European Society of Computational Methods in Sciences and Engineering (ESCMSE)

Journal of Numerical Analysis, Industrial and Applied Mathematics (JNAIAM)
vol. 1, no. 1, 2006, pp. 113-130
ISSN 1790–8140

Discrete Conservative Vector Fields Induced by the Trapezoidal Method

Felice Iavernaro* and Donato Trigiante†

* *Dipartimento di Matematica, Università di Bari, Via Orabona 4, I-70125 Bari (Italy), felix@dm.uniba.it, fax:+39 080 5460612*
† *Dip. di Energetica, Università di Firenze, via C. Lombroso 6/17, I-50134 Firenze (Italy), trigiant@unifi.it.*

Received 3 January, 2006; accepted in revised form 7 March, 2006

Abstract: The use of symmetric schemes has revealed interesting stability properties for the long time simulation of conservative, and in particular Hamiltonian systems. By using one of the simplest symmetric formulae, namely the trapezoidal method, we show that, under certain circumstances, one can attach to the discrete system a *discrete energy function* and therefore obtain a discrete conservative dynamical system.

Keywords: Conservative vector fields, Hamiltonian systems, symplecticity, symmetric methods.

Mathematics Subject Classification: 65P10, 65L05, 37M15.

1 Introduction

There is a great deal of experimental evidence that symmetric methods display good stability properties when applied to different classes of dynamical systems that possess a first integral (generally the energy function)[3, 9]. A systematic analytical study to give a theoretical account for their good long time behavior is more recent and generally based upon a backward error analysis approach. This means that one tries to fit a continuous system (the so called modified equation) to the given numerical method, to an arbitrary order of accuracy, and then retrieves information on the numerical solution by looking at the continuous problem. This technique has proven successful for the investigation of symplectic methods and later on it has been extended to the study of symmetric methods [1, 5, 6].

For many interesting situations, the diagram that displays the energy function evaluated at the points of the numerical solution versus the time is similar to that reported in Figure 1 where the trapezoidal method has been applied to solve the nonlinear pendulum equation ($H(p,q) = 1/2p^2 - cos(q)$). The sequence $H(p_n, q_n)$ undergoes regular bounded oscillations whose amplitude is proportional to h^p, where h is the stepsize of integration and p the order of the method (in [4] the authors find out a condition symmetric schemes must satisfy in order to share such behavior).

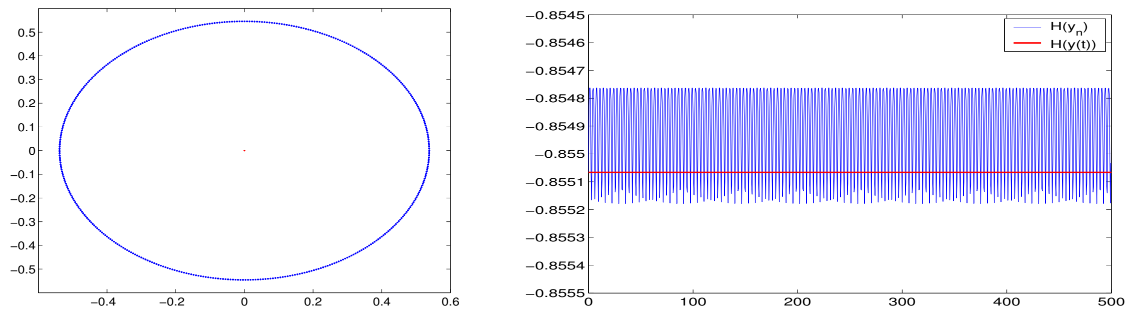

Figure 1: The trapezoidal method applied to solve the nonlinear pendulum equation nearly preserves the energy of the continuous system (right plot), as a comparison between $H(y_n)$ and $H(y(t))$ reveals. Initial condition: $(p_0, q_0) = (0.5, 0.2)$; integration time interval: $[0, 500]$; stepsize $h = 0.5$.

A more specific issue that one can pose is whether, to the discrete dynamical system obtained by the application of a given numerical method to a conservative (in particular Hamiltonian) continuous system, it is possible to attach a term that is *exactly* preserved by the numerical solution and may be interpreted as a *discrete energy* term that is $O(h^p)$-close to the original one. The present paper focusses on this aspect. In the next section we introduce the discrete line integral induced by symmetric schemes and the definition of conservative discrete vector fields. Subsequently we confine our investigation to 2D systems and to the trapezoidal method: this choice will make the analysis simpler without diminishing the richness of the problem.

2 Discrete conservative vector fields

We consider a continuously differentiable vector field $f = [f_1(y), f_2(y), \dots, f_r(y)]^T$ defined on $\Omega \subset \mathbb{R}^r$ and its projection $\{f(y_i)\}$ onto an orthogonal grid $D \subset \Omega$. D is made up of a finite number of points $y_i = [y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(r)}]^T$ defined as the intersection of orthogonal hyperplanes of dimension $r-1$ spanned by $r-1$ unit vectors $e_i = [0, 0, \dots, 1, \dots, 0]^T$.

The generic cell of the grid is a hypercube identified by its 2^r vertices $y_i + \sum_{j=1}^n \omega_j h_{ij} e_j$ where h_{ij} is the length of the jth edge, and ω_j is either 0 or 1 (the left picture of Figure 2 plots a cell in $\mathbb{R}^3$). Two adjacent points on the grid are two points belonging to the same cell, while an oriented path γ on D will be a sequence of adjacent points: $\gamma = (y_1, y_2, \dots, y_n)$ (if $y_n = y_1$, γ defines a closed path). γ may also be thought of as a path on Ω by joining each couple of adjacent vertices (y_i, y_{i+1}) by means of a continuous curve which, for the moment, we assume to be the segment $\gamma_i(t) = y_i + t(y_{i+1} - y_i)$ (central plot of Figure 2).

As is well known, the line integral of the continuous vector field on the segment γ_i is defined as

$$\int_{\gamma_i} f \cdot dy \equiv \int_0^1 \dot{\gamma}_i(t)^T f(\gamma_i(t)) dt = \int_0^1 (y_{i+1} - y_i)^T f(\gamma_i(t)) dt, \tag{1}$$

while the line integral on the path γ (which is indeed a broken line) is retrieved by summing up the above integrals along the segments γ_i:

$$\int_{\gamma} f \cdot dy \equiv \sum_{i=1}^{n-1} \int_{\gamma_i} f \cdot dy. \tag{2}$$

Approximating the integral on the right hand side of (1) by means of a quadrature formula, allows us to introduce the definition of discrete line integral on the path γ:

$$\sum_{\gamma} f \cdot \Delta y \equiv \sum_{i=1}^{n-1} \left((y_{i+1} - y_i)^T \sum_{j=0}^{\ell} A_j f(y_i + c_j(y_{i+1} - y_i)) \right), \tag{3}$$

where A_j are the weights of the quadrature formula corresponding to the abscissae $c_j \in [0, 1]$.

One property of (2) we want to be retained by (3) is its reversal symmetry, that is, reversing the orientation of the path only yields a change of sign in the discrete line integral:

$$\sum_{-\gamma} f \cdot \Delta y = -\sum_{\gamma} f \cdot \Delta y, \tag{4}$$

where $-\gamma$ is the path obtained by reversing the order of the vertices: $-\gamma = (y_n, y_{n-1}, \dots, y_1)$. It is easy to verify that this condition, for a generic vector field, is equivalent to the symmetry of the quadrature formula:

$$c_j = 1 - c_{\ell-j}, \quad A_j = A_{\ell-j}, \qquad j = 1, \dots, \ell.$$

For example the trapezoidal rule yields

$$\sum_{-\gamma_i} f \cdot \Delta y = \frac{1}{2}(y_i - y_{i+1})^T (f(y_{i+1}) + f(y_i)) = -\frac{1}{2}(y_{i+1} - y_i)^T (f(y_i) + f(y_{i+1})) = -\sum_{\gamma_i} f \cdot \Delta y$$

while the rectangle formula would give

$$\sum_{\gamma_i} f \cdot \Delta y = (y_{i+1} - y_i)^T f(y_i) \qquad \text{and} \qquad \sum_{-\gamma_i} f \cdot \Delta y = (y_i - y_{i+1})^T f(y_{i+1}).$$

The importance of the above property and the consequent choice of symmetric quadrature formulae in the definition (3) is well understood by considering the following definition.

Definition 2.1 *A discrete vector field is conservative if its circulation vanishes over any closed path:*

$$\sum_{\gamma} f \cdot \Delta y = 0, \qquad \text{for any closed path } \gamma. \tag{5}$$

Choosing as a closed path the path $\sigma = \gamma \cup -\gamma$, we see that a necessary condition for a vector field to be conservative is that the underlying quadrature formula be symmetric.

We remark that, apart from the trapezoidal rule (for which $\ell = 1,\ c_0 = 0,\ c_1 = 1$), (3) requires the evaluation of f at points $y_i + c_j(y_{i+1} - y_i)$ not belonging to D, located over the segment (y_i, y_{i+1}). Such extra points give rise to a finer grid $\tilde{D}$ which is embedded in the coarser grid D: they are solely used as a tool to retrieve a more accurate approximation in the quadrature formula.

Of course, generalizations of the discrete line integral (3) are easily obtained by acting on either the choice of the curve γ_i or the specific quadrature formulae. For example, one possibility is to allow the points of the internal grid $\tilde{D}$ not to be aligned and to take γ_i as the broken line joining such points (see the right plot of Figure 2). Another important consequence of property (4) (independent of whether or not the vector field is conservative) is that the circulation along any arbitrary closed path may be seen as a sum of circulations on elementary closed paths made up of triangles, because all the sides γ_i that do not belong to the given contour, contribute with both the terms $\sum_{\gamma_i} f \cdot \Delta y$ and $\sum_{-\gamma_i} f \cdot \Delta y$ (see Figure 3). This allows us to confine our investigation to such elementary contours to verify properties that are valid for any wider contour. To a certain extent, this is the analogue of what happens in the continuous case when the local character of

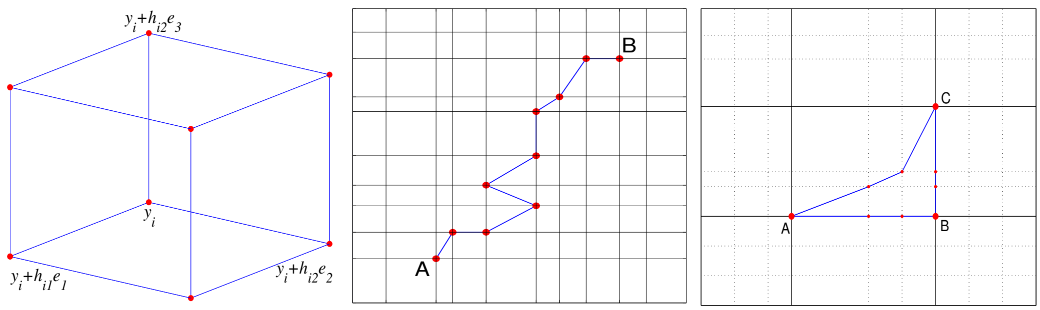

Figure 2: Left plot: a cell in $\mathbb{R}^3$. Central plot: a 2-dimensional path on D (dots) and on Ω (broken line) joining the points $A, B \in D$. Right plot: a triangle with two internal abscissae on each side (small dots); the solid lines represent the coarse grid D while the dotted lines form the internal grid $\tilde{D}$.

the vector field (its divergence, curl, etc.), is used to investigate its behaviour over a given finite region. Even better, the classical approach (see for example [2]) is that of first computing the flow (or the circulation) over infinitesimal regions (contours) where the vector field may be assumed of constant value and then obtaining its local properties as the limit of a certain discretization parameter that makes the region shrink to a point.

In addition, the circulation over an arbitrary triangle in $\mathbb{R}^r, r > 2$, may be computed as sums of circulations of triangles in $\mathbb{R}^{r-1}$. This is a useful tool also to extend properties proven in a given dimension (in the simplest case in $\mathbb{R}^2$), to a higher dimension simply using an inductive procedure as the following example shows.

Example 2.2 *Let us compute the circulation of a three-dimensional vector field* $f = [f_1(p,q,s), f_2(p,q,s), \; f_3(p,q,s)]^T$ *over the triangle ABC of Figure 3 (right plot). As already remarked, we could work separately on the triangles OBC, OCA and OAB and then sum up the three contributions. In any case, assuming* $O(p,q,s)$, $A(p,q,s+\Delta s)$, $B(p+\Delta p, q, s)$, $C(p, q+\Delta q, s)$, *we get*

$$\sum_\gamma f \cdot \Delta y = \frac{\Delta q \Delta s}{2}\left[\frac{\Delta_q f_3}{\Delta q} - \frac{\Delta_s f_2}{\Delta s}\right] + \frac{\Delta p \Delta s}{2}\left[\frac{\Delta_s f_1}{\Delta s} - \frac{\Delta_p f_3}{\Delta p}\right] + \frac{\Delta p \Delta q}{2}\left[\frac{\Delta_p f_2}{\Delta p} - \frac{\Delta_q f_1}{\Delta q}\right],$$

which we recognize to be the discrete analogue of Stokes' theorem after defining the discrete curl vector of a three-dimensional discrete vector field $F = [F_1, F_2, F_3]^T$ *on the grid D as the vector with orthogonal components:*

$$(curlF)_p = \frac{\Delta_q F_3}{\Delta q} - \frac{\Delta_s F_2}{\Delta s}, \quad (curlF)_q = \frac{\Delta_s F_1}{\Delta s} - \frac{\Delta_p F_3}{\Delta p}, \quad (curlF)_s = \frac{\Delta_p F_2}{\Delta p} - \frac{\Delta_q F_1}{\Delta q}.$$

Of course also in this case one has to repeat the computation over all the possible triangles that may occur in an arbitrary closed path.

The next two subsections show better the closeness between the discrete and continuous approaches. We confine our study to the trapezoidal formula because it contains all the most important features of the other symmetric formulae, allowing an easier understanding of the behaviour

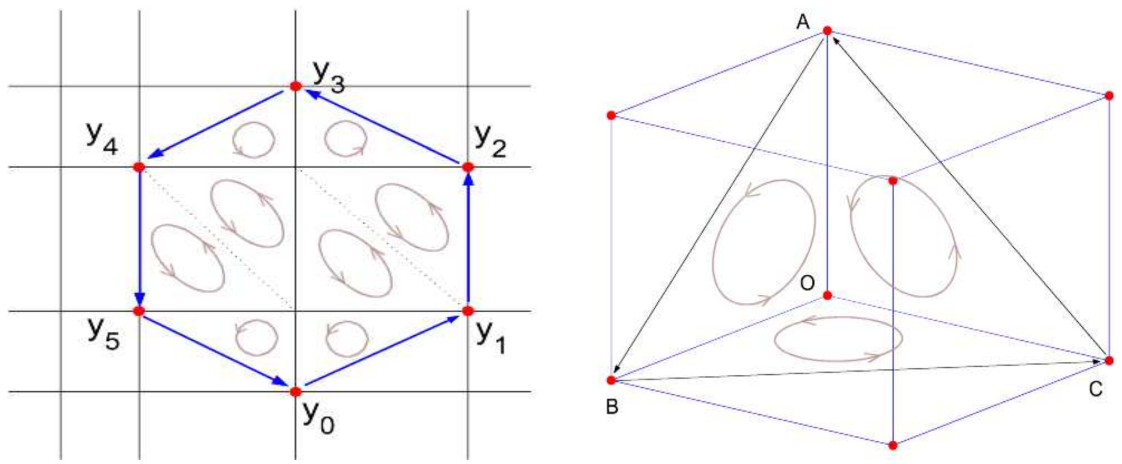

Figure 3: Adding up the individual circulations of suitable triangles, results in a net circulation over a given contour. In the right picture the circulation over the triangle ABC, embedded in $\mathbb{R}^3$ is computed as the sum of the circulations of three triangles embedded in $\mathbb{R}^2$: OAB, OBC and OCA.

of the associated discrete vector field. For this method, (3) reads

$$\sum_{\gamma} f \cdot \Delta y = \frac{1}{2} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^T (f(y_i) + f(y_{i+1})). \tag{6}$$

3 Two dimensional vector fields

Let us compute the circulation of a two-dimensional vector field $f = [f_1(p,q), f_2(p,q)]^T$ over the generic triangle T_1 having a contour $\gamma_1 = (y_0, y_1, y_2, y_0)$ with vertices $y_0 = (p,q)$, $y_1 = (p+\Delta p, q)$, $y_2 = (p, q+\Delta q)$ (see Figure 4). Applying formula (6) yields

$$\begin{aligned}\sum_{\gamma_1} f \cdot \Delta y = \; & \frac{1}{2} [\Delta p(f_1(y_0) + f_1(y_1)) - \Delta p(f_1(y_1) + f_1(y_2)) + \\ & \Delta q(f_2(y_1) + f_2(y_2)) - \Delta q(f_2(y_2) + f_2(y_0))] \\ = \; & \frac{\Delta p \Delta q}{2} \left[\frac{\Delta_p f_2}{\Delta p} - \frac{\Delta_q f_1}{\Delta q} \right],\end{aligned}$$

where $\Delta_p f(p,q) = f(p+\Delta p, q) - f(p,q)$ and $\Delta_q f(p,q) = f(p, q+\Delta q) - f(p,q)$ are the increments of f along the orthogonal directions p and q evaluated at y_0, respectively.

We have obtained the analogue of the divergence (or Gauss) theorem for continuous vector fields. This may be understood by considering that the scalar product $f \cdot \Delta y$ coincides with $F \cdot \Delta n$ where $F \equiv [F_1, F_2]^T = [-f_2, f_1]^T$ and $\Delta n = [-\Delta y_2, \Delta y_1]^T$ are orthogonal to f and Δy respectively and therefore Δn lies in the orthogonal direction of the path (we can assume that it defines an outward direction with respect to the closed oriented curve γ because, otherwise, we simply need to change the signs of Δn and F). Therefore, by defining the net normal flow of a discrete 2-dimensional vector field $F(p,q)$ through the region S bounded by a closed path γ as:

$$\sum_{\gamma} F \cdot \Delta n \equiv \sum_{\gamma} JF \cdot \Delta s, \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{7}$$

that is as the circulation of the vector JF along the path γ, for the triangle of this example we arrive at

$$\sum_{\gamma} F\cdot\Delta n = \frac{\Delta p\Delta q}{2}\left[\frac{\Delta_p F_1}{\Delta p}+\frac{\Delta_q F_2}{\Delta q}\right],$$

which coincides with the discrete version of the Gauss theorem, provided that we define the discrete divergence operator on the grid D as:

$$\mathrm{div}F = \frac{\Delta_p F_1}{\Delta p}+\frac{\Delta_q F_2}{\Delta q}.$$

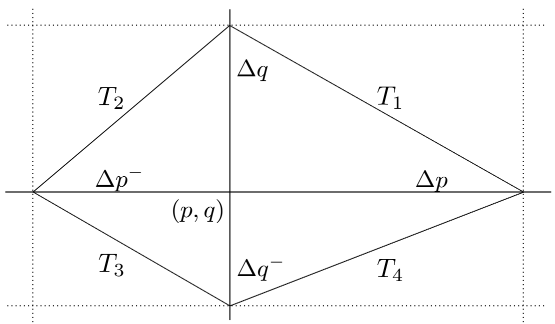

Figure 4: The four shapes of a triangle.

Of course, one obtains the formulation of the Gauss theorem corresponding to an arbitrary region enclosed in a given contour, by just summing up the individual flows through all the triangles that form a partition of the region itself and canceling out all the terms related to the internal segments. To this end, we notice that any region interior to a grid, may be partitioned into triangles whose shapes are like those of the triangles $T_i,\ i=1,\dots,4$ in Figure 4. Repeating the computation of the circulation on the boundaries γ_i of $T_i,\ i=2,3,4$, we arrive at the following four formulas pertaining a generic point (p,q) of the grid:

$$\sum_{\gamma_1} f\cdot\Delta y = \frac{\Delta p\Delta q}{2}\left[\frac{\Delta_p f_2}{\Delta p}-\frac{\Delta_q f_1}{\Delta q}\right],\qquad \sum_{\gamma_2} f\cdot\Delta y = \frac{\Delta p^-\Delta q}{2}\left[\frac{\Delta_p^- f_2}{\Delta p^-}-\frac{\Delta_q f_1}{\Delta q}\right],$$
$$\sum_{\gamma_3} f\cdot\Delta y = \frac{\Delta p^-\Delta q^-}{2}\left[\frac{\Delta_p^- f_2}{\Delta p^-}-\frac{\Delta_q^- f_1}{\Delta q^-}\right],\qquad \sum_{\gamma_4} f\cdot\Delta y = \frac{\Delta p\Delta q^-}{2}\left[\frac{\Delta_p f_2}{\Delta p}-\frac{\Delta_q^- f_1}{\Delta q^-}\right],$$

where, for example, $\Delta_p^- f_2(p,q) = f_2(p,q)-f_2(p-\Delta p^-,q)$. Conservativity of the vector field is then equivalent to the relations

$$\frac{\Delta_p f_2}{\Delta p}=\frac{\Delta_q f_1}{\Delta q},\qquad \frac{\Delta_p^- f_2}{\Delta p^-}=\frac{\Delta_q f_1}{\Delta q},\qquad \frac{\Delta_p f_2}{\Delta p}=\frac{\Delta_q^- f_1}{\Delta q^-},\qquad \frac{\Delta_p f_2}{\Delta p}=\frac{\Delta_q^- f_1}{\Delta q^-},$$

which, in turn, lead to the following characterization of two-dimensional conservative vector fields.

Theorem 3.1 *A vector field $f=[f_1(p,q),f_2(p,q)]^T$ is conservative with respect to the grid D if it satisfies*

(a) $\dfrac{\Delta_q f_1}{\Delta q}=\dfrac{\Delta_p f_2}{\Delta p}$,

(b) $\dfrac{\Delta_q f_1}{\Delta q}=\dfrac{\Delta_q^- f_1}{\Delta q^-}$ *and* $\dfrac{\Delta_p f_2}{\Delta p}=\dfrac{\Delta_p^- f_2}{\Delta p^-}$,

for each point $(p,q)\in D$ for which the above operations are well-defined (internal grid points).

Condition *(b)* requires that the left and right increments of f_1 and f_2 along the q and p directions respectively are the same, while condition *(a)* is the discrete analogue of $\partial f_1/\partial q=\partial f_2/\partial p$ which corresponds to the closeness of the differential form $f_1(p,q)dp + f_2(p,q)dq$ (provided f is of class C^1). Hereafter we gather a number of properties valid for regular conservative continuous vector fields, and then we determine their discrete counterpart (see Theorem 3.2). Our aim is to state that, as is the case for the continuous setting, a conservative discrete vector field always comes from a suitable potential function.

A closed first-order differential form with domain an open, simply connected set $\Omega \subset \mathbb{R}^2$ is exact, that is its circulation vanishes over any closed oriented curve $\gamma \subset \Omega$. In particular, the four following conditions are equivalent (here and in the sequel we assume regularity of f; ∇ denotes the gradient operator):

(i) $\dfrac{\partial f_1}{\partial q} = \dfrac{\partial f_2}{\partial p}$;

(ii) $\displaystyle\int_\gamma \big(f_1(p,q)dp + f_2(p,q)dq\big) = 0$ for every closed oriented curve $\gamma \subset \Omega$;

(iii) there exists a scalar function $H(p,q)$ defined on Ω such that
$$\nabla^T H(p,q) = [f_1(p,q), f_2(p,q)];$$

(iv) there exists a scalar function $H(p,q)$ defined on Ω such that
$$\frac{H(p+\Delta p,q) - H(p,q)}{\Delta p} = \int_0^1 f_1(p+t\Delta\, p,q)dt,$$
and
$$\frac{H(p,q+\Delta q) - H(p,q)}{\Delta q} = \int_0^1 f_2(p,q+t\Delta\, q)dt,$$
for every $\Delta p,\ \Delta q$ such that $(p+\Delta p, q+\Delta q) \in \Omega$.

Properties (ii)–(iv) are three equivalent ways to express exactness of a differential form. The usefulness of the equivalence between (iii) and (iv) will be clear hereafter. It is an easy matter to find out the corresponding equivalences valid for the discrete case (we analyze the two dimensional case here even if the generalization to higher dimensions is straightforward). We begin with defining a difference form as a function with domain which is a grid $D \subset \Omega$ and a range which is the set of linear transformations from $\mathbb{R}^2$ to $\mathbb{R}$, that we denote by $f_1(p,q)\Delta p + f_2(p,q)\Delta q$. Defining the line integral of a difference form (induced by the trapezoidal rule) like in (6), closeness as the conditions *(a)* and *(b)* of Theorem 3.1, we arrive at the following result.

Theorem 3.2 *For a difference form $\omega = f_1(p,q)\Delta p + f_2(p,q)\Delta q$ defined on a grid $D \subset \Omega$, with Ω simply connected, the following conditions are equivalent:*

$(i)'$ *ω is closed;*

$(ii)'$ $\displaystyle\sum_\gamma \omega = 0$ *for every closed path $\gamma \subset D$;*

$(iii)'$ *there exists a scalar function $\widetilde{H}(p,q)$ defined on the grid D such that*
$$\frac{\widetilde{H}(p+\Delta p,q) - \widetilde{H}(p,q)}{\Delta p} = \frac{1}{2}\left(f_1(p,q) + f_1(p+\Delta p,q)\right),$$
and
$$\frac{\widetilde{H}(p,q+\Delta q) - \widetilde{H}(p,q)}{\Delta q} = \frac{1}{2}\left(f_2(p,q) + f_2(p,q+\Delta q)\right),$$
for every (positive or negative) $\Delta p,\ \Delta q$, such that $(p+\Delta p, q+\Delta q) \in D$.

Proof. The equivalence between $(i)'$ and $(ii)'$ comes from Theorem 3.1. Fixing a point $(p_0, q_0) \in D$, the scalar function $\widetilde{H}(p, q)$ is naturally defined as

$$\widetilde{H}(p, q) = C + \sum_{\gamma} \omega,$$

where γ is any path on the grid joining the points (p_0, q_0) and (p, q) and C is any constant number ($\widetilde{H}(p, q)$ is well defined due to $(ii)'$). Then, considering the segments $\rho = [(p, q), (p + \Delta p, q)]$ and $\sigma = [(p, q), (p, q + \Delta q)]$, we get

$$\widetilde{H}(p + \Delta p, q) - \widetilde{H}(p, q) = \sum_{\rho} \omega, \qquad \text{and} \qquad \widetilde{H}(p, q + \Delta q) - \widetilde{H}(p, q) = \sum_{\sigma} \omega,$$

that give, by definition, $(iii)'$. The implication $(iii)' \Rightarrow (ii)'$ is straightforward. □

We remark that $(iii)'$ may be interpreted as the discrete counterpart of (iv) via the approximation of the integrals in (iv) by means of the trapezoidal formula. The correspondence between $(iii)'$ and (iii) could be settled by defining the discrete gradient operator as the vector containing the increments of $\widetilde{H}(p, q)$ along the two axes but this is a more ticklish question since, in general, such increments would depend upon the signs of Δp and Δq as shown by the left hand sides of the equalities in $(iii)'$. On the basis of that, we are led to define the four discrete gradient operators (refer again to Figure 4):

$$\begin{aligned}
\nabla^T_{++}\widetilde{H}(p, q) &= \left[\frac{\widetilde{H}(p + \Delta p, q) - \widetilde{H}(p, q)}{\Delta p}, \quad \frac{\widetilde{H}(p, q + \Delta q) - \widetilde{H}(p, q)}{\Delta q}\right],\\
\nabla^T_{-+}\widetilde{H}(p, q) &= \left[\frac{\widetilde{H}(p - \Delta p^-, q) - \widetilde{H}(p, q)}{\Delta p^-}, \quad \frac{\widetilde{H}(p, q + \Delta q) - \widetilde{H}(p, q)}{\Delta q}\right],\\
\nabla^T_{--}\widetilde{H}(p, q) &= \left[\frac{\widetilde{H}(p - \Delta p^-, q) - \widetilde{H}(p, q)}{\Delta p^-}, \quad \frac{\widetilde{H}(p, q - \Delta q^-) - \widetilde{H}(p, q)}{\Delta q^-}\right],\\
\nabla^T_{+-}\widetilde{H}(p, q) &= \left[\frac{\widetilde{H}(p + \Delta p, q) - \widetilde{H}(p, q)}{\Delta p}, \quad \frac{\widetilde{H}(p, q - \Delta q^-) - \widetilde{H}(p, q)}{\Delta q^-}\right].
\end{aligned}$$

Which one of the four operators will come into play, will depend upon the direction in which we are integrating: their role will become clearer starting from the next section.

At least two drawbacks arise when searching for discrete conservativity:

1. in view of Theorem 3.2, the projection of a conservative continuous vector field does not lead, in general, to a conservative discrete vector field. For example, the difference form $(p^2 - q^2)\,\Delta p - 2pq\,\Delta q$, coming from an exact differential form, does not satisfy any of the conditions for closeness;

2. for discrete vector fields the conservativity property depends, in general, on the choice of the grid.

Both the above problems disappear for two important classes of vector fields:

- *Linear vector fields.* If f_1 and f_2 depend linearly on p and q, (iv) and $(iii)'$ represent exactly the same condition, whatever the grid, because the trapezoidal rule is exact for linear functions. It follows as well that $H(p, q) = \widetilde{H}(p, q)$ for any point on the grid: they are scalar quadratic functions in p and q.

- *Separable vector fields.* They are of the form $f_1 = f_1(p)$, $f_2 = f_2(q)$, that is f_1 and f_2 do not depend on q and p respectively. In such a case f satisfies

$$\frac{\partial f_1}{\partial q} = \frac{\partial f_2}{\partial p} = \frac{\Delta_q f_1}{\Delta q} = \frac{\Delta_p f_2}{\Delta p} = \frac{\Delta_q^- f_1}{\Delta q^-} = \frac{\Delta_p^- f_2}{\Delta p^-} = 0,$$

independently of the chosen grid.

In the next section we give a first application of the results presented so far, namely 2D separable Hamiltonian systems. Although we assume these two restrictions (minimal dimension and separability), the results we achieve suggest that this new approach can throw new light on the study of the long time behavior of symmetric methods, because they provide a closer relationship between the techniques employed to investigate the behavior of continuous and discrete dynamical systems.

4 2D discrete Hamiltonian problems

Hamiltonian systems have even dimension and, when this is two, they are defined by imposing that the trajectories they generate are always orthogonal to a conservative vector field or, equivalently, to the gradient of a scalar function (the Hamiltonian function). Such geometric specification applies to both continuous and discrete vector fields even if, as we will see, in this latter case we need an extra ingredient.

In the former case we get $y' = J\nabla H(y)$, with J defined as in (7) and $y = (p, q)^T$, while to investigate the latter we suppose that the state vector at times t_n and t_{n+1} is $y_n = (p, q)^T$ and $y_{n+1} = (p + \Delta p, q + \Delta q)^T$ respectively, where the increments Δp and Δq may be either positive, negative or zero. Then $y_{n+1} - y_n$ gives the local direction of the trajectory. By imposing orthogonality to the gradient of the discrete Hamiltonian function, we obtain $y_{n+1} - y_n = h_n J\nabla\widetilde{H}(y)$ where $\nabla\widetilde{H}(p, q)$ must be chosen according to the signs of Δp and Δq as described in the previous section (h_n is any nonzero real number). Anyway, whatever the signs are, exploiting $(iii)'$ of Theorem 3.2, we finally get

$$y_{n+1} - y_n = \frac{h_n}{2} J \left(\nabla H(y_n) + \nabla H(y_{n+1})\right). \tag{8}$$

Summarizing, the discrete dynamical system defined (in analogy with the continuous case) by imposing the trajectories to be orthogonal to the gradient of the discrete Hamiltonian function, is just the trapezoidal method applied to $y' = J\nabla H(y)$. In (8) we have made explicit the dependance of the stepsize on n to make the approach as general as possible (see the next subsection for details). Can we say that (8) represents a discrete Hamiltonian system? Unfortunately a delicate question arises at this point: nothing can *a priori* assure us that, starting at a point y_0 on a given grid, the subsequent points y_n lie on that grid for any n. Of course one could not fix the grid beforehand, but rather define it as the solution advances in time. But even in this case it is not guaranteed that a grid, in the sense of the definition we gave at the beginning of the paper, might exist. To better elucidate this aspect, we consider again the *nonlinear pendulum equation* ($H(p, q) = 1/2p^2 - cos(q) + 1$ and hence $H(0, 0) = 0$), that from now on will serve as a test problem to carry out the further steps of our approach.

Example 4.1 *We choose $y_0 = (p_0, q_0) = (0.7, 0)$ as initial value. In the left plot of Figure 5, the drawback outlined above is shown: the orbit does not lie on a grid unless the stepsize h is suitably tuned. For example, the right plot of Figure 5 displays a periodic orbit with period* 12. *In the case of periodic orbits lying on a grid (and therefore closing after one cycle), the expression for*

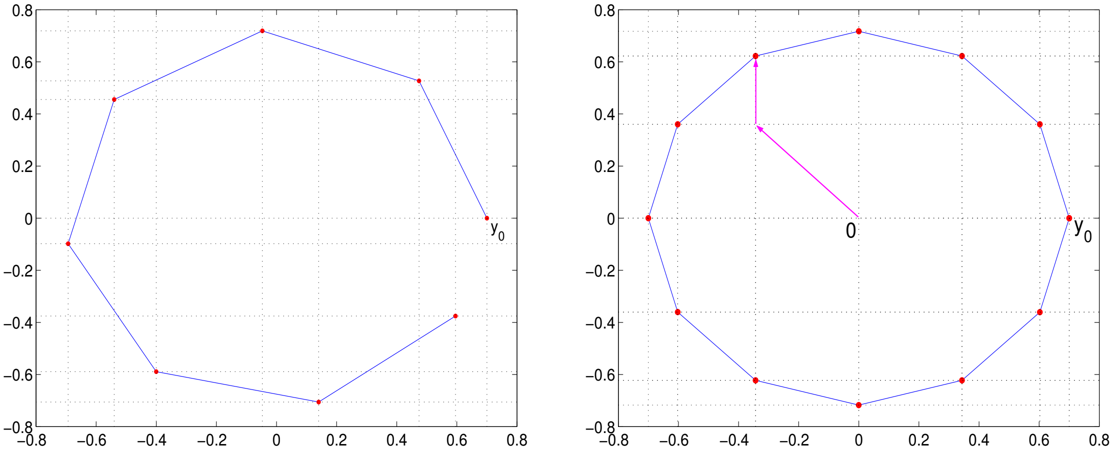

Figure 5: Nonlinear pendulum dynamics in the phase space: a grid is consistently defined only for those stepsizes that produce periodic orbits.

$\widetilde{H}(p_n, q_n)$ may be simplified as follows:

$$\widetilde{H}(p_n, q_n) = \frac{1}{2}p_n^2 + \frac{1}{2}q_n \sin(q_n) + \frac{1}{2}\sum_{i=0}^{n-1}\left(q_{i+1}\sin(q_i) - q_i \sin(q_{i+1})\right), \tag{9}$$

where $(p_0, q_0), (p_1, q_1), \ldots, (p_n, q_n)$ is any path on D joining the origin to the point y_n (for example the one shown in the picture). In the present example, $\widetilde{H}(p_n, q_n)$ assumes the constant value 0.245, just like the continuous Hamiltonian function $H(p(t), q(t))$ because we have no error in the p direction, since the variable p contributes a quadratic term in the expression of $H(p, q)$. Figure 6 (left picture) also reports the function $\widetilde{H}(p_n, q_n)$: it acts like a Liapunov function in that it admits a level curve coinciding with the orbit of the discrete system.

The class of problems we analyze are those having the Hamiltonian function in separable form $H(p, q) = T(p) + U(q)$ with $H(p, q)$ regular and strictly convex in a domain surrounding the origin. In particular we assume that $f_1(p) = T'(p)$ and $f_2(q) = U'(q)$ satisfy the following relations in the rectangle $R = \{(p, q) \in (-r_p, r_p) \times (-r_q, r_q)\}$:

$$f_1(p) > 0 \text{ for } p > 0,\; f_1(-p) = -f_1(p), \quad f_2(q) > 0 \text{ for } q > 0,\; f_2(-q) = -f_2(q), \tag{10}$$

that provide closed, marginally stable orbits symmetric with respect to the origin. Problems from mechanics with a central force, including the nonlinear pendulum equation, fall in this class ($T(p)$ and $U(q)$ are the kinetic and potential energies respectively). Our aim is to investigate the long time behaviour in the phase plane of the orbits provided by the trapezoidal method and in particular to check whether their stability properties can be considered as being of the same kind as those of the continuous problem. As already observed, in the discrete case periodicity of the solutions does not arise naturally, but needs additional requirements. We discuss the periodic and non-periodic cases separately.

5 Periodic orbits

The assumption that the orbit should lie on a grid is exceptional but not impossible as the previous example shows. As we will see, this may happen under very general assumptions and, more

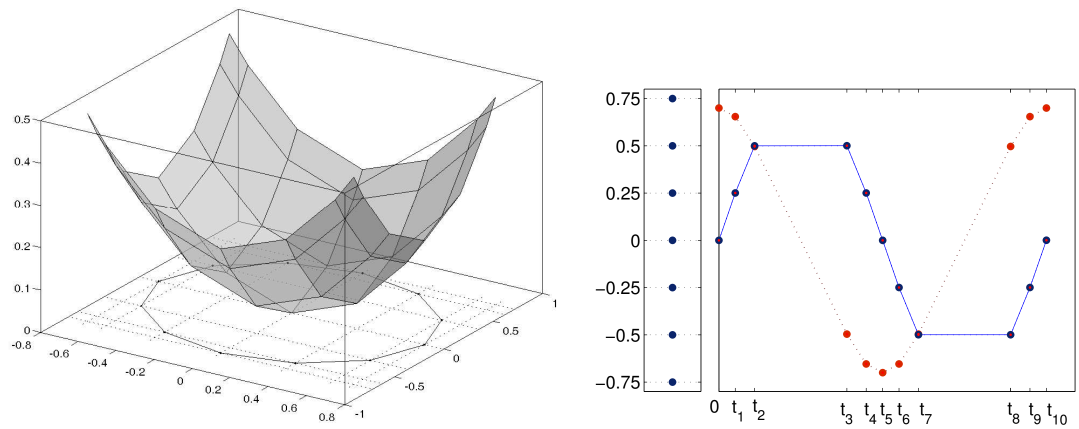

Figure 6: Left picture: a plot of $\widetilde{H}(p,q)$ on the grid D. Right picture: time evolution of the discrete pendulum (14). The q_i's are the dots joined by the solid line, while the p_i's are the dots joined by the dashed line.

importantly, without changing the standard techniques that select the stepsize to control the error. We begin with describing a simple technique to obtain periodic solutions lying on a grid that contains the equilibrium point O. For simplicity of exposition (but without loss of generality) we assume, like in the previous example, $y_0 = (p_0, 0)$ with $p_0 > 0$. Suppose we advance the solution using a stepsize that may be constant or chosen at every step by any technique for the error control. In order for the grid to contain the origin we impose the condition that the orbit passes through a point located on the q axis. As soon as the current state vector $y_n = (p_n, q_n)$ is such that $p_n < 0$, we reject the last computed stepsize and tune it in order to have $p_n = 0$ (we assume continuity of $y_n = y_n(h)$). If $q_0 \neq 0$ we could redo the same procedure using positive and negative times, until we reconstruct the orbit in the sector that contains y_0. At this point, the symmetry of the vector field and of the method imply that the orbit in the remaining sectors of the phase space is obtained by a simple reflection of the piece of orbit already computed. For example, in the fourth sector, by reversing the sequence of stepsizes used to construct the orbit in the first sector (from $y_0 = (p_0, 0)^T$ to $y_n = (0, q_n)^T$) one would get $y_{-m} = (p_{-m}, q_{-m}) = (p_m, -q_m)^T$ in conformity with the analogous property $y(-t) = (p(-t), q(-t))^T = (p(t), -q(t))^T$ of the continuous system (with $y(0) = y_0$).

The success of this procedure, is guaranteed using any stepsize, provided f satisfies some mild assumptions. Exploiting the symmetry argument, it is enough to restrict our study to the first sector $S_1 = \{(p,q) : p \geq 0,\ q \geq 0\}$. We denote by $y_n = (p_n, q_n)$ the solution at time t_n provided by the trapezoidal method applied to the Hamiltonian system $y' = J\nabla H(y)$ satisfying (10).

Lemma 5.1 *If y_n and y_{n+1} belong to S_1, then $p_{n+1} < p_n$ and $q_{n+1} > q_n$.*

Proof. The assertion comes directly from the expression of the trapezoidal method, by exploiting the positivity of f_1 and f_2 in the first sector. □

This result states the monotonicity of the orbit in each sector which, in turn, assures us that the technique described above to generate the grid works well provided the solution exists at each step and includes a point on the p axis and a point on the q axis: the following results are concerned with these two questions.

Theorem 5.2 *Let $y_0 = [p_0, q_0]^T$, $y = [p, q]^T$, with $y_0, y \in S_1$, and f be a vector field satisfying (10). The following propositions are equivalent:*

(a) $y = y_0 + \frac{h}{2} J(f(y_0) + f(y))$, *for some stepsize* $h \in \mathbb{R}$;

(b) $(y - y_0)^T (f(y_0) + f(y)) = 0$.

Furthermore, the trapezoidal method will admit a continuous solution curve $y(h) = [p(h), q(h)]^T \in S_1$, *with* $h \in [-h_1, h_2]$, *satisfying* $y(0) = y_n$, $p(h_2) = 0$ *and* $q(-h_1) = 0$ *if*

$$\max_{0 \le p \le p_0} \frac{1}{2}(p_0 - p)(f_1(p_0) + f_1(p)) < \sup_{q_0 \le q < r_q} \frac{1}{2}(q - q_0)(f_2(q_0) + f_2(q)) \tag{11}$$

and

$$\max_{0 \le q \le q_0} \frac{1}{2}(q_0 - q)(f_2(q_0) + f_2(q)) < \sup_{q_0 \le q < r_q} \frac{1}{2}(p - p_0)(f_1(p_0) + f_1(p)) \tag{12}$$

Proof. Both (a) and (b) express orthogonality between $y - y_0$ and $f(y_0) + f(y)$. The equivalence comes from the fact that the latter term never vanishes on $S_1 \setminus O$. Therefore all the solutions of the trapezoidal method that lie on S_1 may be retrieved by equation (b) that, in terms of the unknowns p and q, reads

$$(p - p_0)((f_1(p_0) + f_1(p)) + (q - q_0)(f_2(q_0) + f_2(q)) = 0. \tag{13}$$

Fix any $\bar{p} \in [0, p_0]$. From Lemma 5.1 a solution $(\bar{p}, \bar{q}) \in S_1$ of (13) must satisfy $\bar{q} \ge q_0$ and therefore it will exist if condition (11) is satisfied. Similarly, (12) provides a sufficient condition for a solution $(\bar{p}, \bar{q}) \in S_1$ of (13) to exist when $\bar{q} \in [0, q_0]$. ◻

Condition (b) or equivalently equation (13) describe the locus of points in the phase space that make the discrete line integral on the segment (y_0, y) vanish. The minimum period that one can ask for is four, with y_0 and y_1 located on the positive p and q axes[1]. All periodic solutions that lie on a grid containing O must have period $4(n+1)$ since, by symmetry, all the points that are strictly inside S_1 also recur in the other sectors. However, apart from the solution of period 4, one can obtain infinitely many solutions of a given period (when the period is $4(n+1)$, these are ∞^n).

In principle, conditions (11) and (12) should be checked for any point of the piece of orbit in S_1 but, in most cases of interest, the existence of period 4 implies that of any greater period. For example, one easily realizes that (11) and (12) are satisfied whatever the choice of (p_0, q_0), if $\lim_{p \to r_p^-} p f_1(p) = \lim_{q \to r_q^-} q f_2(q) = +\infty$. The following result considers a less restrictive sufficient condition, that gives account for the dynamics in a neighborhood of the equilibrium point O.

Corollary 5.3 *Suppose there exist* $\bar{p} > 0$ *and* $\bar{q} > 0$ *such that* $\bar{p} f_1(\bar{p}) = \bar{q} f_2(\bar{q}) (\equiv M)$ *(condition for the existence of a solution with period 4). Assume* $f_1'(p) \ge 0$, $f_1''(p) \le 0$ *for* $p \in [0, \bar{p}]$ *and* $f_2'(q) \ge 0$, $f_2''(q) \le 0$ *for* $q \in [0, \bar{q}]$. *Then, for any point* $(p_0, q_0) \in S_1$ *such that* $p_0 f_1(p_0) + q_0 f_2(q_0) \le M$, *conditions (11) and (12) are satisfied.*

Proof. The functions $g_1(p) = (p - p_0)(f_1(p_0) + f_1(p))$ and $g_2(q) = (q - q_0)(f_2(q_0) + f_2(q))$ are strictly increasing; therefore:

$$\max_{0 \le p \le p_0} (p_0 - p)(f_1(p_0) + f_1(p)) = p_0 f_1(p_0)$$

and

$$\max_{q_0 \le q \le \bar{q}} (q - q_0)(f_2(q_0) + f_2(q)) = (\bar{q} - q_0)(f_2(q_0) + f_2(\bar{q})).$$

Finally,

$$p_0 f_1(p_0) \le M - q_0 f_2(q_0) = \bar{q} f_2(\bar{q}) - q_0 f_2(q_0) \le (\bar{q} - q_0)(f_2(q_0) + f_2(\bar{q})).$$

[1] here we require that the periodic orbit has points on the p and q axes; without this constraint, one can obtain period 2 for $h \to \infty$.

The last inequality derives from the fact that $f_2(q)$ is positive and concave; in fact $A_1 = (\bar{q} - q_0)(f_2(q_0) + f_2(\bar{q}))$, $A_2 = \bar{q} f_2(\bar{q})$ and $A_3 = q_0 f_2(q_0)$ are obtained by the trapezoidal formula applied in the intervals $[q_0, \bar{q}]$, $[0, \bar{q}]$ and $[0, q_0]$ respectively, and therefore they satisfy $A_2 \leq A_1 + A_3$. This proves condition (11). The condition for (12) can be derived in an analogous way. □

The nonlinear pendulum equation satisfies the above assumptions with $\bar{q} = \pi/2$ and $\bar{p} = \sqrt{\pi/2}$. This means, for example, that starting at $(p_0, 0)$, $0 < p_0 \leq \bar{p}$, solutions of period $4(n+1)$ may be retrieved for any $n \geq 0$. We emphasize that the stepsize h does not need to remain constant during the integration: the only constraints required are that the p and q axis contain a point of the orbit. On the other hand, working with fixed stepsize and exploiting continuity of the solutions with respect to the parameter h, it is easy to state the existence of infinitely many periodic orbits defining a grid. More precisely, starting from the stepsize h_0 that provides period four, and decreasing the value of h, one can pick a sequence of stepsizes h_n approaching zero and providing periods $4(n+1)$.

5.1 An application: dynamics on lattices

The fact that only a discrete set of stepsizes is admissible in order to get periodic solutions should not be viewed as a surprise. This is due to the discrete nature of our solutions. From a physical point of view we naturally fall into the case described in the previous sections as soon as we let the space variable q assume values in a discrete set. The analysis of dynamical systems defined on lattices is now extensively researched both in theoretical and applied physics (see for example [10] for recent studies on this topic). For example, we may think of the periodic solution in Example 4.1 as the solution of the nonlinear discrete pendulum system

$$\begin{cases} p_{n+1} = p_n - \dfrac{h_n}{2}(\sin(q_n) + \sin(q_{n+1})), \\ q_{n+1} = q_n + \dfrac{h_n}{2}(p_n + p_{n+1}), \end{cases} \tag{14}$$

where the values of the q_n's locate the positions of adjacent atoms forming a linear lattice. More precisely, the dynamics of (14) has to be interpreted as follows. Fix a discrete set of admissible values for the space variable q centred at 0: $q \in \{0, \pm Q_1, \pm Q_2, \dots\}$. Let the space variable of the solution (p_n, q_n) at time t_n lie at the point Q_s. Then choose the stepsize h_n as the minimum positive value such that, at time $t_{n+1} = t_n + h_n$, $q_{n+1} \in \{Q_{s-1}, Q_s, Q_{s+1}\}$. Which among $Q_j,\ j = s-1, s, s+1$ is selected, will depend on the sign of the velocity $(p_n + p_{n+1})/2$. The right plot in Figure 6 shows one such solution (we choose $Q_s = \pm\frac{1}{4}s,\ s \in \mathbb{N}$ and $(p_0, q_0) = (0.7, 0)$). It follows that the transition time from one state to the next one will not remain constant but will be tuned in order that the transition itself may be realized. For this example we get a solution of period 10 and $[h_1, \dots h_{10}] \simeq [0.37, 0.43, 2.07, 0.43, 0.37, 0.37, 0.43, 2.07, 0.43, 0.37]$.

Looking at the figure, we see that $q_2 = q_3$ and $q_7 = q_8$. This means that in the time intervals $[t_2, t_3]$ and $[t_7, t_8]$ there are no transitions at all but just a reversal of the velocities: $p_3 = -p_2$ and $p_8 = -p_7$. If we assume the additional hypothesis that a transition must necessarily result in a change of the value of q at each interval time $[t_i, t_{i+1}]$, we realize that not all the initial velocities p_0 are allowed, but only those (forming a discrete set) yielding a solution for which the maximum value of $|q_i|$ is reached at a null velocity. From (9) it then follows that only a discrete set of energies are allowed, which typically is the case in quantum physics.

6 Non-periodic orbits

As already emphasized in the introduction, the study of non-periodic orbits, which represents the general case if no restriction on the choice of the stepsize h is imposed, has been thoroughly studied in the literature and also specified for 2D problems [4]. For the class (10) we wonder if, even in this case, we can attach to the discrete problem a discrete energy. A number of properties concerning the qualitative behaviour of the solution over long times may be derived once we state for the trapezoidal rule a result of measure preservation.

As is well known, a two-dimensional mapping $(p_1, q_1) = \Phi(p_0, q_0)$ defined on $\Omega \subset \mathbb{R}^2$ is area preserving iff its Jacobian matrix satisfies the relation

$$(\Phi'(p_0, q_0))^T J\, \Phi'(p_0, q_0) = J, \quad \forall (p,q) \in \Omega, \qquad \text{with } J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{15}$$

The trapezoidal method is not symplectic but, being conjugate to a symplectic method (the implicit midpoint formula), it is a Poisson map and therefore it preserves a different (non-Euclidean) measure. Instead of (15), a 2D Poisson map $(p_1, q_1) = \Phi(p_0, q_0)$ satisfies

$$\mu(p_1, q_1)\, (\Phi'(p_0, q_0))^T J\, \Phi'(p_0, q_0) = \mu(p_0, q_0)\, J$$

for some scalar function $\mu(p,q)$ (referred to as integrating factor), and hence preserves the measure

$$\mathcal{M}(S) = \iint_S \mu(p,q)\, dp\, dq, \tag{16}$$

where $S \subset \Omega$ is any measurable subset of Ω (we assume that μ is a smooth function on Ω), that is, $\mathcal{M}(\Phi(S)) = \mathcal{M}(S)$ (in [8] the definition of *state dependent symplecticity* was introduced as a generalization of Poisson maps to investigate the behavior of general symmetric one-step methods). The integrating factor for the trapezoidal method is

$$\mu(p,q) = 1 + \frac{h^2}{4} f_1'(p) f_2'(q).$$

Assuming boundedness of the orbits in a region containing the equilibrium point, an immediate consequence of (16) is the possibility of applying Poincaré recurrence theorem which, in our situation, states that almost all points (p_n, q_n) of the orbit are recurrent, namely for each neighborhood B of (p_n, q_n) there exists an integer $r > 0$ such that $\Phi^r(p_n, q_n) \in B$.

In the non-periodic case the orbit will densely cover a closed (invariant) curve $\mathcal{C}$ around the origin like in the left plot of Figure 1. One way to attack the problem is to exploit Birkhoff's ergodic theorem which, again, requires boundedness of the orbits and a measure preserving mapping Φ. Under these assumptions the theorem says that for any $f \in L^1(\mu)$, the set of μ-integrable functions, there exists $f^* \in L^1(\mu)$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\Phi^k(p,q)) = f^*(p,q). \tag{17}$$

The interest in (17) in relation with the numerical simulation of conservative systems resides in the following fact: if one can state that there is a smooth quantity $f(p,q)$ which is *exactly preserved* by the continuous flow Φ and *nearly preserved* by the discrete flow $\Phi_h = \Phi + O(h^p)$ ($p = 2$ for the trapezoidal method), then the function f_h^*, defined by Φ_h through (17), may be assumed as the discrete counterpart of f because Birkhoff's theorem also states that f_h^* is Φ_h-invariant, that is $f_h^* \circ \Phi_h = f_h^*$ (for recent results regarding the use of Birkhoff's ergodic theorem to Hamiltonian systems see [11]).

Example 6.1 *We consider the solution of the nonlinear pendulum equation reported in Figure 1 and assume $f \equiv H(p,q) = 1/2p^2 - \cos(q)$, the Hamiltonian function of the continuous system. As a result of* (17)*, the mean value of $H(y_n)$ will admit a limit value $H_h^*(p_0, q_0)$ that may be interpreted as a discrete energy.*

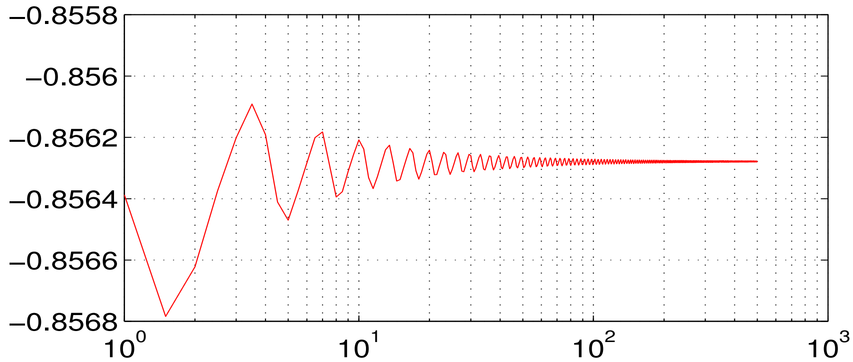

Figure 7: Mean value of $H(y_n)$ for $t_n = nh \in [0, 500]$.

We observe that any other expression which could be interpreted as an energy term would work in (17) and would lead to a discrete energy. If for example the solution is periodic then, as seen previously, it is more appropriate to consider as a discrete energy term the one, say $\widetilde{H}_h(y_n)$, introduced in the previous sections which, although on the one hand would differ from the above defined mean value of $H(y_n)$, on the other hand would have the advantage to be an invariant of the discrete system h^2-close to H (see also the next section; of course these energy terms will coincide when $h \to 0$).

One remark is in order: the use of the mean value of $H(y_n)$ as a discrete energy would indeed introduce a discontinuity in the argument presented in the above sections. More precisely, fix the initial point y_0 and assume that $h \in [\underline{h}, \bar{h}]$, where $\underline{h}$ and $\bar{h}$ are stepsizes giving rise to periodic solutions lying on a grid and such that in $(\underline{h}, \bar{h})$ there are no other periodic solutions on a grid. Then, exploiting a continuity argument, the value of the discrete energy term to attach to the solution with stepsize h is expected to be intermediate between the values $\widetilde{H}_{\underline{h}}$ and $\widetilde{H}_{\bar{h}}$ which, as seen, are the discrete line integrals on paths joining the origin to any point of the (periodic) solution.

The above point would therefore lead us to prefer an expression for the discrete energy term $\widetilde{H}_h$, which reduces to the one already considered in the previous section in the case where the solution lies on a grid. Stated differently, the discrete energy function $\widetilde{H}_h$ is already known for all the values of h leading to a solution lying on a grid (these are indeed infinite and admit zero as a limit point), and the problem is how to reconstruct it for all other values of h. It would not be erroneous to use any kind of interpolation procedure, the simplest ones being a piecewise constant polynomial or a linear spline. In the particular case where the Hamiltonian function depends quadratically on p (that is $f_1(p) = 1/2p^2$), as is the case for the nonlinear pendulum, it would be quite natural to take the value of the discrete energy equal to $1/2\bar{p}^2$, where $\bar{p}$ is the intersection of the curve $\mathcal{C}$ with the p-axis: this is because the discrete line integral along the p axis would be the same whatever the decomposition of the path used, due to the fact that (6) is exact for quadratic vector fields.

7 Order of approximation

Starting from the trapezoidal rule and the continuous (conservative) vector field $\nabla H(y)$, we have derived a discrete conservative vector field $\nabla \widetilde{H}(y_n)$ defined on an arbitrary orthogonal grid. In both cases conservativity means that the line integral vanishes on circuits. The key point here is that the symmetry of the formula has allowed us to deduce the conservative property of the discrete vector field as soon as we replace the continuous line integral by its discrete approximation. Another relevant question is to investigate the order of approximation, namely to obtain an estimation of $H(y_n) - \widetilde{H}(y_n)$, where y_n is any point on the considered grid. More precisely, since our final goal is the study of the behaviour of the numerical solution y_n generated by a symmetric scheme (assuming by now that it lies on a certain grid), the error $H(y_n)-\widetilde{H}(y_n)$ ($\widetilde{H}(y_n)$ is indeed constant) is useful to get information about the displacement of the numerical solution from the manifold that contains the dynamics of the continuous one. The result may be trivially obtained by looking at the error in the quadrature formula (we assume again $H(0) = 0$). Let $\sigma_n = (y_0, y_1, \dots, y_n)$ be a path on the grid, joining the origin ($y_0 = 0$) with the point y_n. The restriction of σ_n to the edge (y_i, y_{i+1}) may be parameterized as follows:

$$\sigma_n(s) = y_i + s(y_{i+1} - y_i), \qquad \text{with } s \in [0,1].$$

Its derivative in that segment is $\dot{\sigma}_n(s) = y_{i+1} - y_i$, and therefore

$$H(y_n) = \int_{\sigma_n} \nabla H \, ds = \sum_{i=0}^{n-1} \left[(y_{i+1} - y_i)^T \int_0^1 \nabla H \left(y_i + s(y_{i+1} - y_i) \right) \, ds \right],$$

while, by definition,

$$\widetilde{H}(y_n) = \sum_{\sigma_n} \nabla H \, \Delta y = \frac{1}{2} \sum_{i=0}^{n-1} (y_{i+1} - y_i)^T \left(\nabla H(y_i) + \nabla H(y_{i+1}) \right).$$

Subtracting these two expressions and considering that the latter one is obtained by solving the integral in the former one by means of the trapezoidal rule, yields

$$H(y_n) - \widetilde{H}(y_n) = -\frac{1}{12} \sum_{i=0}^{n-1} (y_{i+1} - y_i)^T \bar{f}_i''(\xi_i), \tag{18}$$

where $f_i(s) = \nabla H \left(y_i + s(y_{i+1} - y_i) \right), \; i = 0, \dots, n-1$, and the bar over f means that actually the value of ξ_i generally changes along the components of the vector function f_i, namely

$$\bar{f}_i''(\xi_i) = \left[\frac{d^2 f_i^{(1)}}{ds^2}(\xi_i^{(1)}), \frac{d^2 f_i^{(2)}}{ds^2}(\xi_i^{(2)}), \dots, \frac{d^2 f_i^{(2m)}}{ds^2}(\xi_i^{(2m)}) \right]^T.$$

From the definition of f_i, the jth component of $\bar{f}_i''(\xi_i)$ takes the form

$$\frac{d^2 f_i^{(j)}}{ds^2}(\xi_i^{(j)}) = (y_{i+1} - y_i)^T \cdot J_{\nabla H^{(j)}}(y_i + \xi_i^{(j)}(y_{i+1} - y_i)) \cdot (y_{i+1} - y_i), \tag{19}$$

where $J_{\nabla H^{(j)}}(y_i + \xi_i^{(j)}(y_{i+1} - y_i))$ is the Jacobian of the jth component of ∇H evaluated at a suitable point of the segment (y_i, y_{i+1}). Once again, either from (18) or (19), we see that for quadratic Hamiltonians the error vanishes in accordance with the fact that the trapezium method is exact for linear functions.

Let M be an upper bound for $||J_{\nabla H^{(j)}}(y)||$, $j = 1, \dots, 2m$, for any y belonging to a given domain D containing the origin, the points y_n we are interested in and, of course, the paths σ_n. From (19) we deduce the bound

$$||\frac{d^2 f_i^{(j)}}{ds^2}(\xi_i^{(j)})|| \le M ||y_{i+1} - y_i||^2,$$

that inserted in (18) yields

$$|H(y_n) - \widetilde{H}(y_n)| \leq \frac{M}{12} \sum_{i=0}^{n-1} ||y_{i+1} - y_i||^3. \tag{20}$$

Introducing the size of the grid on the domain D as $\bar{h} = \max_{y_i \in D} ||y_{i+1} - y_i||$, if the domain is bounded, from (20) we steadily get $H(y_n) = \widetilde{H}(y_n) + O(\bar{h}^2)$. In particular, if the grid is deduced by the numerical solution, as in Example 4.1, we can assume $\bar{h} = h$ where h is the stepsize of integration.

8 Conclusions

The use of symmetric methods to approximate the line integral of a continuous conservative vector field, allowed us to define discrete vector fields having the special feature of maintaining a conservative character which, analogously to the continuous case, was expressed in terms of discrete line integral over a (discrete) closed path. Each p-order symmetric method yields a corresponding $O(h^p)$-close approximation of the original continuous field. By using, for simplicity, the trapezoidal method, we introduced the discrete counterpart of a number of operators, like the divergence, the gradient and the curl operators, which turn out to be very useful to describe the properties of conservative vector fields. In particular, the definition of closed difference forms was introduced and exploited in Theorem 3.2 to state that a conservative discrete vector field is nothing but the discrete gradient of a suitable scalar potential function.

We exploited such result to introduce discrete Hamiltonian systems, by simply imposing orthogonality between the discrete velocity and the (discrete) gradient of the field itself (provided the solution lies on a grid). Confining the analysis to 2D dimensional fields, two scenarios may occur in the phase space, according to whether the numerical solution lies on a grid or otherwise. In the former case, which includes the dynamics on lattices, we obtain an explicit expression of the *discrete energy*, that is a quantity approximating the continuous energy up to the order of the method and which remains exactly constant along a numerical solution. In the latter case, we exploited the measure preservation property of the trapezoidal method to obtain the discrete energy function via Birkhoff's ergodic theorem.

Acknowledgment

This work was supported by COFIN-PRIN 2004 (project "Metodi numerici e software matematico per le applicazioni").

References

[1] P. Amodio, F. Iavernaro and D. Trigiante, *Symmetric schemes and Hamiltonian perturbations of Hamiltonian systems*, Numer. Linear Algebra Appl., **12** (2005) (2-3), 171-179.

[2] J. Edwards, *A treatise on the integral calculus*, Chelsea Pub. Co., New York, 1930.

[3] N. W. Evans and S. Tremaine, *Linear multistep methods for integrating reversible differential equations*, Astron. J., **118** (1999), 1888-1899.

[4] E. Faou, E. Hairer, T.-L. Pham, *Energy conservation with non-symplectic methods: examples and counter-examples*, BIT Numerical Mathematics, **44** (2004), 699-709.

[5] E. Hairer and C. Lubich, *Symmetric multistep methods over long times*, Numer. Math., **97** (2004), 699-723.

[6] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations.*, Springer Series in Computational Mathematics, **31**, Springer-Verlag, Berlin, 2002.

[7] F. Iavernaro and D. Trigiante, *On some conservation properties of the trapezoidal method applied to Hamiltonian systems*, J. Comput. Appl. Math., (in press).

[8] F. Iavernaro and D. Trigiante, *State dependent symplecticity and area preserving numerical methods*, submitted.

[9] G. D. Quinlan and S. Tremaine, *Symmetric multistep methods for the numerical integration of planetary orbits*, Astron. J., **100** (1990), 1694-1700.

[10] V. Talasila, J. Clemente-Gallardo, A. J. van der Schaft, *Geometry and Hamiltonian mechanics on discrete spaces*, J. Phys. A **37** (2004) (41), 9705–9734.

[11] P. F. Tupper, *Ergodicity and the Numerical Simulation of Hamiltonian Systems*, SIAM J. Appl. Dyn. Syst., **4**(3) (2005), 563-587.