Chapter 14

First Integrals

Main concepts: First integrals, linear, quadratic, discrete gradients.

14.1 First Integrals

In Chapter 5 we said defined the space $\mathbb{R}^d$ to be the phase space of the differential equation (1.13). Each point in $\mathbb{R}^d$ defines a state of the system, and a solution of (1.13) traces a curve in $\mathbb{R}^d$. Given a function $I(y) : \mathbb{R}^d \to \mathbb{R}$, we can evaluate this function along a solution to (1.13), i.e. sampling $I(y(t))$. For some special functions I we may find that that I is constant along solutions to (1.2). Such a function is called a *first integral* of (1.2).

For example, many models from physics are designed to incorporate energy conservation. In the case of the harmonic oscillator (5.4) the energy is just

$$E(x,p) = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 x^2.$$

You can easily check that this is constant along solutions of (5.4) by differentiation of the expression with respect to time, and use of the differential equations:

$$\begin{aligned} \frac{d}{dt}E(x(t),p(t)) &= p\dot{p} + \omega^2 x\dot{x} \\ &= p(-\omega^2 x) + \omega^2 xp = 0. \end{aligned}$$

The energy of a conservative mechanical system is one example of a first integral.

We say that for any differential equation of the form (1.13), a function $I(y)$ is a **first integral** if the function I is constant when evaluated along any solution of the differential equation. Using the equation (1.13), we can easily see that I is a first integral provided, whenever $y = y(t)$ is a solution of (1.13),

$$\frac{d}{dt}I(y(t)) \equiv 0.$$

Using the chain rule, this just means

$$\nabla I(y(t)) \cdot \frac{dy}{dt}(t) \equiv 0,$$

or (since $dy/dt = f(y)$ for any solution of (1.13)), simply

$$\nabla I \cdot f(y) = 0 \quad \text{or,} \quad (\nabla I)^T f(y) = 0.$$

Other names for first integrals include: *invariant*, *constant of motion*, and *conserved quantity*.

First integrals play several crucial roles in studies of dynamical systems. For example it is often the case that first integrals help to define the problem; their values can be viewed as constants that have specific physical significance. First integrals may confine the solution to a bounded region of phase space. In many systems, the first integrals are the only computable measures of the performance of the numerical method – we cannot plot the error in the solution unless the exact solution is available, but we can almost always compute and monitor the energy or another first integral, and use this to assess the quality of the approximation.

Example. The rigid body equations describe the motion of a rotating object in the absence of external forces. The equations are

$$\frac{dy}{dt} = y \times \Gamma y, \tag{14.1}$$

where $y(t) = (y_1(t), y_2(t), y_3(t))^T \in \mathbb{R}^3$ and $\Gamma = \text{diag}(I_1^{-1}, I_2^{-1}, I_3^{-1})$ is a diagonal matrix and $I_j > 0$, $j = 1, \ldots, 3$ are the moments of inertia about the principle axes of the object.

For any solution to the rigid body equations the functions:

$$H(y_1, y_2, y_3) = \frac{1}{2}\left(\frac{y_1^2}{I_1} + \frac{y_2^2}{I_2} + \frac{y_3^2}{I_3}\right)$$

and

$$K(y_1, y_2, y_3) = y_1^2 + y_2^2 + y_3^2$$

are constant (and equal to $H(y(0))$ and $K(y(0))$, respectively). Note that the solution is constrained to the intersection of the surfaces of a sphere and an ellipsoid, i.e. a closed one-dimensional submanifold of $\mathbb{R}^3$. The preservation of either of these first integrals guarantees the boundedness of the solution as $t \to \infty$.

In this chapter, we discuss the conservation of first integrals by numerical methods.

14.2 First Integrals Under Discretization

Suppose $d = 1$ in (1.13) then preservation of a first integral means that the scalar equation $I(y) = I_0$ is satisfied. Such a nonlinear equation, for smooth I, generally has at most just some discrete set of point solutions there can be no dynamics. Thus it is clear that in order to have interesting dynamics with a first integral we must consider systems of dimension $d \geq 2$. In this case, $\{y|I(y) = I_0\}$ defines a manifold, a point set with a continuous structure. The solution of the differential equations evolves on this manifold. We can write this, compactly, as

$$I \circ \Phi_t = I.$$

Now consider the preservation of such a first integral by a numerical method, which we assume to be represented as a map Ψ_h. Under what circumstances do we have

$$I \circ \Psi_h = I?$$

We will discuss several examples to show that it is possible to have exact conservation. First suppose a system has a linear first integral. Such an integral can be represented by

$$I(y) = r + s \cdot y,$$

where r is a scalar, and s a vector in $\mathbf{R}^d$. We assume that $\nabla I \cdot f(y) = 0$ which just means that $s \cdot f(y) = 0$.

If we are given a Runge-Kutta method described by the equations

$$y_{n+1} = y_n + h\sum_i b_i F_i, \qquad F_i = f(Y_i) = f(y_n + h\sum_j a_{ij} F_j), \qquad i = 1, \ldots, s,$$

we can compute the first integral at time t_{n+1}:

$$\begin{aligned} I(y_{n+1}) &= r + s \cdot y_{n+1} \\ &= r + s \cdot (y_n + h \sum_i b_i F_i) \\ &= r + s \cdot y_n + h \sum_i b_i s \cdot F_i \\ &= I(y_n). \end{aligned}$$

Thus, *Runge-Kutta methods preserve arbitrary linear first integrals.*

Next consider the Kepler problem

$$\begin{aligned} \frac{dx}{dt} &= u \\ \frac{dy}{dt} &= v \\ \frac{du}{dt} &= -\frac{x}{(x^2+y^2)^{3/2}} \\ \frac{dv}{dt} &= -\frac{y}{(x^2+y^2)^{3/2}} \end{aligned}$$

We know that the angular momentum

$$L = xv - yu$$

is a conserved quantity. To see this, differentiate L to obtain

$$\frac{dL}{dt} = xv' + x'v - yu' - y'u,$$

where prime is used to denote differentiation with respect to t. Using the differential equation, we have

$$\frac{dL}{dt} = -\frac{xy}{(x^2+y^2)^{3/2}} + uv + \frac{yx}{(x^2+y^2)^{3/2}} - vu = 0.$$

The Kepler problem can be written more compactly by introducing vectors $q = (x, y)$, $p = (u, v)$, then the equations of motion become

$$\begin{aligned} \frac{dq}{dt} &= p \\ \frac{dp}{dt} &= -\frac{q}{\|q\|^3} \end{aligned}$$

Define

$$L = q \times p$$

as the cross product of q and p, where these vectors are viewed as extended to vectors in $\mathbf{R}^3$. Only the third component here is nonzero, and

$$\frac{d}{dt}L = q \times p' + q' \times p = 0$$

Discretizing with, say, Euler's method, does not result in conservation of L. (Check this!) But if we use the one-sided Euler method, we have

$$\begin{aligned} q_{n+1} &= q_n + h p_{n+1} \\ p_{n+1} &= p_n - h \frac{q_n}{\|q_n\|^3} \end{aligned}$$

Now compute

$$\begin{aligned} L_{n+1} &= q_{n+1} \times p_{n+1} \\ &= (q_n + hp_{n+1}) \times p_{n+1} \\ &= q_n \times p_{n+1} + hp_{n+1} \times p_{n+1} \\ &= q_n \times p_{n+1} \\ &= q_n \times (p_n - h\|q_n\|^{-3} q_n) \\ &= q_n \times p_n - h\|q_n\|^{-3} q_n \times q_n \\ &= q_n \times p_n \\ &= L_n \end{aligned}$$

Thus the one-sided Euler method exactly conserves the angular momentum in the Kepler problem. This is quite a remarkable property in itself, but does this hold in more general cases?

14.2.1 Conservation of quadratic first integrals by Runge-Kutta methods

The angular momentum above is an example of a quadratic first integral. It is natural to ask if the one-sided Euler method preserves all such integrals. Unfortunately this cannot be the case. For example for the harmonic oscillator with frequency $\omega = 1$, the exact solutions are circles satisfying $x^2 + p^2 = 2E$, where E is the energy. It can be easily checked that the solution using the one-sided Euler method is not a circle, however. Applying the method, we have $x_{n+1} = x_n + hp_{n+1}$, $p_{n+1} = p_n - hx_n$, so

$$\begin{aligned} x_{n+1}^2 + p_{n+1}^2 &= (x_n + hp_n - h^2 x_n)^2 + (p_n - hx_n)^2 \\ &= x_n^2 + 2hx_n p_n + h^2 p_n^2 + h^4 x_n^2 - 2h^2 x_n^2 - 2h^3 x_n p_n + p_n^2 - 2hx_n p_n + h^2 x_n^2 \\ &= x_n^2 + p_n^2 + h^2(p_n^2 - x_n^2) + O(h^3). \end{aligned}$$

The second order term does not vanish identically, so the energy cannot be conserved.

Remarkably, however, the implicit midpoint rule (8.5) *does* conserve all quadratic invariants

$$I(y) = y^T C y, \quad y^T C f(y) = 0, \quad C^T = C$$

This is easily seen by writing (8.5) as

$$\frac{y_{n+1} - y_n}{\tau} = f(\frac{y_{n+1} + y_n}{2})$$

and multiplying both sides with $\frac{1}{2}(y_{n+1} + y_n)^T C$ to get

$$\frac{1}{2\tau}(I(y_{n+1}) - I(y_n)) = (\frac{y_{n+1} + y_n}{2})^T C f(\frac{y_{n+1} + y_n}{2}) = 0.$$

More generally, we prove the following theorem:

Theorem 14.2.1 (Quadratic First Integrals) *Let $dy/dt = f(y)$ have a quadratic first integral. A Runge-Kutta method preserves this quadratic first integral if*

$$b_i b_j - b_i a_{ij} - b_j a_{ji} = 0 \tag{14.2}$$

We prove this by assuming that $I(y) = r + s \cdot y + y^T T y$ is such a first integral, where r is a scalar, s is a vector of dimension d and T is a symmetric $d \times d$ matrix. Note that the condition for I to be a first integral is that $\nabla I \cdot f \equiv 0$, or

$$s \cdot f + 2y^T T f \equiv 0.$$

This implies that

$$s \cdot f \equiv -2y^T T f. \tag{14.3}$$

Applying an RK method to the system we have

$$y_{n+1} = y_n + h\sum_i b_i F_i; \qquad F_i = f(y_n + h\sum_j a_{ij} F_j)$$

thence

$$I(y_n{+}1) = r + s \cdot (y_n + h\sum_i b_i F_i) + (y_n + h\sum_i b_i F_i)^T T (y_n + h\sum_i b_i F_i)$$

Expanding this out, we have

$$I(y_{n+1}) = I(y_n) + h\sum_i b_i s \cdot F_i + 2h\sum_i b_i y_n^T T F_i + h^2(\sum_i b_i F_i) T (\sum_i b_i F_i)$$

We need to show that, assuming (14.2),

$$J = +\sum_i b_i s \cdot F_i + 2\sum_i b_i y_n^T T F_i + h(\sum_i b_i F_i) T (\sum_i b_i F_i) = 0$$

Now, by virtue of (14.3),

$$s \cdot F_i = s \cdot f(Y_i) = -2Y_i^T T f(Y_i) = -2Y_i^T T F_i$$

hence

$$\begin{aligned} J &= \sum_i b_i(-2) Y_i^T T F_i + 2\sum_i b_i y_n^T T F_i + h(\sum_i b_i F_i) T (\sum_i b_i F_i) \\ &= \sum_i b_i (y_n - Y_i)^T T F_i + h(\sum_i b_i F_i) T (\sum_i b_i F_i) \\ &= -h\sum_i b_i \sum_j a_{ij} F_j^T T F_i + h(\sum_i b_i F_i) T (\sum_i b_i F_i) \end{aligned}$$

Let $F_i^T F_j = B_{ij}$. In general we assume that these terms, which depend on the vector field, are uncorrelated. In that case J will vanish identically provided each of the coefficients multiplying the B_{ij} vanishes. This coefficient is

$$b_i a_{ij} + b_j a_{ji} - b_i b_j$$

which concludes the proof.

One class of method which satisfies the conditions (14.2) is the class of Gauss-Legendre methods, which we have already learned are the optimal order methods for a given number of stages (they are also fully implicit, and therefore, often, too expensive to use in many practical examples).

In terms of finding methods that conserve *all* first integrals of a given class, this is as good as it gets: we do not have broad classes of methods which preserve cubics, quartics, etc. Fortunately, quadratics are an important case.

If we are concerned with preserving a first integral, an alternative approach is to design a method that has this as its specific goal.

14.3 Discrete Gradient Methods

The conservation of quadratic invariants by Gauss Runge-Kutta methods is automatic. Any and all such invariants will be conserved whether or not one is aware of them. In this section we describe a method for conserving a known invariant $I(y)$.

Define the exterior product of two vectors f and v:

$$f \wedge v = fv^T - vf^T, \qquad (f \wedge v)_{ij} = f_i v_j - v_i f_j.$$

If $I(y)$ is a first integral of (1.3), then, except at points y^* where $\nabla I(y^*) = 0$, f can be written as

$$f(y) = B(y)\,\nabla I(y) = \frac{1}{\|\nabla I(y)\|^2}(f \wedge \nabla I)\,\nabla I(y),$$

where $B(y)$ is skew-symmetric: $B^T = -B$.

Let $\bar{B}(y, z)$ satisfy:

- skew-symmetric structure: $\bar{B}(y, z)^T = -\bar{B}(y, z)$,
- symmetry with respect to its arguments: $\bar{B}(y, z) = \bar{B}(z, y)$,
- consistency $\bar{B}(y, y) = B(y)$.

Similarly, define $v(y) := \nabla I(y)$ and let $\bar{v}(y, z)$ be a *discrete gradient* of I, i.e. satisfying

$$\bar{v}(y, z)^T(y - z) = I(y) - I(z),$$

and consistent $\bar{v}(y, y) = v(y)$.

Then the method

$$y_{n+1} = y_n + \tau \bar{B}(y_{n+1}, y_n)\,\bar{v}(y_{n+1}, y_n)$$

satisfies

$$I(y_{n+1}) = I(y_n),$$

since

$$\begin{aligned} I(y_{n+1}) - I(y_n) &= \bar{v}(y_{n+1}, y_n)^T(y_{n+1} - y_n) \\ &= \tau \bar{v}(y_{n+1}, y_n)^T \bar{B}(y_{n+1}, y_n)\bar{v}(y_{n+1}, y_n) = 0. \end{aligned}$$

One discrete gradient is the *midpoint discrete gradient*. Denote $y_{n+1/2} := (y_{n+1} + y_n)/2$ and $\Delta y := y_{n+1} - y_n$. Then

$$\bar{v}(y_{n+1}, y_n) := v(y_{n+1/2}) + \frac{I(y_{n+1}) + I(y_n) - v(y_{n+1/2})^T \Delta y}{\|\Delta y\|^2}\,\Delta y.$$

A possible realization of the matrix $\bar{B}$ is given by

$$\bar{B}(y, z) = B(\frac{y + z}{2}).$$

Example. Mechanical systems (12.3) can be written (using a change in variables to eliminate the mass matrix M) in the form

$$\dot{q} = p, \qquad \dot{p} = \nabla V(q),$$

where $F(q) = -\nabla V(q)$. For $y = (q^T, p^T)^T$, this defines a gradient system with skew-symmetric, constant matrix

$$B = \begin{bmatrix} 0 & id \\ -id & 0 \end{bmatrix},$$

and first integral

$$I(q, p) = \frac{p \cdot p}{2} + V(q).$$

Define the discrete gradient $\bar{v}(\bar{q}, p)$ by

$$\bar{v}_q(q_{n+1}, p_{n+1}, q_n, p_n) = \frac{I(q_{n+1}, p_n) - I(q_n, p_n)}{q_{n+1} - q_n},$$

and

$$\bar{v}_p(q_{n+1}, p_{n+1}, q_n, p_n) = \frac{I(q_{n+1}, p_{n+1}) - I(q_{n+1}, p_n)}{p_{n+1} - p_n},$$

where in both cases the division is understood component-wise.

14.4 Splitting Methods

Consider a skew-gradient system

$$y' = B(y)\nabla I(y),$$

where $B(y)^T = -B(y)$, and suppose we can find a splitting of the form

$$y' = (B_1(y) + \cdots + B_k(y))\nabla I(y),$$

where $B_i(y)^T = -B_i(y)$, and each subproblem

$$y' = B_i(y)\nabla I(y), \quad i = 1, \ldots, k$$

is exactly integrable.

Then the splitting method (13.1) with

$$f_i = B_i(y)\nabla I(y), \quad i = 1, \ldots, k$$

exactly conserves I.

Example. Consider the rigid body equations (14.1). Let us derive a splitting method that exactly conserves the energy

$$H(y_1, y_2, y_3) = \frac{1}{2}\left(\frac{y_1^2}{I_1} + \frac{y_2^2}{I_2} + \frac{y_3^2}{I_3}\right)$$

The rigid body equations can be written in the form

$$y' = B(y)\nabla H(y),$$

with skew-symmetric $B(y)$ split as

$$B(y) = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -y_1 \\ 0 & y_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & y_2 \\ 0 & 0 & 0 \\ -y_2 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -y_3 & 0 \\ y_3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = B_1(y) + B_2(y) + B_3(y).$$

Each of the split sub-systems

$$y' = B_i(y)\nabla H(y), \quad i = 1, \ldots, 3$$

is exactly solvable as a harmonic oscillator, and preserves the manifold $H = const$.

14.5 Exercises

1. What is the energy function for the pendulum? Use the fact that energy is conserved (and Matlab or a graphing calculator) to sketch the phase portrait for the pendulum.
2. Prove that the Lotka-Volterra system has a first integral:
$$K(Y, R) = C \log Y - DY + A \log R - BR.$$
3. Assume an autonomous differential equation system is given (**??**)and is known to have a first integral. Show that any time-rescaling of this differential equation has the same first integral. To rescale time, we introduce a time-transformation, such as
$$\frac{dt}{d\tau} = g(y),$$
where g is a scalar-valued function.
4. Show that the implicit midpoint method and the 2-stage Gauss-Legendre scheme satisfy the condition (14.2).
5. Can you think of a splitting that conserves the other first integral of the rigid body, i.e. the total angular momentum $I = \frac{1}{2}y^T y$? Can you recall a method that conserves both H and I?