Inequality constraints in discrete variational (fluid) dynamics

Onno Bokhove
Leeds Institute for Fluid Dynamics, Leeds, UK

February 17, 2025

Abstract

1 Introduction

2 Variational modified-midpoint time-stepping scheme

A time-discrete variational principle over one time-slab is derived in order to obtain the well-known second-order modified-midpoint time-stepping scheme for a classical Hamiltonian system. The time-discrete variational principle is simplified further to facilitate automated numerical implementation thereof, directly into the *Firedrake* framework. Subsequently, this *variationalised* modified-midpoint time discretisation will be used in the discretisation of augmented Lagrangian variational principles for dynamics subject to inequality constraints.

Consider the variational principle (VP) for an autonomous one-degree-of-freedom Hamiltonian system in time

$$0 = \delta \int_0^T p\dot{q} - H(q,p)\,\mathrm{d}t \tag{1}$$

with time $t \in [0,T]$ $(T>0)$, $\dot{q} \equiv \frac{\mathrm{d}q}{\mathrm{d}t}$, generalised coordinate q and momentum p as well as Hamiltonian (energy) $H(q,p)$. We start from an Ansatz based on the discontinuous Galerkin finite-element time-stepping scheme (3.25) in Gagarina's thesis [8]. The only stable limit therein, in Gagarina's (3.25), emerges when the initial condition and jumps at the time nodes are taken to be continuous and when Gagarina's stability criterion for a linear weighing parameter of value $\alpha = 1/2$ is imposed *a priori*.

Consequently, in each time-slab $t \in [t^n, t^{n+1} = t^n + \Delta t]$ the second-order finite-element expansions and approximation of the integral become

$$q = \frac{2(t-t^n)}{\Delta t} q^{n+1/2} + \frac{t^n + t^{n+1} - 2t}{\Delta t} q^n, \tag{2a}$$

$$p = \frac{2(t-t^n)}{\Delta t} p^{n+1/2} + \frac{t^n + t^{n+1} - 2t}{\Delta t} p^n, \tag{2b}$$

$$\int_{t^n}^{t^{n+1}} H(p,q)\,\mathrm{d}t \approx \Delta t H(q^{n+1/2}, p^{n+1/2}), \tag{2c}$$

in which $\{q^n, q^{n+1}, p^n, p^{n+1}\}$ are the variables $\{q, p\}$ evaluated at the regularly space time nodes $\{t^n, t^{n+1}\}$, whereas $\{q^{n+1/2}, p^{n+1/2}\}$ are $\{q, p\}$ evaluated in the middle $t^{n+1/2} = (t^n + t^{n+1})/2$ of the time-slab. When we use this continuous-Galerkin approach rather than the discontinuous-Galerkin set-up in [8], Gagarina's expression (3.25) for the time-discrete Lagrangian becomes

$$\begin{aligned} 0 = \delta \sum_{m=0}^{N-1} 2p^{m+1/2}(q^{m+1/2} - q^m) - \Delta t H(p^{m+1/2}, q^{m+1/2}) \\ + \delta \sum_{m=-1}^{N-1} (q^{m+1} + q^m - 2q^{m+1/2}) \left(p^{m+1/2} + \frac{1}{2}(p^{m+1} - p^m) \right), \end{aligned} \tag{3}$$

wherein we have used the symbol m to emphasize that it is a summation dummy.

By collecting the contributions $\{q^{n+1/2}, p^{n+1/2}, q^n, p^n\}$ herein (3) relevant to a time level n within the summation over m, the time-discrete VP over the involved time-slabs becomes (by carefully avoiding one term to count twice)

$$\begin{aligned} 0 = \delta \Big(& p^{n+1/2}(q^{n+1} - q^n) - q^{n+1/2}(p^{n+1} - p^n) \\ & + \frac{1}{2}(p^{n+1} - p^n)(q^{n+1} + q^n) + \frac{1}{2} q^n (p^n - p^{n-1}) \\ & + q^n p^{n-1/2} - q^{n-1/2} p^n + \frac{1}{2} p^n q^{n-1} \\ & - \Delta t H(q^{n+1/2}, p^{n+1/2}) \Big). \end{aligned} \tag{4}$$

Considering variations of (4) with respect to $\{q^{n+1/2}, p^{n+1/2}, q^n, p^n\}$ yields the discrete equations

$$\delta p^{n+1/2}: \quad q^{n+1} - q^n - \Delta t \frac{\partial H(q^{n+1/2}, p^{n+1/2})}{\partial p^{n+1/2}} = 0 \tag{5a}$$

$$\delta q^{n+1/2}: \quad p^{n+1} - p^n + \Delta t \frac{\partial H(q^{n+1/2}, p^{n+1/2})}{\partial q^{n+1/2}} = 0 \tag{5b}$$

$$\begin{aligned} \delta q^n: \quad & - p^{n+1/2} + \frac{1}{2}(p^{n+1} - p^n) + p^{n-1/2} \\ & + \frac{1}{2}(p^n - p^{n-1}) = 0 \end{aligned} \tag{5c}$$

$$\begin{aligned} \delta p^n: \quad & q^{n+1/2} - \frac{1}{2}(q^{n+1} + q^n) - q^{n-1/2} \\ & + \frac{1}{2}(q^n + q^{n-1}) = 0, \end{aligned} \tag{5d}$$

which equals Gagarina's (3.26) for $\alpha = 1/2$ and for continuous values with no jumps taken at the time nodes. Alternatively, we can directly derive the above from VP (3) but the interim VP (4) has been provided because it is of direct use in the numerical implementation over one time-slab. For $n = 0$, the last two expressions yield that

$$\delta q^0: \quad - p^{1/2} + \frac{1}{2}(p^1 - p^0) + p^{-1/2} + \frac{1}{2}(p^0 - p^{-1}) = 0 \tag{6}$$

$$\delta p^0: \quad q^{1/2} - \frac{1}{2}(q^1 + q^0) - q^{-1/2} + \frac{1}{2}(q^0 + q^{-1}) = 0. \tag{7}$$

At $n = 0$, imposing that

$$q^{-1/2} = \frac{1}{2}(q^0 + q^{-1}) \quad \text{and} \quad p^{-1/2} = \frac{1}{2}(p^0 + p^{-1}), \tag{8}$$

we therefore find from (6) that

$$p^{1/2} = \frac{1}{2}(p^1 + p^0),\ q^{1/2} = \frac{1}{2}(q^1 + q^0), \tag{9}$$

which expressions we subsequently use in (5) for $n = 1$ and so forth in a bootstrap process. The initial condition will be imposed at the end, in part because at this stage it is unclear how to do so. This bootstrap process is less than ideal since we do not have a closed-form VP over one time-slab without having to impose the starting expressions (8) at $n = 0$.

Instead, in order to facilitate implementation, consider the incomplete time-discrete VP

$$\begin{aligned}
0 = \delta \Big(& p^{n+1/2}(q^{n+1} - q^n) - q^{n+1/2}(p^{n+1} - p^n) \\
& -\Delta t H(q^{n+1/2}, p^{n+1/2})\Big) && \text{(10a)}\\
= &\left(q^{n+1} - q^n - \frac{\partial H(q^{n+1/2}, p^{n+1/2})}{\partial p^{n+1/2}}\right) \delta p^{n+1/2} + \\
& - \left(p^{n+1} - p^n + \frac{\partial H(q^{n+1/2}, q^{n+1/2})}{\partial p^{n+1/2}}\right) \delta q^{n+1/2}, && \text{(10b)}
\end{aligned}$$

but take variations therein with respect to $\delta q^{n+1/2}, \delta p^{n+1/2}$ only. Subsequently, impose the midpoint averages after taking these variations

$$q^{n+1/2} = \frac{1}{2}(q^n + q^{n+1}) \quad \text{and} \quad p^{n+1/2} = \frac{1}{2}(p^n + p^{n+1}) \tag{10c}$$

and rewrite these such that we can eliminate the unknowns

$$q^{n+1} = 2q^{n+1/2} - q^n,\ p^{n+1} = 2p^{n+1/2} - p^n \tag{11}$$

from (10), whence we can exclusively solve for the (nonlinear) coupled system of variables $\{q^{n+1/2}, p^{n+1/2}\}$. Once we have solved for $\{q^{n+1/2}, p^{n+1/2}\}$, the updates $\{q^{n+1}, p^{n+1}\}$ in time follow from (11). At the initial time $n = 0$, we enforce the initial conditions $\{p^0 = p(0), q^0 = q(0)\}$, such that at every stage $\{q^n, p^n\}$ are known values.

The differences with the derivation of the modified midpoint scheme in Brown [4] are as follows: our derivation stems from a discontinuous Galerkin finite-element time-stepping scheme, cf. Gagarina [8], uses exclusively $q^{n+1/2}, p^{n+1/2}$ in the Hamiltonian while Brown [4] takes the mean of $\partial H/\partial q$ since his $H = H(p^{n+1/2}, (q^n + q^{n+1})/2)$ and, in the last simplified implementational version (10), we only take variations with respect to $q^{n+1/2}, p^{n+1/2}$ while imposing the definitions of q^{n+1}, p^{n+1} afterwards. The outcome of both of our derivations is therefore somewhat different from Brown's. However, it what follows it turns out that Brown's midpoint scheme is more compact and robust. It reads

$$\begin{aligned}
0 =& \delta \sum_{n=1}^{N} p^{n+1/2}(q^{n+1} - q^n) - \Delta t H\left(\frac{1}{2}(q^n + q^{n+1}), p^{n+1/2}\right) && \text{(12a)}\\
=& \sum_{n=1}^{N} q^{n+1} - q^n - \frac{\partial H\left(\frac{1}{2}(q^n + q^{n+1}), p^{n+1/2}\right)}{\partial p^{n+1/2}}) \delta p^{n+1/2} && \text{(12b)}\\
& - \sum_{n=2}^{N-1} \Big(p^{n+1/2} - p^{n-1/2} + \frac{1}{2}\Delta t \frac{\partial H\left(\frac{1}{2}(q^n + q^{n+1}), p^{n+1/2}\right)}{\partial q^n} \\
& + \frac{1}{2}\Delta t \frac{\partial H\left(\frac{1}{2}(q^n + q^{n-1}), p^{n-1/2}\right)}{\partial q^n}\Big) \delta q^n + \dots, && \text{(12c)}
\end{aligned}$$

such that the discrete equations become:

$$q^{n+1} - q^n = \frac{\partial H\left(\frac{1}{2}(q^n + q^{n+1}), p^{n+1/2}\right)}{\partial p^{n+1/2}} \tag{12d}$$

$$\begin{aligned} p^{n+1/2} - p^{n-1/2} =& -\frac{1}{2}\Delta t \frac{\partial H\left(\frac{1}{2}(q^n + q^{n+1}), p^{n+1/2}\right)}{\partial q^n} \\ & -\frac{1}{2}\Delta t \frac{\partial H\left(\frac{1}{2}(q^n + q^{n-1}), p^{n-1/2}\right)}{\partial q^n}. \end{aligned} \tag{12e}$$

We note that the above approach starting from (10) immediately extends to multiple degrees of freedom by taking the variables $\{q, p\}$ as vectors rather than scalars. We have successfully used the above modified-midpoint time-stepping scheme (10) to numerically implement and solve VPs of canonical water-wave dynamics, in which the (vector) $\{q, p\}$-variables are the free-surface height and surface/bottom velocity-potential fields [7, 9]. In such more complicated applications, it is essential to have a time-discrete VP over one time-slab. Particularly within the finite-element framework *Firedrake* [10], such (time-discrete) VPs can be implemented directly, upon which (partial) variational derivatives can be used to automatically derive and subsequently solve the weak formulations, cf. Alnaes et al. [1, 2]. Such a direct computational implementation of VPs, using (10), is efficient in that it reduces time-to-development and errors.

3 Inequality constraints in time-continuous finite-dimensional canonical variational principles

Inequality constraints will be considered next for (canonical) dynamical systems governed by a variational principle or Hamiltonian formalism. The general augmented Lagrangian considered (cf. [6]) reads

$$0 = \delta \int_0^T \boldsymbol{p} \cdot \dot{\boldsymbol{q}} - \frac{1}{2}|\boldsymbol{p}|^2 - V(\boldsymbol{q}) - \frac{1}{2\gamma}\left(F_+(-\gamma G(\boldsymbol{q}) - \lambda)^2 - \lambda^2\right) \mathrm{d}t \tag{13}$$

with time t, generalised coordinates $\boldsymbol{q}$ and momentum $\boldsymbol{p}$, potential energy $V(\boldsymbol{q})$, large constant $\gamma \gg 0$, Lagrange multiplier λ and inequality constraint $-G(\boldsymbol{q}) \le 0$ (as is the sign convention, or $G(\boldsymbol{q}) \ge 0$). The function $F_+(q) = q_+ = \max(q, 0)$ of argument q is zero or positive, or a (monotonic) smooth approximation thereof later used in the numerical approximations. After using endpoint conditions $\delta\boldsymbol{q}(0)\delta\boldsymbol{q}(T) = 0$, the variations of (13) are as follows

$$\delta\boldsymbol{p}: \quad \dot{\boldsymbol{q}} = \boldsymbol{p} \tag{14a}$$

$$\delta\boldsymbol{q}: \quad \dot{\boldsymbol{p}} = -\frac{\partial V}{\partial \boldsymbol{q}} + \frac{\partial G(\boldsymbol{q})}{\partial \boldsymbol{q}} F_+(-\gamma G(\boldsymbol{q}) - \lambda) F_+'(-\gamma G(\boldsymbol{q}) - \lambda) \tag{14b}$$

$$\delta\lambda: \quad \lambda = -F_+\left(-\gamma G(\boldsymbol{q}) - \lambda\right) F_+'\left(-\gamma G(\boldsymbol{q}) - \lambda\right). \tag{14c}$$

At every time t, the following Karush-Kuhn-Tucker conditions (cf. [6]) are satisfied

$$-G(\boldsymbol{q}) \le 0, \quad \lambda \le 0, \quad \lambda G(\boldsymbol{q}) = 0, \tag{15}$$

since these can be shown to be equivalent to

$$\lambda = -\left[-\gamma G(\boldsymbol{q}) - \lambda\right]_+ = -F_+\left(-\gamma G(\boldsymbol{q}) - \lambda\right). \tag{16}$$

That is, to prove this equivalence, consider in turn that:

- if $\lambda < 0$ then (15)c yields that $G(\boldsymbol{q}) = 0$ such that in (16) $\lambda = -F_+(-\lambda)$; so when $\lambda < 0$ one finds $\lambda = \lambda$ identically;
- if $\lambda = 0$ then in (16) $0 = -F_+\left(-\gamma G(\boldsymbol{q})\right)$ such that we must have $-G(\boldsymbol{q}) \leq 0$ or $G(\boldsymbol{q}) \geq 0$;
- if $\lambda > 0$ then (15)c yields that $G(\boldsymbol{q}) = 0$ such that in (16) $\lambda = -F_+(-\lambda) = 0$, giving a contradiction in that λ cannot be positive but must be zero.

Approximations of $F_+(\cdot)$ include:

$$
\begin{aligned}
F_+(q) =& \frac{1}{2}q + \sqrt{b^2 + \frac{1}{4}q^2} \to_{b\to 0} \max(q,0) \quad \text{with} \\
F_+'(q) =& \frac{1}{2} + \frac{\frac{1}{4}q}{\sqrt{b^2 + \frac{1}{4}q^2}} \to_{b\to 0} \Theta(q) \\
F_+(q)F_+'(q) =& \frac{1}{2}q + \frac{1}{2}\sqrt{b^2 + \frac{1}{4}q^2} + \frac{\frac{1}{8}q^2}{\sqrt{b^2 + \frac{1}{4}q^2}} \to_{b\to 0} F_+(q) \qquad \text{(18a)}
\end{aligned}
$$

$$
F_+(q) = b\ln\left(1 + e^{q/b}\right), \quad F_+(q)F_+'(q) = \frac{b\ln\left(1 + e^{\frac{x}{b}}\right)}{\left(1 + e^{\frac{x}{b}}\right)} e^{\frac{x}{b}} \qquad \text{(18b)}
$$

$$
F_+(q) = \begin{cases} 0 & q < -b \\ (q+b)^2/(4b) & else \\ q & q > b \end{cases} \qquad \text{(18c)}
$$

$$
F_+(q) = \begin{cases} 0 & q < -b \\ \sqrt{(q+b)^3/(6b)} & else \\ \sqrt{q^2 + b^2/3} & q > b \end{cases} \qquad \text{(18d)}
$$

for small $b > 0$. The last approximation is an integrated version of the smooth quadratic one such that the smooth quadratic one equals $F_+(q)F_+'(q)$. When $F_+(q) = \max(q,0)$, one finds that $F_+(q)F_+'(q) = F_+(q)$, since $F_+'(q)$ is the Heaviside function $\Theta(q)$, but when $F_+(q)$ is a smooth approximation then generally $F_+(q)F_+'(q) \approx F_+(q)$ and $F_+(q)F_+'(q)$ is less smooth than $F_+(q)$.

Following (10a), the matching modified midpoint discrete VP is

$$
\begin{aligned}
0 =& \delta\Big(\boldsymbol{p}^{n+1/2} \cdot (\boldsymbol{q}^{n+1} - \boldsymbol{q}^n) - \boldsymbol{q}^{n+1/2} \cdot (\boldsymbol{p}^{n+1} - \boldsymbol{p}^n) - \frac{1}{2}\Delta t |\boldsymbol{p}^{n+1/2}|^2 \\
& - \Delta t V(\boldsymbol{q}^{n+1/2}) - \Delta t \frac{1}{2\gamma}\left(F_+(-\gamma G(\boldsymbol{q}^{n+1/2}) - \lambda)^2 - \lambda^2\right)\Big) \qquad \text{(19)}
\end{aligned}
$$

The variations of (19) are as follows, after subsequent elimination of the $(n+1)$-level variables using the midpoint definitions,

$$
\begin{aligned}
\delta\boldsymbol{p}^{n+1/2}: \quad & \boldsymbol{q}^{n+1/2} - \boldsymbol{q}^n = \frac{1}{2}\Delta t \boldsymbol{p}^{n+1/2} \qquad \text{(20a)} \\
\delta\boldsymbol{q}^{n+1/2}: \quad & \boldsymbol{p}^{n+1/2} - \boldsymbol{p}^n = -\frac{1}{2}\Delta \frac{\partial V}{\partial \boldsymbol{q}^{n+1/2}} \\
& + \frac{1}{2}\Delta t \frac{\partial G(\boldsymbol{q}^{n+1/2})}{\partial \boldsymbol{q}^{n+1/2}} F_+(-\gamma G(\boldsymbol{q}^{n+1/2} - \lambda) F_+'(-\gamma G(\boldsymbol{q}^{n+1/2}) - \lambda) \qquad \text{(20b)} \\
\delta\lambda: \quad & \frac{\Delta t}{\gamma}\lambda = -\frac{\Delta t}{\gamma} F_+\left(-\gamma G(\boldsymbol{q}^{n+1/2}) - \lambda\right) F_+'\left(-\gamma G(\boldsymbol{q}^{n+1/2}) - \lambda\right). \qquad \text{(20c)}
\end{aligned}
$$

To assess the Jacobian of (20) for preconditioning, we note that (20)c is an approximation of $\lambda G(\boldsymbol{q})$ with $\lambda < 0, G(\boldsymbol{q}) > 0$. It is not clear whether and which approximations preserve these inequality requirements in the $\{\lambda, G\}$-plane. We can rewrite the last equation (20)c as $\lambda - (\gamma G(\boldsymbol{q}) + \lambda)\,\Theta\,(-\gamma G(\boldsymbol{q}) - \lambda) = 0$. After adjusting for the prefactor $\Delta t/\gamma$, the leading components in the Jacobian then become (for unknowns $\{\boldsymbol{q}^{n+1/2}, \boldsymbol{p}^{n+1/2}, \lambda\}$, with Heaviside $\Theta(q)$ and δ-functions)

$$J = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\gamma \frac{\partial G}{\partial \boldsymbol{q}}\left(\Theta(q) - q\delta(q)\right) & 0 & 1 - \Theta(q) + q\delta(q) \end{pmatrix} \approx \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\gamma \frac{\partial G}{\partial \boldsymbol{q}}\Theta(q) & 0 & 1 - \Theta(q) \end{pmatrix}, \tag{21}$$

with $q \equiv -\gamma G(\boldsymbol{q}) - \lambda$. in which a (troublesome) δ-functions emerge. However, when they emerge under an integral these contributions with a δ-function cancel due to their multiplication with $(\gamma G(\boldsymbol{q}) + \lambda)$. It therefore seems useful to impose this latter adapted Jacobian directly into Firedrake such as to avoid these subtle terms with a δ-function:

$$J = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\gamma \frac{\partial G}{\partial \boldsymbol{q}}\Theta(q) & 0 & 1 - \Theta(q) \end{pmatrix}. \tag{22}$$

For the piecewise constant, linear and quadratic approximation, we find

$$J = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\gamma \frac{3(q+b)^2}{8b^2} \frac{\partial G}{\partial \boldsymbol{q}} & 0 & 1 - \frac{3(q+b)^2}{8b^2} \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\gamma \frac{\partial G}{\partial \boldsymbol{q}} & 0 & 0 \end{pmatrix}, \tag{23}$$

in which the terms $\{0, \frac{3(q+b)^2}{8b^2}, 1\}$ are poorly "approximating" the exact Heaviside function $\Theta(\boldsymbol{q})$, while the (approximation of the) δ-function is absent. Note that $|\frac{3(q+b)^2}{8b^2}| < 3/2$, since $q \le |b|$. By construction, the zero, quadratic and linear approximation to F_+F_+' (in the provided expression (18d) of $F_+(q)$) yields a piecewise linear and constant approximation of $\Theta(q)$ in the Jacobian.

4 Finite-dimensional test cases

4.1 Ball bouncing off surface under gravity

The variational principle (VP) for the dimensionless dynamics of a particle moving under gravity in the vertical direction Z and bouncing off a table such that $Z \ge 0$ or $-Z \le 0$ is the following

$$0 = \delta \int_0^T W\dot{Z} - \frac{1}{2}W^2 - Z - \frac{1}{2\gamma}\left(F_+(-\gamma Z - \lambda)^2 - \lambda^2\right) \mathrm{d}t \tag{24}$$

with unit mass and unit gravity after a scaling, positive constant $\gamma > 0$ and Lagrange multiplier λ. The velocity of the particle is W.

The VP (24) involves an augmented Lagrangian with function $F_+(q) = q_+ = \max(q, 0)$, or a (monotonic) smooth approximation thereof, with Lagrange multiplier

λ. The variations of (24) are as follows

$$\begin{aligned}
\delta W: \quad && \dot{Z} =& W && \text{(25a)}\\
\delta Z: \quad && \dot{W} =& -1 + F_+(-\gamma Z - \lambda)F'_+(-\gamma Z - \lambda) && \text{(25b)}\\
\delta\lambda: \quad && \lambda =& -F_+(-\gamma Z - \lambda)F'_+(-\gamma Z - \lambda) &&\\
&& =& -F_+(-\gamma Z - \lambda) && \text{(25c)}
\end{aligned}$$

after using the end-point conditions $\delta Z(0) = \delta Z(T) = 0$.

Note that the last step follows because $F'_+(-\gamma Z - \lambda) = \Theta(-\gamma Z - \lambda)$ and $F_+(-\gamma Z - \lambda)\Theta(-\gamma Z - \lambda) = F_+(-\gamma Z - \lambda)$. We keep this more complicated expression involving $F_+(\cdot)$ for later usage in both analysis and a numerical discretisation. At every time t, the following Karush-Kuhn-Tucker conditions are satisfied

$$-Z \leq 0, \quad \lambda \leq 0, \quad \lambda Z = 0, \tag{26}$$

since these can be shown to be equivalent to

$$\lambda = -\left[-\gamma Z - \lambda\right]_+ = -F_+(-\gamma Z - \lambda). \tag{27}$$

That is, to prove this equivalence, consider in turn that:

- if $\lambda < 0$ then (26)c yields that $Z = 0$ such that in (27) $\lambda = -F_+(-\lambda)$; so when $\lambda < 0$;
- if $\lambda = 0$ then in (27) $0 = -F_+(-\gamma Z)$ such that we must have $-Z \leq 0$ or $Z \geq 0$;
- if $\lambda > 0$ then (26)c yields that $Z = 0$ such that in (27) $\lambda = -F_+(-\lambda) = 0$, giving a contradiction in that λ cannot be positive.

Using the above approximations and that $F_+(q)F'_+(q) \approx F_+(q)$, involving that $\max(q, 0) \approx F_+(q)$, we can solve the equation governing λ's behaviour for the particular choice (18), such that

$$\begin{aligned}
\lambda =& -\left(-\frac{1}{2}(\gamma Z + \lambda) + \sqrt{b^2 + (\gamma Z + \lambda)^2/4}\right) && \text{(29a)}\\
\iff \frac{1}{2}(\lambda - \gamma Z) =& -\sqrt{b^2 + (\gamma Z + \lambda)^2/4} \implies && \text{(29b)}\\
-\gamma Z \lambda =& b^2 \quad \text{for} \quad Z \geq 0 \iff && \text{(29c)}\\
\lambda =& -\frac{b^2}{\gamma Z} \quad \text{for} \quad Z \geq 0. && \text{(29d)}
\end{aligned}$$

We cannot (yet) solve λ in a closed-form expression, when we use the full (18a). We note that there is no real-valued solution for $Z < 0$ and that the solution holds for $Z \geq 0$ with positive $\gamma \gg 1$. The equations of motion can be combined with this result such that the phase portrait of the Hamiltonian dynamics can be found:

$$\begin{aligned}
\dot{Z} = W, \dot{W} = -1 + \frac{b^2}{\gamma Z} \iff \ddot{Z} =& -1 + \frac{b^2}{\gamma Z} \iff && \text{(30a)}\\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}W^2 + Z - \frac{b^2}{\gamma}\ln Z\right) =& 0 \iff && \text{(30b)}\\
\frac{1}{2}W^2 + Z - \frac{b^2}{\gamma}\ln Z =& H_0 && \text{(30c)}
\end{aligned}$$

with the integration constant and energy $H_0 = H(t)$. When $W = 0$ the maximum $Z = Z_{max}$ satisfies $Z_{max} - b^2/(\gamma Z_{max}) = H_0$ with as first approximation $Z_{max} \approx H_0 = H_1 - b^2/(\gamma)\ln H_1$. Note that for the free-falling ball without constraint $Z \geq 0$, we obtain

$$\frac{1}{2}W^2 + Z = H_1 \tag{31}$$

with $Z_{max} = H_1$. Likewise for the constrained case the minimum Z_{min} satisfies

$$Z_{min} - b^2/(\gamma Z_{min}) = H_0 \Longrightarrow Z^0_{min} = e^{-\gamma H_0/b^2} \tag{32a}$$

$$\Longrightarrow Z^{(k)}_{min} - b^2/(\gamma Z^{(k+1)}_{min})) = H_0 \Longleftrightarrow \tag{32b}$$

$$Z^{(k+1)}_{min} = e^{-\gamma H_0/b^2} e^{Z^{(k)}_{min}} \tag{32c}$$

with the indicated (iterative) approximations $Z^{(k)}_{min}$ using that $Z_{min} \approx 0$ will be small and therefore that $e^{Z^{(k)}_{min}} \approx 1$. The phase portraits of the free falling particle and constrained particle dynamics can be compared in the W, Z–plane and explored for different values of the ratio b^2/γ. For $b^2/\gamma = 0.34$, the constrained dynamics takes place on closed contours, see Fig. 1.

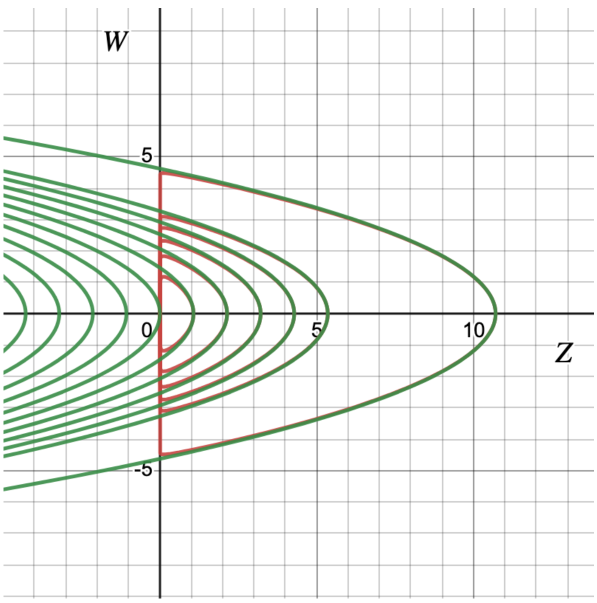

Figure 1: Phase portraits of the free and constrained Hamiltonian dynamics of a particle falling under unit gravity. Free falling case: energy H_1 and green lines; constrained case: energy $H_0 = H_1 - b^2 \ln H_0/(\gamma)$ and red lines. Labels for energy values: $H_1 = -5, -4, \ldots, 4, 5$; $b^2/\gamma = 0.34^2$.

Following and/or extending Bueler [5], the alternative formulation in terms of variational inequalities (VI) reads

$$\delta \int_0^T W\dot{Z} - \frac{1}{2}W^2 - Z\,\mathrm{d}t \leq 0 \Longrightarrow \tag{33a}$$

$$\int_0^T \delta W(\dot{Z} - W) - \delta Z(\dot{W} + 1)\,\mathrm{d}t \leq 0, \tag{33b}$$

such that we obtain the VI-system

$$\dot{Z} = W \tag{33c}$$

$$\dot{W} + 1 \geq 0 \tag{33d}$$

$$Z \geq 0. \tag{33e}$$

The system (25) can be rewritten as

$$\dot{Z} = W \tag{34}$$

$$\dot{W} + 1 + \lambda = 0 \tag{35}$$

$$Z \geq 0, \lambda \leq 0, \lambda Z = 0, \tag{36}$$

whence

$$\delta Z(\dot{W} + 1) = -\lambda \delta Z = |\lambda| \delta Z \geq -|\lambda| Z \tag{37}$$

since $Z + \delta Z \geq 0$. Hence, since $\min_{Z\geq 0} Z = 0$ one finds that

$$\delta Z(\dot{W} + 1) = -\lambda \delta Z \geq 0. \tag{38}$$

(Is that reasoning in red right?).

Using (10), a second-order modified-midpoint time-discrete version of VP (24) reads

$$\begin{aligned} 0 =& \delta\Big(W^{n+1/2} \frac{Z^{n+1} - Z^n}{\Delta t} - Z^{n+1/2} \frac{W^{n+1} - W^n}{\Delta t} \\ & - \frac{1}{2}(W^{n+1/2})^2 - Z^{n+1/2} - \frac{1}{2\gamma}\left(F_+(-\gamma Z^{n+1/2} - \lambda)^2 - \lambda^2\right)\Big) \end{aligned} \tag{39}$$

with additional relations

$$Z^{n+1} = 2Z^{n+1/2} - Z^n \text{ and } W^{n+1} = 2W^{n+1/2} - W^n. \tag{40}$$

The variations of (39) with respect to $\{W^{n+1/2}, Z^{n+1/2}, \lambda\}$ are

$$\delta W^{n+1/2}: \; Z^{n+1} = Z^n + \Delta t W^{n+1/2} \Longrightarrow$$

$$Z^{n+1/2} = Z^n + \frac{1}{2}\Delta t W^{n+1/2} \tag{41a}$$

$$\begin{aligned} \delta Z^{n+1/2}: \; W^{n+1} =& W^n - \Delta t \\ & + \Delta t F_+(-\gamma Z^{n+1/2} - \lambda) F_+'(-\gamma Z^{n+1/2} - \lambda) = W^n - \Delta t(1+\lambda) \end{aligned}$$

$$\begin{aligned} \Longrightarrow \frac{4(Z^{n+1/2} - Z^n)}{\Delta t} =& 2W^n - \Delta t(1+\lambda) \\ & + \Delta t F_+(-\gamma Z^{n+1/2} - \lambda) F_+'(-\gamma Z^{n+1/2} - \lambda) \end{aligned} \tag{41b}$$

$$\delta\lambda: \; \lambda = -F_+(-\gamma Z^{n+1/2} - \lambda) F_+'(-\gamma Z^{n+1/2} - \lambda); \tag{41c}$$

noting that these last two equations (41)(b,c) are coupled equations for the unknowns $\{Z^{n+1/2}, \lambda\}$. Given that we can isolate $Z^{n+1/2}$, we can write down one equation for λ as follows

$$Z^{n+1/2} = Z^n + \frac{\Delta t W^n}{2} - \frac{(1+\lambda)\Delta t^2}{4} \quad \text{such that} \tag{42}$$

$$\begin{aligned} \lambda =& -F_+\left(-\gamma Z^n - \gamma \frac{\Delta t W^n}{2} + \gamma(1+\lambda)\frac{\Delta t^2}{4}\right) \times \\ & F_+'\left(-\gamma Z^n - \gamma \frac{\Delta t W^n}{2} + \gamma(1+\lambda)\frac{\Delta t^2}{4}\right). \end{aligned} \tag{43}$$

Two versions with inequality constraints have been implemented: one in which the weak formulation of λ is explicitly coded up using the approximation $F_+(q)F'_+(q) \approx F_+(q)$; and in the other one, the time-discrete VP is implemented directly with the weak formulations generated automatically within Firedrake. In the latter case, the approximation is rougher than in the former case. The former case is strictly speaking not shown to be variational anymore, except that we have shown that λ can be explicitly eliminated and that the system has an unconstrained formulation with a matching unconstrained Hamiltonian involving an explicit potential. Furthermore, a mtaching $F_+(q)$ can likely be found by integral quadrature. In both cases, the time step is chosen heuristically in order to resolve the particle turning smoothly and with sufficient resolution near $Z = 0$. Standalone symplectic Euler (SE) and Störmer-Verlet (SV) discretisations have also been applied and implemented for comparison.

Results shown in Figs. 2 and 3 are revealing that the MMP discretisations works best and that all discretisations require suffuciently small time steps in order to resolve the particle dynamics. We note that the full VP-based MMP-discretisation has a smoother cushioning of the dynamics near Z and is more accurate further away from $Z = 0$, as is apparent from the shape of the phase portraits in Fig. 2 (note that the values of b are different and have been heuristically tuned by visual comparison). The relation between the choice of an adequate time step and the parameter values $\{b, \gamma\}$ is as yet unknown. To date, the current choices are based in trial and error.

Firedrake's solver settings used in the MMP discretisations are as follows:

```
solver_parameters = {
    'snes_type': 'ksponly',  # Use KSP for nonlinear solve
    'ksp_type': 'preonly',   # Use preconditioned KSP
     # 'pc_type': 'bjacobi', # Use block Jacobi preconditioner both bjacobi and none work
    'pc_type': 'none',       # None
    'sub_pc_type': 'ilu',    # works with and without with none; does not work with bjacobi
    'mat_type': 'nest',      # AIJ matrix type
    'snes_max_it': 100,      # Maximum number of Picard iterations
    'ksp_rtol': 1e-8,        # Relative tolerance for linear solve
}
```

The VP

$$0 = \delta \int_0^T W\dot{Z} - \frac{1}{2}W^2 - Z + \lambda^2 Z - 2a\ln\lambda \,\mathrm{d}t \tag{44a}$$

yields (after scaling one equation with $1/\Delta t$

$$\dot{Z} - W = 0 \tag{44b}$$

$$\dot{W} + Z - \lambda^2 = 0 \tag{44c}$$

$$2\lambda Z - 2a/\lambda = 0 \quad \Longrightarrow \quad \lambda^2 = a/Z. \tag{44d}$$

With $a = b^2/\gamma$ this yields the approximate solution analysed earlier. Substituting $\lambda = \sqrt{a/Z}$ back into the VP yields the energy $H = W^2/2 + Z - a\ln(Z)$ used in the phase plots found before. The MMP-implementation diverges though. The leading terms in the Jacobian of the system within (44a) are:

$$J \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2\lambda \\ 2\lambda & 0 & 2(Z + a/\lambda^2) \end{pmatrix}. \tag{45}$$

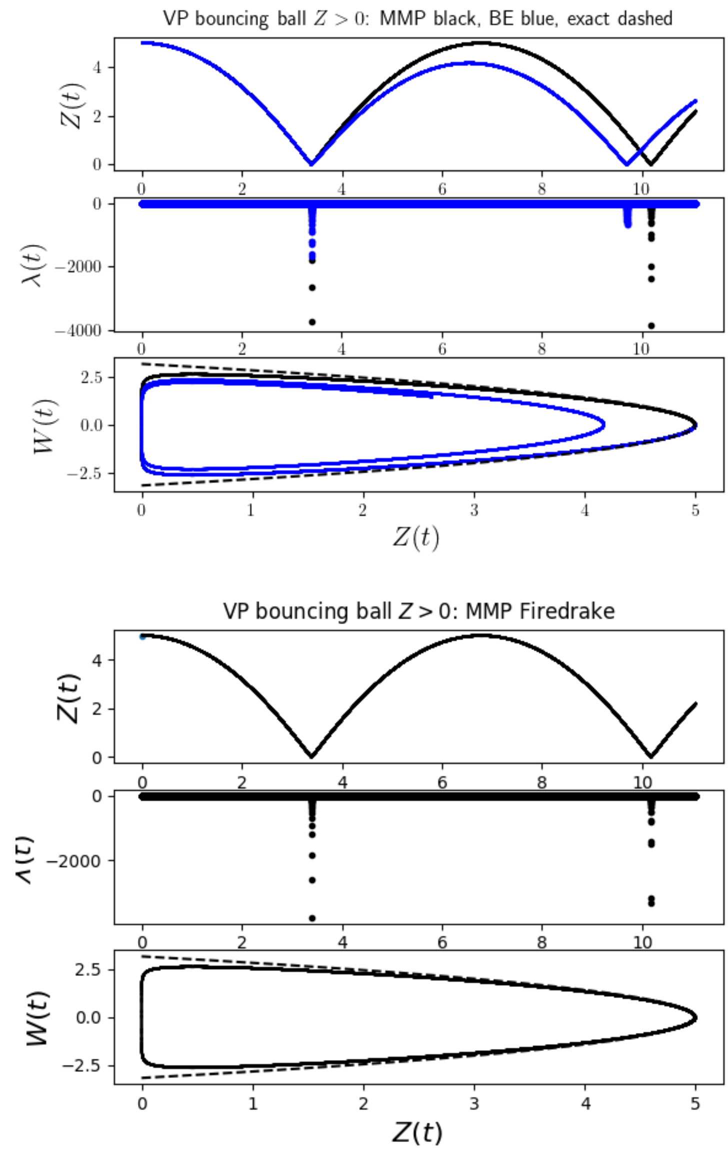

Figure 2: Phase portraits of the free and constrained Hamiltonian dynamics of a particle falling under unit gravity. Top: MMP (black line) and SE (blue line) standalone code; SE is seen to be overstable and energy is decreasing. The dashed line denotes the phase portrait in the free-falling case. Bottom: MMP Firedrake simulations $b^2/\gamma = 0.68^2$. *Made with codes SE/SV/MMP: plainball.py and FD: wavemmpball2024errorcopy.py*

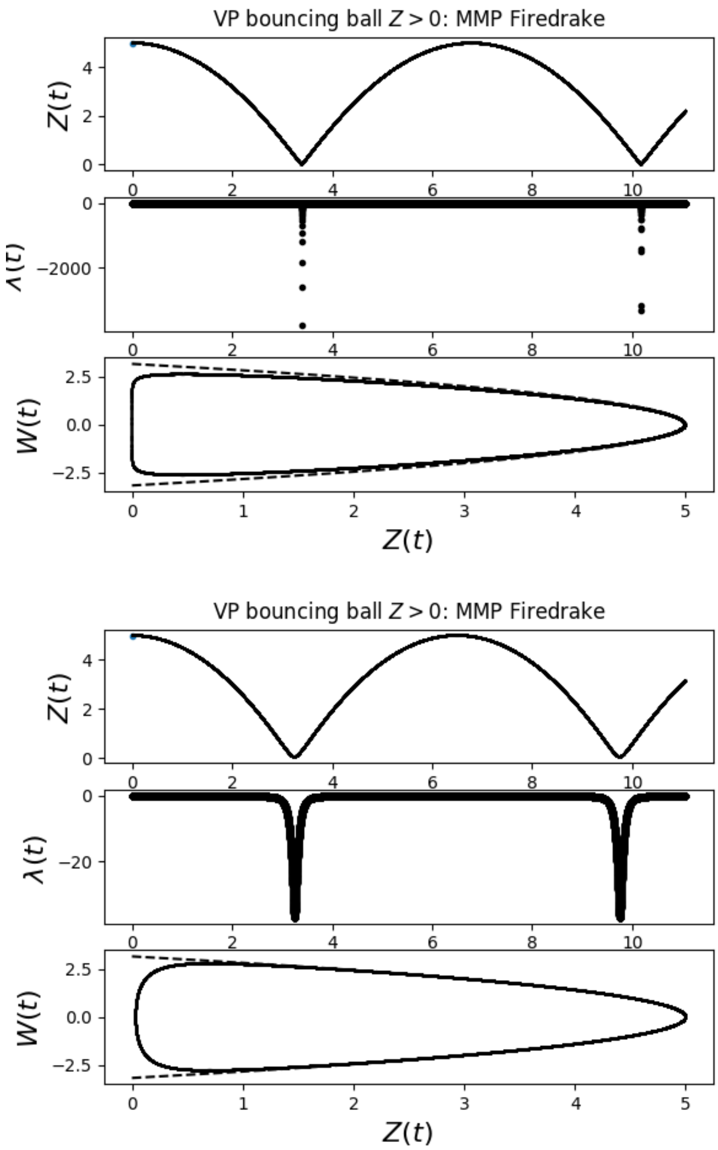

Figure 3: Phase portraits of the free (dashed) and constrained (solid) Hamiltonian dynamics of a particle falling under unit gravity. Top: numerics for MMP weak formulation for λ-parts using $F_+(q)$ (black line) and bottom: MMP based on time-discrete VP using $F_+(q)$ yielding $\lambda = -F_+(q)F'_+(q)$ (black line). Firedrake implementations. Parameters $\gamma = 100$, top: $b = 0.68\sqrt{\gamma}$ and bottom: b is four time larger; $\Delta t = 0.0005$. Note that displays of $F_+(q)$ and $F_+(q)F'_+(q)$ are visually similar for these different values of b. *Made with code FD: wavemmpball2024errorcopy.py constraint setting with weak formulation for* λ *or the full time-discrete VP.*

4.2 Free particle on a squircle billiard table

The following VP describes the two-dimensional dynamics of a free particle of unit mass moving within a billiard table defined by a hard-walled squircle boundary

$$\begin{aligned} 0 =& \delta \int_0^T U\dot{X} + V\dot{Y} - \frac{1}{2}(U^2 + V^2) \\ & - \frac{1}{2\gamma}\left(F_+\Big(\gamma((X/L_x)^{2m} + (Y/L_y)^{2m} - 1) - \lambda\Big)^2 - \lambda^2\right)\mathrm{d}t, \end{aligned} \tag{46}$$

with particle coordinates (X, Y) and velocities (U, V), wherein the squircle billiard is imposed as the following inequality constraint

$$-G(X,Y) = (X/L_x)^{2m} + (Y/L_y)^{2m} - 1 \leq 0, \tag{47}$$

with integer value $m > 0$ such that for large values of m the billiard approaches the rectangular shape: $|X| \leq L_x, |Y| \leq L_y$ with side lengths L_x and L_y. Variations of (46) yield the equations

$$\delta U, \delta V: \quad \dot{X} = U = \frac{\partial H}{\partial U}, \quad \dot{Y} = V = \frac{\partial H}{\partial V}, \tag{48a}$$

$$\delta X: \quad \dot{U} = -\frac{2m}{L_x}(X/L_x)^{2m-1}F_+(q)F_+'(q) = \frac{2m}{L_x}(X/L_x)^{2m-1}\lambda, \tag{48b}$$

$$\delta Y: \quad \dot{V} = -\frac{2m}{L_y}(Y/L_y)^{2m-1}F_+(q)F_+'(q) = \frac{2m}{L_y}(Y/L_y)^{2m-1}\lambda, \tag{48c}$$

$$\delta\lambda: \quad \lambda = -F_+(q)F_+'(q) \approx -F_+(q) \quad \text{with}$$
$$q \equiv \gamma((X/L_x)^{2m} + (Y/L_y)^{2m} - 1) - \lambda. \tag{48d}$$

The Hamiltonian or energy $H = H(X, Y, U, V, \lambda)$ is defined as

$$H = \frac{1}{2}(U^2 + V^2) - \frac{1}{2\gamma}\left(F_+\Big(\gamma((X/L_x)^{2m} + (Y/L_y)^{2m} - 1) - \lambda\Big)^2 - \lambda^2\right). \tag{48e}$$

A reduced Hamiltonian or energy $H = H(X, Y, U, V)$ is defined below after involving the approximation (18).

Alternatively, given or inspired by Bueler's work [5], consider (47)

$$\delta \int_0^T U\dot{X} + V\dot{Y} - \frac{1}{2}(U^2 + V^2)\,\mathrm{d}t, \leq 0 \Longrightarrow \tag{49a}$$

$$\int_0^T \delta U(\dot{X} - U) + \delta V(\dot{X} - V) - \dot{\mathbf{U}} \cdot \delta\mathbf{X} \leq 0. \tag{49b}$$

Hence, one finds the system rephrased in a variational ineqiality (VI) style

$$\delta\mathbf{U} \cdot (\dot{\mathbf{X}} - \mathbf{U}) = 0 \tag{49c}$$

$$\delta\mathbf{X} \cdot \dot{\mathbf{U}} \geq 0 \tag{49d}$$

$$\delta G\left(G(\mathbf{X}) - 1 + (X/L_x)^{2m} + (Y/L_y)^{2m}\right) = 0 \tag{49e}$$

$$G \geq 0. \tag{49f}$$

The system (48) with Lagrange multiplier can be rewritten as

$$\delta\mathbf{U}\cdot)\dot{\mathbf{X}} - \mathbf{U}0 = 0 \tag{50a}$$

$$\delta\mathbf{x} \cdot (\dot{\mathbf{U}} + \lambda\nabla G) = 0 \tag{50b}$$

$$\delta\lambda\lambda G(\mathbf{X}) = 0 \quad \text{with} \tag{50c}$$

$$G(\mathbf{X}) \geq 0, \lambda \leq 0, \tag{50d}$$

whence what follows after multiplication of the momentum equations is

$$\delta\mathbf{X} \cdot \dot{\mathbf{U}} = -\lambda\delta\mathbf{X} \cdot \nabla G = |\lambda|\delta\mathbf{X} \cdot \nabla G \geq -|\lambda|G(\mathbf{X}) \Longrightarrow \delta\mathbf{X} \cdot \dot{\mathbf{U}} \geq 0, \tag{51}$$

since $|\lambda| \geq 0$ and $G(\mathbf{X} + \delta\mathbf{X}) = G(\mathbf{X}) + \nabla G \cdot \delta\mathbf{X} \geq 0$ but also since the smallest value of $\min_{\mathbf{X}} G(\mathbf{X} = 0$. (Is that reasoning in red right?).

With this approximation (18) and $F_+(q)F'_+(q) \approx F_+(q)$, the Lagrange multiplier can again be solved exactly to be

$$\lambda = \frac{b^2}{\gamma\left((X/L_x)^{2m} + (Y/L_y)^{2m} - 1\right)} = -\frac{b^2}{\gamma\left(1 - (X/L_x)^{2m} - (Y/L_y)^{2m}\right)}, \tag{52}$$

such that the momentum equations become

$$\delta X: \quad \dot{U} = -\frac{2m}{L_x}(X/L_x)^{2m-1}\frac{b^2}{\gamma\left(1 - (X/L_x)^{2m} - (Y/L_y)^{2m}\right)} = -\frac{\partial H}{\partial X} \tag{53}$$

$$\delta Y: \quad \dot{V} = -\frac{2m}{L_y}(Y/L_y)^{2m-1}\frac{b^2}{\gamma\left(1 - (X/L_x)^{2m} - (Y/L_y)^{2m}\right)} = -\frac{\partial H}{\partial Y} \tag{54}$$

with Hamiltonian

$$H = \frac{1}{2}(U^2 + V^2) - \frac{b^2}{\gamma}\ln\left(1 - (X/L_x)^{2m} - (Y/L_y)^{2m}\right). \tag{55}$$

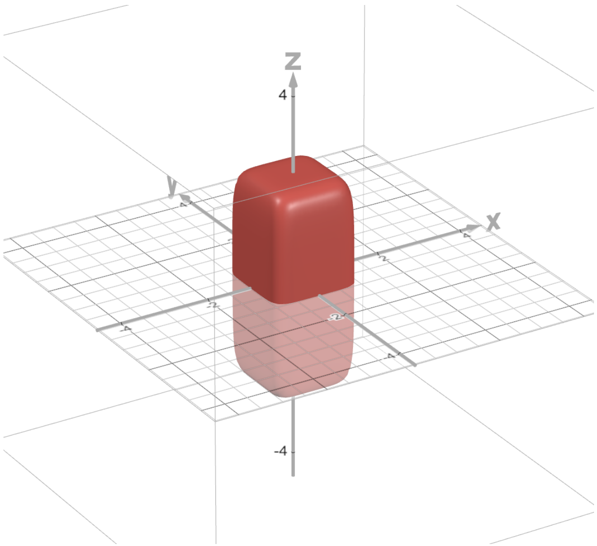

Figure 4: Illustrative (cross-section of) phase portrait of the constrained and smoothed Hamiltonian dynamics of a particle travelling in a squircle billiard. A cross-section in the four-dimensional $\{X, Y, U, V\}$-space is made by considering $\frac{1}{2}z^2 - a^2\ln\left(1 - (X/Lx)^{2m} - (Y/L_y)^{2m}\right) = H$ with $z^2 = U^2 + V^2$. The sign reversals of z occur when there has been an interaction with the squircle walls (and has no other meaning). Parameter values: $m = 6, a = 0.84, L_x = 1, L_y = 2, H = 2.7$.

Using (10), the modified-midpoint time discretisation of (46) is

$$\begin{aligned} 0 =& U^{n+1/2}\frac{(X^{n+1} - X^n)}{\Delta t} - X^{n+1/2}\frac{(U^{n+1} - U^n)}{\Delta t} \\ &+ V^{n+1/2}\frac{(Y^{n+1} - Y^n)}{\Delta t} - Y^{n+1/2}\frac{(V^{n+1} - V^n)}{\Delta t} - \frac{1}{2}({U^{n+1/2}}^2 + {V^{n+1/2}}^2) \\ &- \frac{1}{2\gamma}\left(F_+\left(\gamma((X^{n+1/2}/L_x)^{2m} + (Y^{n+1/2}/L_y)^{2m} - 1) - \lambda\right)^2 - \lambda^2\right). \end{aligned} \tag{56}$$

After eliminating the future $(n+1)$-level variables and multiplying by Δt, the resulting discrete system in terms of the midpoint unknowns reads

$$
\begin{aligned}
&(X^{n+1/2}-X^n)-\frac{1}{2}\Delta t U^{n+1/2}=0 && \text{(57a)}\\
&(Y^{n+1/2}-Y^n)-\frac{1}{2}\Delta t V^{n+1/2}=0 && \text{(57b)}\\
&(U^{n+1/2}-U^n)+\frac{1}{2}\frac{2m}{L_x}(X^{n+1/2}/L_x)^{2m-1}\Delta t F_+(q)F_+'(q)=0 && \text{(57c)}\\
&(V^{n+1/2}-V^n)+\frac{1}{2}\frac{2m}{L_y}(Y^{n+1/2}/L_x)^{2m-1}\Delta t F_+(q)F_+'(q)=0 && \text{(57d)}\\
&\lambda+F_+(q)F_+'(q)=0 && \text{(57e)}\\
q=&\gamma((X^{n+1/2}/L_x)^{2m}+(Y^{n+1/2}/L_y)^{2m}-1)-\lambda, && \text{(57f)}
\end{aligned}
$$

after scaling the λ equation with the large factor $\gamma/\Delta t$.

This time-discrete variational principle is directly implemented in Firedrake, whereupon the five weak formulations are generated automatically. An unconstrained modified-midpoint time discretisation of (48) is also implemented, while using the approximate solution for λ in order to find a classical Hamiltonian system. These two discretisation work in either X or Y direction with initial conditions $X(0)=Y(0)=0, U(0)=0$ or $V(0)=0$, but fail to converge when both coordinate directions are involved (default Firedrake solution solver settings have been used). Finally, an SE time discretisation for the unconstrained billiard dynamics (48) defined as weak formulations yields the results in Fig. 5. It needs a small time step, otherwise the particle just passes the squircle boundary.

Firedrake's solver settings used for the MMP discetisations are the following:

```
solver_parameters6 = {
    'mat_type': 'nest',
}
```

Using (12b), Brown's modified-midpoint time discretisation of (46) is

$$
\begin{aligned}
0=\sum_{n=1}^{N} & U^{n+1/2}\frac{(X^{n+1}-X^n)}{\Delta t}+V^{n+1/2}\frac{(Y^{n+1}-Y^n)}{\Delta t}-\frac{1}{2}({U^{n+1/2}}^2+{V^{n+1/2}}^2)\\
&-\frac{1}{2\gamma}\Big(F_+\Big(\gamma((X^n/L_x)^{2m}+(Y^n/L_y)^{2m}-1)-\lambda^n\Big)^2-(\lambda^n)^2\Big). \qquad \text{(58)}
\end{aligned}
$$

After variations with respect to $\{X^n, Y^n, U^{n+1/2}, V^{n+1/2}\}$, the resulting discrete system in terms of the midpoint unknowns reads

$$
\begin{aligned}
&\lambda^n+F_+(q^n)F_+'(q^n)=0 && \text{(59a)}\\
q^n=&\gamma((X^n/L_x)^{2m}+(Y^n/L_y)^{2m}-1)-\lambda^n, && \text{(59b)}\\
&(U^{n+1/2}-U^{n-1/2})+\frac{2m}{L_x}(X^n/L_x)^{2m-1}\Delta t F_+(q^n)F_+'(q^n)=0 && \text{(59c)}\\
&(V^{n+1/2}-V^{n-1/2})+\frac{2m}{L_y}(Y^n/L_y)^{2m-1}\Delta t F_+(q^n)F_+'(q^n)=0 && \text{(59d)}\\
&(X^{n+1}-X^n)-\Delta t U^{n+1/2}=0 && \text{(59e)}\\
&(Y^{n+1}-Y^n)-\Delta t V^{n+1/2}=0 && \text{(59f)}\\
&\lambda^{n+1}+F_+(q^{n+1})F_+'(q^{n+1})=0 && \text{(59g)}\\
q^{n+1}=&\gamma((X^{n+1}/L_x)^{2m}+(Y^{n+1}/L_y)^{2m}-1)-\lambda^{n+1}, && \text{(59h)}
\end{aligned}
$$

after scaling the λ^n equation with the large factor $\gamma/\Delta t$.

4.2.1 Energy conserving discretisation

Extending Quispel and Turner's method [11], a time discretisation is sought such that

$$E = H(X^{n+1}, Y^{n+1}, U^{n+1}, V^{n+1}, \lambda^{n+1}) - H(X^n, Y^n, U^n, V^n, \lambda^n) = 0. \tag{60}$$

Defining $Z \equiv (X, Y, U, V)^T$, a discrete gradient $\bar{\boldsymbol{\nabla}} H$ of the Hamiltonian is defined such that the vector product

$$\bar{\boldsymbol{\nabla}} H(Z^n, \lambda^n; Z^{n+1}, \lambda^{n+1}) \cdot \begin{pmatrix} X^{n+} - X^n \\ Y^{n+} - Y^n \\ U^{n+} - U^n \\ V^{n+} - V^n \\ \lambda^{n+2} - \lambda^n \end{pmatrix} = \\ H(X^{n+1}, Y^{n+1}, U^{n+1}, V^{n+1}, \lambda^{n+1}) - H(X^n, Y^n, U^n, V^n, \lambda^n). \tag{61}$$

This gradient is defined such that (61) is valid via a telescoping sum; i.e. the gradient $\bar{\boldsymbol{\nabla}} H_1 = \bar{\boldsymbol{\nabla}} H(Z^n, \lambda^n; Z^{n+1}, \lambda^{n+1})$ taken reads

$$\bar{\boldsymbol{\nabla}} H_1 = \begin{pmatrix} \frac{(H(X^{n+1}, Y^n, U^n, V^n, \lambda^n) - H(X^n, Y^n, U^n, V^n, \lambda^n))}{(X^{n+1} - X^n)} \\ \frac{(H(X^{n+1}, Y^{n+1}, U^n, V^n, \lambda^n) - H(X^{n+1}, Y^n, U^n, V^n, \lambda^n))}{(Y^{n+1} - Y^n)} \\ \frac{(H(X^{n+1}, Y^{n+1}, U^{n+1}, V^n, \lambda^n) - H(X^{n+1}, Y^{n+1}, U^n, V^n, \lambda^n))}{(U^{n+1} - U^n)} \\ \frac{(H(X^{n+1}, Y^{n+1}, U^{n+1}, V^{n+1}, \lambda^n) - H(X^{n+1}, Y^{n+1}, U^{n+1}, V^n, \lambda^n))}{(V^{n+1} - V^n)} \\ \frac{(H(X^{n+1}, Y^{n+1}, U^{n+1}, V^{n+1}, \lambda^{n+1}) - H(X^{n+1}, Y^{n+1}, U^{n+1}, V^{n+1}, \lambda^n))}{(\lambda^{n+1} - \lambda^n)} \end{pmatrix}. \tag{62}$$

Hence, defining $Z \equiv (X, Y, U, V)^T$ one finds that

$$\begin{aligned} E =& H(Z^{n+1}, \lambda^{n+1}) - H(Z^n, \lambda^n) && \text{(63a)} \\ =& (H(Z^{n+1}, \lambda^{n+1}) - H(Z^{n+1}, \lambda^n) + \bar{\boldsymbol{\nabla}}_Z H \cdot (Z^{n+1} - Z^n) && \text{(63b)} \\ =& (H(Z^{n+1}, \lambda^{n+1}) - H(Z^{n+1}, \lambda^n) + \bar{\boldsymbol{\nabla}}_Z H J \bar{\boldsymbol{\nabla}}_Z H && \text{(63c)} \\ =& (H(Z^{n+1}, \lambda^{n+1}) - H(Z^{n+1}, \lambda^n) && \text{(63d)} \\ =& 0, && \text{(63e)} \end{aligned}$$

when the equation for λ^{n+1} is taken as $H(Z^{n+1}, \lambda^{n+1}) - H(Z^{n+1}, \lambda^n) = 0$ or

$$F_+(-\gamma G(X^{n+1}, Y^{n+1}) - \lambda^{n+1})^2 - (\lambda^{n+1})^2 = F_+(-\gamma G(X^{n+1}, Y^{n+1}) - \lambda^n)^2 - (\lambda^n)^2 \tag{63f}$$

and with skew-symmetric 4×4 cosymplectic matrix J. A second-order method emerges when one takes

$$\bar{\boldsymbol{\nabla}} H_2 = \frac{1}{2} \left(\bar{\boldsymbol{\nabla}} H(Z^n, \lambda^n; Z^{n+1}, \lambda^{n+1}) + \bar{\boldsymbol{\nabla}} H(Z^{n+1}, \lambda^{n+1}; Z^n, \lambda^n) \right). \tag{64}$$

Then the equation for λ^{n+1} becomes $H(Z^{n+1}, \lambda^{n+1}) + H(Z^n, \lambda^{n+1}) - H(Z^{n+1}, \lambda^n) - -H(Z^n, \lambda^n) = 0$ or

$$F_+(-\gamma G(X^{n+1}, Y^{n+1}) - \lambda^{n+1})^2 - (\lambda^{n+1})^2 + F_+(-\gamma G(X^n, Y^n) - \lambda^{n+1})^2 - (\lambda^{n+1})^2 + \\ - F_+(-\gamma G(X^{n+1}, Y^{n+1}) - \lambda^n)^2 - (\lambda^n)^2 - F_+(-\gamma G(X^n, Y^n) - \lambda^n)^2 - (\lambda^n)^2 = 0. \tag{65}$$

These equations for λ^{n+1} are an approximation of $\partial H/\partial\lambda = 0$ and we note that the expression $\bar{\boldsymbol{\nabla}}_Z H J \bar{\boldsymbol{\nabla}}_Z H$ only involves λ^n while $\bar{\boldsymbol{\nabla}}_Z H_2 J \bar{\boldsymbol{\nabla}}_Z H$ involves λ^n and λ^{n+1} in a symmetric manner. Quispel and Turner [11] point out that with Yoshida's method integrators of arbitrary order can be constructed with this (first-order or second-order) method as building block. The method allows a variable timestep since each time step is self-contained energy conserving.

5 Inequality constraint in time-continuous infinite-dimensional canonical variational principles

5.1 Hydrostatic rest flow with triangular buoy in one horizontal dimension

5.2 Dynamic VBM model with with triangular buoy in one horizontal dimension

6 Summary and discussion

References

[1] M.S. Alnaes, UFL: a finite element form language. In: Automated Solution of Differential Equations by the Finite Element Method. Eds: A. Logg and K.-A. Mardal and G. N. Wells, Springer, Berlin. 2011.

[2] M.S. Alnaes, A. Logg, K.B. Oelgaard, M.E. Rognes, G.N. Wells G.N., Unified Form Language: A domain-specific language for weak formulations of partial differential equations. 2013. https://arxiv.org/pdf/1211.4047.pdf

[3] S. Balay, et al., PETSc, Toolkit for Advanced Optimisation: optimisation solvers, bound-constraint optimisation. 2023.

[4] J.D. Brown, Midpoint rule as a variational-symplectic integrator: Hamiltonian systems. Physical Review D 73, 2006, 2400.

[5] E. Bueler PETSc for Partial Differential Equations: Numerical Solutions in C and Python. Inequality constraints. Chapter 12, 2020. https://doi.org/10.1137/1.9781611976311.ch12

[6] E. Burman, P. Hansbo, M.G. Larson, The augmented Lagrangian method as a framework for stabilised methods in computational mechanics. Archives of Computational Methods in Engineering 30, 2023, 2579—2604.

[7] J. Choi, A. Kalogirou, A., Y. Lu, O. Bokhove, M. Kelmanson, A study of extreme water waves using a hierarchy of models based on potential-flow theory. Water Waves **6**, 2024, 225–277.

[8] E. Gagarina, Variational approaches to water wave simulations. PhD Thesis, University of Twente, 2014. https://ris.utwente.nl/ws/portalfiles/portal/6031258/thesis_E_Gagarina.pdf

[9] Y. Lu, F. Gidel, J. Choi, O. Bokhove, M. A. Kelmanson, 2025. Variational numerical-modelling strategies for the simulation of driven free-surface waves. To

be submitted. Preliminary version: `https://eartharxiv.org/repository/view/3411/`

[10] F. Rathgeber et al., Firedrake: automating the finite element method by composing abstractions. ACM Trans. Math. Softw. 43(3), 2016, 27.

[11] G.R.W. Quispel, G.S. Turner, 1996. Discrete gradient methods for solving ODEs numerically while preserving a first integral. *J. Phys. A: Math. Gen.* **29** L341.

[12] H. Yoshida, 1990. Construction of higher order symplectic integrators. *Phys. Lett.* **150A**, 262.

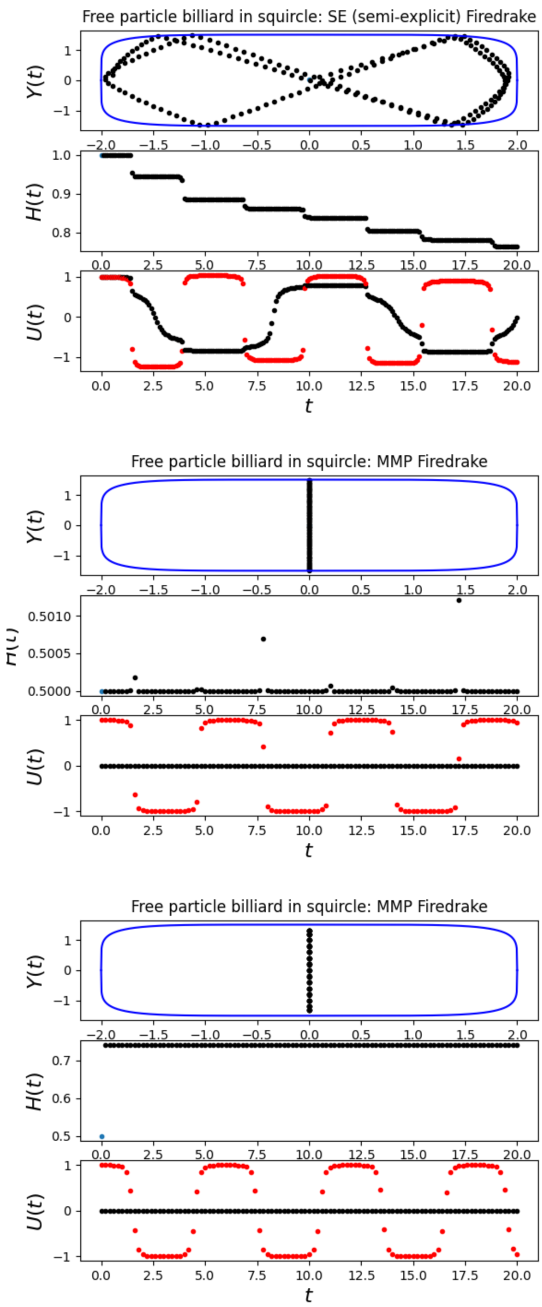

Figure 5: Top: SE discretisation results of unconstrained Hamiltonian billiard dynamics; $m = 6$, $\Delta t = 0.0001, t \in [0, 20]$. Bottom: MMP discretisations of constrained and unconstrained billiard dynamics. Firedrake implementations with standard solver parameters for MMP; SE is semi-explicit. Parameters $\gamma = 100$, top: $b = 0.68\sqrt{\gamma}$ and bottom: b is four time larger, $\Delta t = 0.01, t \in [0, 20]$. *Made with codes FD: wavesebilliard2024.py runs but needs small timestep; wavemmpbilliard2024.py MMP unconstrained fails if not aligned in 1D; wavemmpbilliardlamb2025.py: MMP fully VP fails if not aligned in 1D, same results.*