucker (KKT) conditions, sometimes also referred to as Hertz–Signorini–Moreau (HSM) conditions (Wriggers and Panagiotopoulos, 1999). These are given by

$$g_N = (\boldsymbol{x}_M - \boldsymbol{x})^T \boldsymbol{n} \geq 0 \tag{22}$$

$$\boldsymbol{\lambda}_N = -\boldsymbol{n}_\Omega \cdot \boldsymbol{\sigma}(\boldsymbol{x}) \cdot \boldsymbol{n}_\Omega \geq 0 \tag{23}$$

$$g_N(\boldsymbol{x})\boldsymbol{\lambda}_N = 0 \tag{24}$$

which hold for all $\boldsymbol{x} \in \Gamma_c$, where $\boldsymbol{n}$ is the outer pointing normal on the wall $\Gamma_R$ and $\boldsymbol{n}_\Omega$ the outer-pointing normal on the deformable body $\Omega$. In particular, $\boldsymbol{n}_\Omega = -\boldsymbol{n}$ when a contact occurs. $\Gamma_c$ represents all contact points on $\Omega$, $\boldsymbol{x}_M \in \Gamma_R$ the closest point to $\boldsymbol{x} \in \Gamma_c$ and $\boldsymbol{\sigma}$ is the Cauchy stress field.

Condition (22) enforces the impenetrability, (23) implies that no adhesion occurs upon contact which means the contact-stress is only compressive, and (24) reduces the contact-states to CONTACT ($g_N = 0$ and $\lambda_N > 0$) and NO CONTACT ($g_N > 0$, $\lambda_N = 0$). The subscript $N$ is to emphasise dependence on the *normal* part of the velocity, i.e. the part perpendicular to the contact surface. This subscript will be omitted in the rest of this derivation to simplify notation.

We represent the position of all (single-node) bodies in a $d$-dimensional space by

$$\mathbf{Q} = \left(\boldsymbol{q}^k, k \in \{1, \ldots, K\}\right) \tag{25}$$

where $\mathbf{q}^k \in \mathbb{R}^d$ represents the individual position of body $k$.

To be able to calculate the velocity component each body has towards the impact surface a projection operator

$$\mathbf{L} = \left((\boldsymbol{n}^k), k \in \{1, \ldots, K\}\right) \tag{26}$$

is introduced, where $\boldsymbol{n}^k \in \mathbb{R}^d$ is the surface normal of the object body $k$ is colliding with. The distance between each body and the closest point on the wall or another body is represented by

$$\boldsymbol{g} = \left(g^k, k \in \{1, \ldots, K\}\right), \quad \text{where} \quad g^k(\mathbf{Q}) \in \mathbb{R}. \tag{27}$$

Finally, define $\boldsymbol{c} = \operatorname{sgn}(\boldsymbol{g})$ to be the *contact function*. Note that the equations of motion depend only on $\boldsymbol{g}$ through $\boldsymbol{c}$: we will subsequently replace this function with the contact network $\hat{\boldsymbol{c}}_\theta \in \{0, 1\}$.

**Discrete impact equations.** Given a single impact event at time $t_c$ of a hyperelastic solid, we may write the action integral (Cirak and West, 2005)

$$S(\mathbf{Q}, \dot{\mathbf{Q}}, t_c) = \int_0^{t_c} L(\mathbf{Q}, \dot{\mathbf{Q}}) \mathrm{d}t + \int_{t_c}^{T} L(\mathbf{Q}, \dot{\mathbf{Q}}) \mathrm{d}t \tag{28}$$

where $L(\mathbf{Q}, \dot{\mathbf{Q}})$ is the semi-discrete Lagrangian of the form

$$L(\mathbf{Q}, \dot{\mathbf{Q}}) = \dot{\mathbf{Q}}^T \mathbf{M} \dot{\mathbf{Q}} + V(\mathbf{Q}), \tag{29}$$

where $\mathbf{M}$ represents the mass matrix and $V$ the external potential, e.g. the gravitational field (Cirak and West, 2005). If there are multiple impacts over the interval $[0, T]$ the action integral can be split up further to handle the contact events in a sequential manner. We apply small variation of the path $\delta q$

$$\delta S(\mathbf{Q}, \dot{\mathbf{Q}}, t_c) = \delta \left( \int_0^{t_c} L(\mathbf{Q}, \dot{\mathbf{Q}}) \mathrm{d}t + \int_{t_c}^{T} L(\mathbf{Q}, \dot{\mathbf{Q}}, t_c) \mathrm{d}t \right) \tag{30}$$

$$= \int_0^T \left( \frac{\partial L}{\partial \mathbf{Q}} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{\mathbf{Q}}} \right) \delta \mathbf{Q} \mathrm{d}t - \left[ \frac{\partial L}{\partial \dot{\mathbf{Q}}} \cdot \delta \mathbf{Q} + L \delta t \right]_{t_c^-}^{t_c^+} = 0. \tag{31}$$

Applying the Fundamental Lemma of the Calculus of Variations (Cirak and West, 2005) yields the equations

$$\ddot{\mathbf{Q}} = \mathbf{M}^{-1} \frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}} \tag{32}$$

$$\left[ \mathbf{M} \dot{\mathbf{Q}} \right]_{t_c^-}^{t_c^+} = \mathbf{I}(\boldsymbol{\lambda}(t_c)) \tag{33}$$

$$\left[ \left( \mathbf{M} \dot{\mathbf{Q}} \right)^T \mathbf{M}^{-1} \left( \mathbf{M} \dot{\mathbf{Q}} \right) \right]_{t_c^-}^{t_c^+} = 0 \tag{34}$$

which fully describe the dynamics. Here (32) represents the smooth dynamics, (33) represents the non-smooth dynamics and (34) represents the kinetic energy balance upon impact.

Using the theory of non-smooth dynamics (Moreau, 1988), (32), (33), and (34) can be combined into one equation and the velocity can be expressed as a sum of the smooth and non-smooth parts, given by

$$\mathrm{d}\dot{\mathbf{Q}} = \mathrm{d}\dot{\mathbf{Q}}_{\mathrm{s}} + \mathrm{d}\dot{\mathbf{Q}}_{\mathrm{ns}}. \tag{35}$$

Fekak et al. (2017) show that incorporating the KKT conditions into the above yields the *non-smooth contact dynamics* (NSCD) equations

$$\mathbf{M} \mathrm{d}\dot{\mathbf{Q}} = \frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}} + \mathrm{d}\mathbf{I}(\boldsymbol{\lambda}) \tag{36}$$

$$\forall k \in \{1, \ldots, K\} \quad \begin{cases} 0 \leq \lambda^k \perp v^k \geq 0, & \text{if } g^k = 0 \\ \lambda^k = 0, & \text{if } g^k > 0. \end{cases} \tag{37}$$

By integrating (37) between $t_{n+\frac{1}{2}}$ and $t_{n+\frac{3}{2}}$ and estimating the time integrals using a midpoint-rule, following Fekak et al. (2017), one obtains the semi-discrete equilibrium equation

$$\mathbf{M} \left( \dot{\mathbf{Q}}_{n+\frac{3}{2}} - \dot{\mathbf{Q}}_{n+\frac{1}{2}} \right) = h \frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}} + \mathbf{I}(\boldsymbol{\lambda}_{n+\frac{3}{2}}), \tag{38}$$

which results in an asynchronous update rule for the velocity given by

$$\dot{\mathbf{Q}}_{n+\frac{3}{2}} = \dot{\mathbf{Q}}^S_{n+\frac{3}{2}} + \mathbf{M}^{-1} \mathbf{I}(\boldsymbol{\lambda}_{n+\frac{3}{2}}), \qquad \dot{\mathbf{Q}}^S_{n+\frac{3}{2}} = \dot{\mathbf{Q}}_{n+\frac{1}{2}} + \mathbf{M}^{-1} h \frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}}. \tag{39}$$

Further, $\mathbf{Q}$ can be updated using the previous velocity

$$\mathbf{Q}_{n+1} = \mathbf{Q}_n + \frac{h}{2} \dot{\mathbf{Q}}_{n+\frac{1}{2}} \tag{40}$$

To calculate the impulse $\mathbf{I}$ we combine Newton's restitution law, as proposed by (Fekak et al., 2017), and the law of conservation of linear momentum. Consequently, the impulse between body $i$ and $j$ caused by a collision is defined by

$$\boldsymbol{I}_{ij} = (1 + e) \frac{m_i m_j}{m_i + m_j} \boldsymbol{n}_i \left( \dot{\boldsymbol{q}}^j - \dot{\boldsymbol{q}}^i \right) \boldsymbol{n}_i. \tag{41}$$

Unpacking this equation, the velocity difference between the two bodies is projected onto the normal of the contact surface $n$ and scaled by the mass ratio $m_i m_j/(m_i+m_j)$ and the elasticity $e$.

We introduce the matrix $\mathbf{A}$ which, at impact times, determines which bodies are in contact. This is given by

$$[\mathbf{A}_n]_{ij} = \begin{cases} -1 & \text{if } i=j \text{ and } [\boldsymbol{g}_n]_i < 0 \\ 1 & \text{if body } i \text{ and } j \text{ are in contact} \\ & \text{(implicitly } [\boldsymbol{g}_n]_i = [\boldsymbol{g}_n]_j < 0) \\ 0 & \text{otherwise.} \end{cases}, \quad \sum_{i=1}^{K}\sum_{j=1}^{K}[\mathbf{A}_n]_{ij} = 0. \tag{42}$$

From this, the mass ratio can be expressed as

$$[\mathbf{H}]_{ij} = [\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T]^{-1}_{ij} \tag{43}$$

where the inverse is applied element-wise. This definition of $\mathbf{A}$ works for multiple simultaneous two-body collisions. In case three (or more) bodies collide with each other simultaneously, contacts are resolved sequentially.

Solving equation (38) for $\mathbf{I}$ and using equation (41) to allow two-body impacts yields

$$\boldsymbol{I}^k_{n+1} = \boldsymbol{L}^k_{n+1}\max\left(0, \lambda^k_{n+1}\right) \qquad \lambda^k_{n+\frac{3}{2}} = \left[\mathbf{H}_{n+1}(e\dot{\mathbf{Q}}_{n+\frac{1}{2}} + \dot{\mathbf{Q}}^S_{n+\frac{3}{2}})\mathbf{L}^T_{n+1}\right]_{kk}. \tag{44}$$

where $\boldsymbol{L}^k = \boldsymbol{n}_k$. In equation (44), the operator $\mathbf{H}$ incorporates the mass ratio, and is responsible for selecting the two bodies that are interacting in the collision when applied to $\dot{\mathbf{Q}}$. $\mathbf{L}$ projects the velocity for every body onto the surface normal of the contact surface. Since we work with independent single-node bodies, we are only interested in the diagonal elements of $\mathbf{H}$. The maximum makes sure the velocity of the colliding object is not pointing away from the surface, indicating that the collision already happened in the previous time step. We replace $\boldsymbol{g}(\mathbf{Q})$ with the contact network $\hat{\boldsymbol{c}}_\theta(\mathbf{Q}, \dot{\mathbf{Q}})$. Since the contact network $\hat{\boldsymbol{c}}_\theta$ receives $\mathbf{Q}$ as well as $\dot{\mathbf{Q}}$, it can learn whether the collision already happened or not, and thus we omit the max.

Summarising, the integrator is governed by the equations

$$\mathbf{Q}_{n+1} = \mathbf{Q}_n + h\dot{\mathbf{Q}}_{n+\frac{1}{2}}, \tag{45}$$

$$\mathbf{L}_{n+1} = \mathbf{L}(\mathbf{Q}_{n+1}), \quad \hat{\boldsymbol{c}}_{n+1} = \hat{\boldsymbol{c}}_\theta(\mathbf{Q}_{n+1}, \dot{\mathbf{Q}}_{n+\frac{1}{2}}), \quad \mathbf{H}_{n+1} = \mathbf{H}(\mathbf{Q}_{n+1}, \dot{\mathbf{Q}}_{n+\frac{1}{2}}) \tag{46}$$

$$\mathbf{F}_{n+1} = \frac{\partial V(\mathbf{Q}_{n+1})}{\partial \mathbf{Q}_{n+1}}, \tag{47}$$

$$\mathbf{U}_{n+\frac{3}{2}} = -e\dot{\mathbf{Q}}_{n+\frac{1}{2}}\mathbf{L}^T_{n+1}, \tag{48}$$

$$\begin{cases} \text{if } \hat{\boldsymbol{c}}^k_{n+1} = 1 & \boldsymbol{I}^k_{n+\frac{3}{2}} = \boldsymbol{L}^k_{n+1}\left[\mathbf{H}\left(\mathbf{U}_{n+\frac{3}{2}} + \left(\dot{\mathbf{Q}}_{n+\frac{1}{2}} + h\mathbf{M}^{-1}\mathbf{F}_{n+1}\right)\mathbf{L}^T_{n+1}\right)\right]_{kk} \\ \text{otherwise} & \boldsymbol{I}^k_{n+\frac{3}{2}} = \mathbf{0} \end{cases}, \tag{49}$$

$$\dot{\mathbf{Q}}_{n+\frac{3}{2}} = \dot{\mathbf{Q}}_{n+\frac{1}{2}} + \mathbf{M}^{-1}\left(h\mathbf{F}_{n+1} + \mathbf{I}_{n+1}\right). \tag{50}$$

In our experiments we assume $\mathbf{H}$ is known. $\hat{\boldsymbol{c}}_\theta$ represents the contact-network, which returns 1 if a contact occurs and 0 if not. The potential $\boldsymbol{V}$ is approximated using the potential-network.

We conclude with two remarks on more general settings. Firsly, if we wish to allow multiple nodes per body the operator $\mathbf{H}$ needs to be adapted accordingly, and the internal stress field $\boldsymbol{W}$ has to be added to the force (47), to yield

$$\mathbf{F}_{n+1} = \frac{\partial V(\mathbf{Q}_{n+1})}{\partial \mathbf{Q}_{n+1}} - \frac{\partial W(\mathbf{Q}_{n+1})}{\partial \mathbf{Q}_{n+1}}. \tag{51}$$

Further, it is no longer possible to use only the diagonal elements of the impulse equation (49), as now the interactions between nodes play a role in contact resolution.

Secondly, the force $\mathbf{F}$ can easily be extended to contain other forces. For example, we can add damping to obtain

$$\mathbf{F}_{n+1} = \frac{\partial V(\mathbf{Q}_{n+1})}{\partial \mathbf{Q}} - \mathbf{C}\dot{\mathbf{Q}}, \tag{52}$$

where $\mathbf{C}$ is a learned damping coefficient.

## B Appendix: experimental details

**Learning rate.** We use a learning rate of $10^{-3}$ in combination with the ADAM optimiser in all experiments.

**Potential.** We use the set-up proposed by Sæmundsson et al. (2020), which consists of a two-layer fully connected neural network with tanh activation function between the layers. The input layer has a size of 500 units and outputs a single real scalar. $\ell^2$ regularisation is used in both layers.

**Baseline contact-aware residual network.** To ensure a fair comparison, we use a vanilla residual network (ResNet) as a baseline. It consists of two dense layers with tanh activation function and an input layer size of 500 units. We update the state according to

$$\boldsymbol{s}_{n+1} = \text{ResNet}(\boldsymbol{s}_n, \boldsymbol{c}_n). \tag{53}$$

The residual network also receives the contact information $c_n$ and uses the same architecture as $\hat{c}_\theta$ to predict contact by means of

$$\boldsymbol{c}_{n+1} = \hat{c}_\theta^{\text{ResNet}}(\boldsymbol{s}_n). \tag{54}$$

**Pendulum.** The parameters used in the pendulum experiment are

$$g = 9.81\,\frac{\text{m}}{\text{s}^2} \qquad q_0 = 1 \qquad M = 1\,\text{kg} \qquad h = 0.02\,\text{s} \qquad l = 1\,\text{m}. \tag{55}$$

We add centered Gaussian noise with standard deviation $\sigma = 0.2$ to the positions and velocities. All models are trained for 3000 epochs.

**Bouncing ball.** The projection operator is $L = 1$, and the other parameters are

$$e = 1 \qquad g = 9.81\,\frac{\text{m}}{\text{s}^2} \qquad q_0 = 10\,\text{m} \qquad M = 1\,\text{kg} \qquad h = 0.02\,\text{s} \qquad \sigma = 0.2. \tag{56}$$

All models are trained for 2000 epochs.

**Newton's cradle.** We will use two generalised angles to determine the position of both balls. The initial positions and velocities are

$$\mathbf{Q}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \dot{\mathbf{Q}}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \tag{57}$$

where $(\mathbf{Q}_0^{(1)}, \dot{\mathbf{Q}}_0^{(1)})$ represent body 1, and $(\mathbf{Q}_0^{(2)}, \dot{\mathbf{Q}}_0^{(2)})$ represent body 2. The restitution parameter is $e = 1$, and

$$\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{58}$$

For simplification, we assume both bodies have no volume and are connected with the same joint. This means collisions will always happen perpendicular to the contact surface. The projection operator is thus

$$\mathbf{L} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{59}$$

**Additional experiments.** Here we include additional experimental results for the Newton's cradle without touch feedback in Figure 9, and the bouncing ball with elasticity $e = 0.8$ in Figure 10.

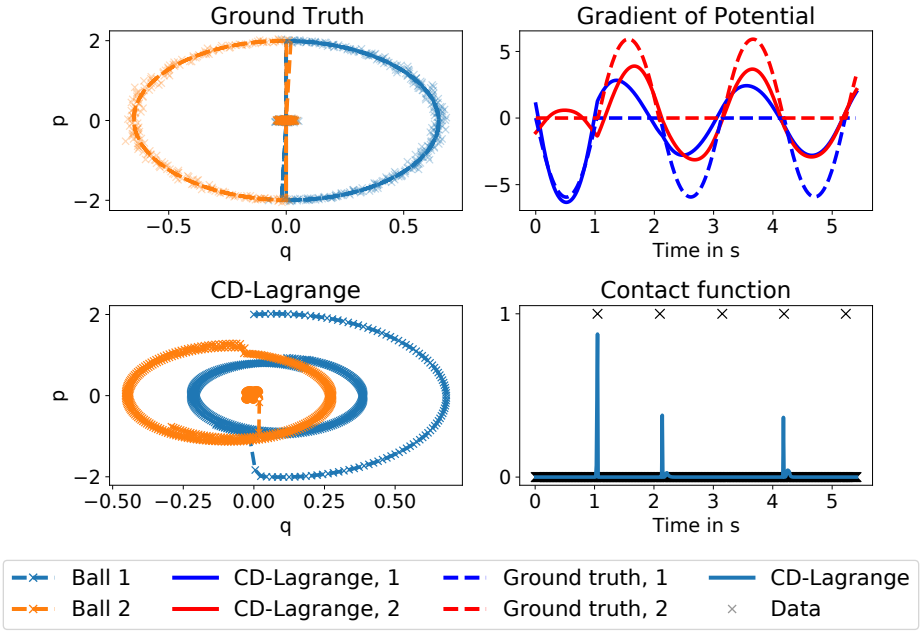


Figure 9: Learning the equations of motion of the Newton's cradle *without touch feedback* and Gaussian noise in the training data with $\sigma = 0.02$. The model tries to approximate the collisions using the potential network and therefore struggles to predict the trajectory accurately. CD-Lagrange RMSE: 0.907.

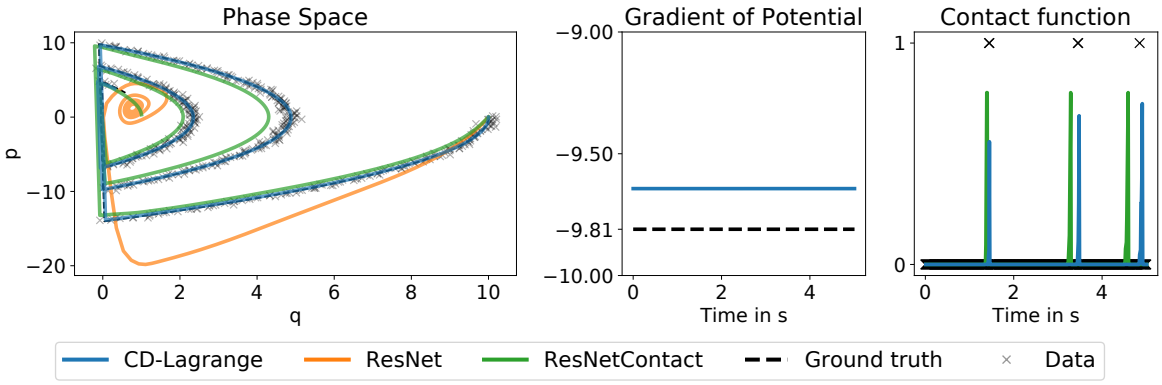


Figure 10: Learning the equations of motion of the bouncing ball with elasticity $e = 0.7$. CD-Lagrange approximates the true trajectory in the most accurate manner, particularly for longer simulation times which correspond to phase space regions closer to the centre of the spiral. CDL RMSE:2.076, ResNet RMSE: 8.291, ResNetContact RMSE: 4.156.