# Chapter 1

# Basic Facts about HBVMs

We consider Hamiltonian problems in the form

$$\dot{y}(t) = J\nabla H(y(t)), \qquad y(t_0) = y_0 \in \mathbb{R}^{2m}, \tag{1.1}$$

where $J$ is a skew-symmetric constant matrix, and the Hamiltonian $H(y)$ is assumed to be sufficiently differentiable. Usually,

$$J = \begin{pmatrix} & I_m \\ -I_m & \end{pmatrix}, \qquad y = \begin{pmatrix} q \\ p \end{pmatrix}, \quad q, p \in \mathbb{R}^m,$$

so that (1.1) assumes the form

$$\dot{q} = \nabla_p H(q,p), \qquad \dot{p} = -\nabla_q H(q,p).$$

The induced dynamical system is characterized by the presence of invariants of motion, among which the Hamiltonian itself:

$$\dot{H}(y(t)) = \nabla H(y(t))^T \dot{y}(t) = \nabla H(y(t))^T J \nabla H(y(t)) = 0,$$

due to the fact that $J$ is skew-symmetric. Such property is usually lost, when numerically solving problem (1.1). This drawback can be overcome by using Hamiltonian BVMs (hereafter, HBVMs).

The key formula which HBVMs rely on, is the *line integral* and the related property of conservative vector fields:

$$H(y_1) - H(y_0) = h\int_0^1 \dot{\sigma}(t_0+\tau h)^T \nabla H(\sigma(t_0+\tau h))\mathrm{d}\tau, \tag{1.2}$$

for any $y_1 \in \mathbb{R}^{2m}$, where $\sigma$ is any smooth function such that

$$\sigma(t_0) = y_0, \qquad \sigma(t_0+h) = y_1. \tag{1.3}$$

Here we consider the case where $\sigma(t)$ is a polynomial of degree $s$, yielding an approximation to the true solution $y(t)$ in the time interval $[t_0, t_0+h]$. The numerical approximation for the subsequent time-step, $y_1$, is then defined by (1.3). After introducing a set of $s$ distinct abscissae

$$0 < c_1, \dots, c_s \le 1, \tag{1.4}$$

we set

$$Y_i = \sigma(t_0 + c_i h), \qquad i = 1, \dots, s, \tag{1.5}$$

so that $\sigma(t)$ may be thought of as an interpolation polynomial, interpolating the *fundamental stages* $Y_i$, $i = 1, \dots, s$. We observe that, due to (1.3), $\sigma(t)$ also interpolates the initial condition $y_0$.

**Remark 1.** *Sometimes, the interpolation at $t_0$ is explicitly required. In such a case, the extra abscissa $c_0 = 0$ is formally added to (1.4). This is the case, for example, of a Lobatto distribution of the abscissae [6].*

Let us consider the following expansions of $\dot\sigma(t)$ and $\sigma(t)$ for $t \in [t_0, t_0+h]$:

$$\dot\sigma(t_0 + \tau h) = \sum_{j=1}^{s} \gamma_j P_j(\tau), \qquad \sigma(t_0 + \tau h) = y_0 + h \sum_{j=1}^{s} \gamma_j \int_0^{\tau} P_j(x)\,\mathrm{d}x, \tag{1.6}$$

where $\{P_j(t)\}$ is a suitable basis of the vector space of polynomials of degree at most $s-1$ and the (vector) coefficients $\{\gamma_j\}$ are to be determined. Because of the arguments in [6, 7, 8], we shall consider an **orthonormal basis** of polynomials on the interval $[0,1]$, i.e.:

$$\int_0^1 P_i(t) P_j(t) \mathrm{d}t = \delta_{ij}, \qquad i, j = 1, \dots, s, \tag{1.7}$$

where $\delta_{ij}$ is the Kronecker symbol, and $P_i(t)$ has degree $i-1$. Such a basis can be readily obtained as

$$P_i(t) = \sqrt{2i-1}\,\hat P_{i-1}(t), \qquad i = 1, \dots, s, \tag{1.8}$$

with $\hat P_{i-1}(t)$ the shifted Legendre polynomial, of degree $i-1$, on the interval $[0,1]$.

**Remark 2.** *From the properties of shifted Legendre polynomials (see, e.g., [1] or the Appendix in [6]), one readily obtains that the polynomials $\{P_j(t)\}$ satisfy the three-terms recurrence:*

$$\begin{aligned} P_1(t) &\equiv 1, \qquad P_2(t) = \sqrt{3}(2t-1), \\ P_{j+2}(t) &= (2t-1)\frac{2j+1}{j+1}\sqrt{\frac{2j+3}{2j+1}} P_{j+1}(t) - \frac{j}{j+1}\sqrt{\frac{2j+3}{2j-1}} P_j(t), \quad j \ge 1. \end{aligned}$$

We shall also assume that $H(y)$ is a polynomial, which implies that the integrand in (1.2) is also a polynomial so that the line integral can be exactly computed by means of a suitable quadrature formula. In general, however, due to the high degree of the integrand function, such quadrature formula cannot be solely based upon the available abscissae $\{c_i\}$: one needs to introduce an additional set of abscissae $\{\hat{c}_1, \ldots, \hat{c}_r\}$, distinct from the nodes $\{c_i\}$, in order to make the quadrature formula exact:

$$\int_0^1 \dot{\sigma}(t_0 + \tau h)^T \nabla H(\sigma(t_0 + \tau h))\mathrm{d}\tau = \qquad (1.9)$$
$$\sum_{i=1}^{s} \beta_i \dot{\sigma}(t_0 + c_i h)^T \nabla H(\sigma(t_0 + c_i h)) + \sum_{i=1}^{r} \hat{\beta}_i \dot{\sigma}(t_0 + \hat{c}_i h)^T \nabla H(\sigma(t_0 + \hat{c}_i h)),$$

where $\beta_i$, $i = 1, \ldots, s$, and $\hat{\beta}_i$, $i = 1, \ldots, r$, denote the weights of the quadrature formula corresponding to the abscissae $\{c_i\}$ and $\{\hat{c}_i\}$, respectively, i.e.,

$$\begin{aligned} \beta_i &= \int_0^1 \left( \prod_{j=1, j\neq i}^{s} \frac{t - c_j}{c_i - c_j} \right) \left( \prod_{j=1}^{r} \frac{t - \hat{c}_j}{c_i - \hat{c}_j} \right) \mathrm{d}t, \qquad i = 1, \ldots, s, \\ \hat{\beta}_i &= \int_0^1 \left( \prod_{j=1}^{s} \frac{t - c_j}{\hat{c}_i - c_j} \right) \left( \prod_{j=1, j\neq i}^{r} \frac{t - \hat{c}_j}{\hat{c}_i - \hat{c}_j} \right) \mathrm{d}t, \qquad i = 1, \ldots, r. \end{aligned} \qquad (1.10)$$

**Remark 3.** *In the case considered in the previous Remark 1, i.e. when* $c_0 = 0$ *is formally considered together with the abscissae (1.4), the first product in each formula in (1.10) ranges from* $j = 0$ *to* $s$*. Moreover, also the range of* $\{\beta_i\}$ *becomes* $i = 0, 1, \ldots, s$*. However, for sake of simplicity, we shall not consider this case further.*

According to [27], the right-hand side of (1.9) is called *discrete line integral*, while the vectors

$$\hat{Y}_i = \sigma(t_0 + \hat{c}_i h), \qquad i = 1, \ldots, r, \qquad (1.11)$$

are called *silent stages*: they just serve to increase, as much as one likes, the degree of precision of the quadrature formula, but they are not to be regarded as unknowns since, from (1.6), they can be expressed in terms of linear combinations of the *fundamental stages* (1.5).

**Definition 1.** *The method defined by substituting the quantities in* (1.6) *into the right-hand side of* (1.9)*, and by choosing the unknown coefficients* $\{\gamma_j\}$ *in order that the resulting expression vanishes, is called* Hamiltonian Boundary Value Method with $k$ steps and degree $s$*, in short* HBVM($k$,$s$)*, where* $k = s + r$ *[6].*

In such a way, one easily obtains, from (1.2)–(1.3),

$$H(\sigma(t_0+h)) = H(y_0),$$

that is, the value of the Hamiltonian is *exactly* preserved at the subsequent approximation, provided by $\sigma(t_0+h)$.

In the sequel, we shall see that HBVMs may be expressed through different, though equivalent, formulations: some of them can be directly implemented in a computer program, the others being of more theoretical interest.

Because of the equality (1.9), we can apply the procedure directly to the original line integral appearing in the left-hand side. With this premise, by considering the first expansion in (1.6), the conservation property reads

$$\sum_{j=1}^{s} \gamma_j^T \int_0^1 P_j(\tau)\nabla H(\sigma(t_0+\tau h))\mathrm{d}\tau = 0, \tag{1.12}$$

which, as is easily checked, is certainly satisfied if we impose the following set of orthogonality conditions

$$\gamma_j = \int_0^1 P_j(\tau) J\nabla H(\sigma(t_0+\tau h))\mathrm{d}\tau, \qquad j=1,\dots,s. \tag{1.13}$$

Then, from the second relation of (1.6) we obtain, by introducing the operator

$$\begin{aligned} &L(f;h)\sigma(t_0+ch) = \qquad\qquad (1.14)\\ &\sigma(t_0) + h\sum_{j=1}^{s}\int_0^c P_j(x)\mathrm{d}x \int_0^1 P_j(\tau) f(\sigma(t_0+\tau h))\mathrm{d}\tau, \qquad c\in[0,1], \end{aligned}$$

that $\sigma$ is the eigenfunction of $L(J\nabla H;h)$ relative to the eigenvalue $\lambda = 1$:

$$\sigma = L(J\nabla H;h)\sigma. \tag{1.15}$$

**Definition 2.** *Equation (1.15) is the* Master Functional Equation *defining $\sigma$ [7].*

**Remark 4.** *From the previous arguments, one readily obtains that the Master Functional Equation (1.15) characterizes HBVM$(k,s)$ methods, for all $k \geq 1$. Indeed, such methods are uniquely defined by the polynomial $\sigma$, of degree $s$, the number of steps $k$ being only required to obtain an exact quadrature formula (see (1.9)).*

To practically compute $\sigma$, we set (see (1.5) and (1.6))

$$Y_i = \sigma(t_0+c_ih) = y_0 + h\sum_{j=1}^{s} a_{ij}\gamma_j, \qquad i=1,\dots,s, \tag{1.16}$$

where

$$a_{ij} = \int_0^{c_i} P_j(x)\mathrm{d}x, \qquad i,j = 1,\dots,s. \tag{1.17}$$

Inserting (1.13) into (1.16) yields the final formulae which define the HBVMs class based upon the orthonormal basis $\{P_j\}$:

$$Y_i = y_0 + h\sum_{j=1}^{s} a_{ij}\int_0^1 P_j(\tau)J\nabla H(\sigma(t_0+\tau h))\mathrm{d}\tau, \qquad i=1,\dots,s. \tag{1.18}$$

For sake of completeness, we report the nonlinear system associated with the HBVM$(k,s)$ method, in terms of the fundamental stages $\{Y_i\}$ and the silent stages $\{\hat{Y}_i\}$ (see (1.11)), by using the notation

$$f(y) = J\nabla H(y). \tag{1.19}$$

In this context, it represents the discrete counterpart of (1.18), and may be directly retrieved by evaluating, for example, the integrals in (1.18) by means of the (exact) quadrature formula introduced in (1.9):

$$Y_i = \tag{1.20}$$
$$y_0 + h\sum_{j=1}^{s} a_{ij}\left(\sum_{l=1}^{s}\beta_l P_j(c_l)f(Y_l) + \sum_{l=1}^{r}\hat{\beta}_l P_j(\hat{c}_l)f(\widehat{Y}_l)\right), \quad i=1,\dots,s.$$

From the above discussion it is clear that, in the non-polynomial case, supposing to choose the abscissae $\{\hat{c}_i\}$ so that the sums in (1.20) converge to an integral as $r = k-s \to \infty$, the resulting formula is (1.18). This implies that HBVMs may be as well applied in the non-polynomial case since, in finite precision arithmetic, HBVMs are indistinguishable from their limit formulae (1.18), when a sufficient number of silent stages is introduced. The aspect of having a *practical* exact integral, for $k$ large enough, was already stressed in [3, 6, 7, 23, 27].

We emphasize that, in the non-polynomial case, (1.18) becomes an operative method, only after that a suitable strategy to approximate the integral is taken into account. In the present case, if one discretizes the *Master Functional Equation* (1.14)–(1.15), HBVM$(k,s)$ are then obtained, essentially by extending the discrete problem (1.20) also to the silent stages (1.11). In order to simplify the exposition, we shall use (1.19) and introduce the following notation:

$$\begin{aligned} &\{\tau_i\} = \{c_i\}\cup\{\hat{c}_i\}, \qquad \{\omega_i\} = \{\beta_i\}\cup\{\hat{\beta}_i\}, \\ &y_i = \sigma(t_0+\tau_i h), \qquad f_i = f(\sigma(t_0+\tau_i h)), \qquad i=1,\dots,k. \end{aligned} \tag{1.21}$$

The discrete problem defining the HBVM$(k,s)$ then becomes,

$$y_i = y_0 + h\sum_{j=1}^{s}\int_0^{\tau_i} P_j(x)\mathrm{d}x \sum_{\ell=1}^{k}\omega_\ell P_j(\tau_\ell) f_\ell, \qquad i=1,\dots,k. \tag{1.22}$$

**Remark 5.** *We also observe that, from (1.13) and the first relation in (1.6), one obtains the equations*

$$\dot\sigma(t_0+\tau_i h) = \sum_{j=1}^{s} P_j(\tau_i)\int_0^1 P_j(\tau)J\nabla H(\sigma(t_0+\tau h))\mathrm{d}\tau, \qquad i=1,\dots,k, \tag{1.23}$$

*which may be viewed as* extended collocation conditions *according to [27, Section 2], where the integrals are (exactly) replaced by discrete sums.*

By introducing the vectors

$$\boldsymbol{y} = (y_1^T,\dots,y_k^T)^T, \qquad e=(1,\dots,1)^T \in \mathbb{R}^k,$$

and the matrices

$$\Omega = \mathrm{diag}(\omega_1,\dots,\omega_k), \qquad \mathcal{I}_s,\ \mathcal{P}_s \in \mathbb{R}^{k\times s}, \tag{1.24}$$

whose $(i,j)$th entry are given by

$$(\mathcal{I}_s)_{ij} = \int_0^{\tau_i} P_j(x)\mathrm{d}x, \qquad (\mathcal{P}_s)_{ij} = P_j(\tau_i), \tag{1.25}$$

we can cast the set of equations (1.22) in vector form as

$$\boldsymbol{y} = e\otimes y_0 + h(\mathcal{I}_s\mathcal{P}_s^T\Omega)\otimes I_{2m}\, f(\boldsymbol{y}), \tag{1.26}$$

with an obvious meaning of $f(\boldsymbol{y})$. Consequently, the method can be seen as a Runge-Kutta method with the following Butcher tableau:

$$\begin{array}{c|c} \tau_1 & \\ \vdots & \mathcal{I}_s\mathcal{P}_s^T\Omega \\ \tau_k & \\ \hline & \omega_1\ \dots\ \omega_k \end{array} \tag{1.27}$$

**Remark 6.** *We observe that, because of the use of an orthonormal basis, the role of the abscissae $\{c_i\}$ and of the silent abscissae $\{\hat c_i\}$ is interchangeable, within the set $\{\tau_i\}$. This is due to the fact that all the matrices $\mathcal{I}_s$, $\mathcal{P}_s$, and $\Omega$ depend on all the abscissae $\{\tau_i\}$, and not on a subset of them and, moreover, they are invariant with respect to the choice of the fundamental abscissae $\{c_i\}$.*

The following result then holds true.

**Theorem 1.** *Provided that the quadrature defined by the weights* $\{\omega_i\}$ *has order at least* $2s$ *(i.e., it is exact for polynomials of degree at least* $2s-1$*), HBVM(k,s) has order* $p = 2s \equiv 2\deg(\sigma)$*, whatever the choice of the abscissae* $c_1, \ldots, c_s$*.*

Proof From the classical result of Butcher (see, e.g., [21, Theorem 7.4]), the thesis follows if the usual simplifying assumptions $C(s)$, $B(p)$, $p \geq 2s$, and $D(s-1)$ are satisfied for the Runge-Kutta method defined by the tableau (1.27). By looking at the method (1.26)–(1.27), one has that the first two (i.e., $C(s)$ and $B(p)$, $p \geq 2s$) are obviously fulfilled: the former by the definition of the method, the second by hypothesis. The proof is then completed, if we prove $D(s-1)$. Such condition can be cast in matrix form, by introducing the vector $\bar{e} = (1, \ldots, 1)^T \in \mathbb{R}^{s-1}$, and the matrices

$$Q = \mathrm{diag}(1, \ldots, s-1), \qquad D = \mathrm{diag}(\tau_1, \ldots, \tau_k), \qquad V = (\tau_i^{j-1}) \in \mathbb{R}^{k \times s-1},$$

(see also (1.25)) as

$$QV^T\Omega\left(\mathcal{I}_s\mathcal{P}_s^T\Omega\right) = \left(\bar{e}\, e^T - V^T D\right)\Omega,$$

i.e.,

$$\mathcal{P}_s\mathcal{I}_s^T\Omega V Q = \left(e\,\bar{e}^T - DV\right). \tag{1.28}$$

Since the quadrature is exact for polynomials of degree $2s-1$, one has

$$\begin{aligned}
\left(\mathcal{I}_s^T\Omega V Q\right)_{ij} &= \left(\sum_{\ell=1}^{k}\omega_\ell\int_0^{\tau_\ell}P_i(x)\mathrm{d}x\,(j\tau_\ell^{j-1})\right) = \left(\int_0^1\int_0^t P_i(x)\mathrm{d}x(jt^{j-1})\mathrm{d}t\right)\\
&= \left(\delta_{i1} - \int_0^1 P_i(x)x^j\mathrm{d}x\right), \qquad i = 1, \ldots, s, \quad j = 1, \ldots, s-1,
\end{aligned}$$

where the last equality is obtained by integrating by parts, with $\delta_{i1}$ the Kronecker symbol. Consequently,

$$\begin{aligned}
\left(\mathcal{P}_s\mathcal{I}_s^T\Omega V Q\right)_{ij} &= \left(1 - \sum_{\ell=1}^{s}P_\ell(\tau_i)\int_0^1 P_\ell(x)x^j\mathrm{d}x\right)\\
&= (1-\tau_i^j), \qquad i = 1, \ldots, k, \quad j = 1, \ldots, s-1,
\end{aligned}$$

that is, (1.28), where the last equality follows from the fact that

$$\sum_{\ell=1}^{s}P_\ell(\tau)\int_0^1 P_\ell(x)x^j\mathrm{d}x = \tau^j, \qquad j = 1, \ldots, s-1. \;\Box$$

Concerning the stability of the methods, the following result holds true.

**Theorem 2.** *For all $k$ such that the quadrature formula has order at least $2s \equiv 2\deg(\sigma)$, HBVM(k,s) is perfectly $A$-stable,*[1] *whatever the choice of the abscissae $c_1, \ldots, c_s$.*

Proof As it has been previously observed, a HBVM$(k, s)$ is fully characterized by the corresponding polynomial $\sigma$ which, for $k$ sufficiently large (i.e., assuming that (1.9) holds true), satisfies the *Master Functional Equation* (1.14)–(1.15), which is independent of the choice of the nodes $c_1, \ldots, c_s$ (since we consider an orthonormal basis). When, in place of $f(y) = J\nabla H(y)$ we put the test equation $f(y) = \lambda y$, we have that the collocation polynomial of the Gauss-Legendre method of order $2s$, say $\sigma_s$, satisfies the *Master Functional Equation*, since the integrands appearing in it are polynomials of degree at most $2s - 1$, so that $\sigma = \sigma_s$. The proof completes by considering that Gauss-Legendre methods are perfectly $A$-stable. $\square$

**Example 1.** *As an example, for the methods studied in [6], based on a Lobatto distribution of the nodes $\{c_0 = 0, c_1, \ldots, c_s\} \cup \{\hat{c}_1, \ldots, \hat{c}_{k-s}\}$, one has that $\deg(\sigma) = s$, so that the order of HBVM(k,s) turns out to be $2s$, with a quadrature satisfying $B(2k)$. Finally, we observe that, with such choice of the abscissae HBVM$(s, s)$ reduces to the Lobatto IIIA method of order $2s$.*

**Example 2.** *For the same reason, when one considers a Gauss distribution for the abscissae $\{c_1, \ldots, c_s\} \cup \{\hat{c}_1, \ldots, \hat{c}_{k-s}\}$, as done in [7], one also obtains a method of order $2s$ with a quadrature satisfying $B(2k)$. Similarly as in the previous example, HBVM$(s, s)$ now reduces to the Gauss-Legendre method of order $2s$.*

**Remark 7.** *A number of remarks are in order, to emphasize relevant features of HBVM$(k, s)$:*

- *From Remark 6, HBVM(k,s) are* symmetric methods *according to the* time reversal symmetry condition *defined in [16, p. 218] (see also [18]), provided that the abscissae $\{\tau_i\}$ (see (1.21)) are symmetrically distributed [6].*
- *By virtue of Theorems 1 and 2, all methods in Examples 1 and 2 are symmetric, perfectly A-stable, and of order $2s$. In particular such HBVM$(k, s)$ are exact for polynomial Hamiltonian functions of degree $\nu$, provided that*

$$k \geq \frac{\nu s}{2}. \tag{1.29}$$

[1] That is, its region of Absolute stability precisely coincides with the left-half complex plane, $\mathbb{C}^-$.

- *For all $k$ sufficiently large so that (1.9) holds, HBVM$(k,s)$ based on the $k$ Gauss-Legendre abscissae in $[0,1]$ are* equivalent *to HBVM$(k,s)$ based on $k+1$ Lobatto abscissae in $[0,1]$ (see [7]), since both methods define the same polynomial $\sigma$ of degree $s$ (i.e., they satisfy the same Master Functional Equation (1.15)–(1.14)).*