Chapter 1
Introduction

1.1 The Birman-Schwinger Principle

The Birman-Schwinger principle is a widely used and well-established tool in mathematical quantum mechanics. It was introduced through the independent works of Birman [10] and Schwinger [81], with the idea of counting or at least estimating the number of eigenvalues of Schrödinger operators on $L^2(\mathbb{R}^n)$. To be more specific, consider (only formal at this point)

$$H = -\Delta + V;$$

to avoid introducing negative parts, we will assume that $V \leq 0$. Then it is not difficult to calculate that

(a) $-e$ is a (negative) eigenvalue of H if and only if 1 is an eigenvalue of the Birman-Schwinger operator

$$B_e = \sqrt{-V}(-\Delta + e)^{-1}\sqrt{-V}; \tag{1.1}$$

   see [53, Section 4.3.1].

Furthermore,

(b) if ϕ is an eigenfunction of H for the eigenvalue $-e$, then $\psi = \sqrt{-V}\phi$ is an eigenfunction of B_e for the eigenvalue 1;

(c) if ψ is an eigenfunction of B_e for the eigenvalue 1, then $\phi = (-\Delta + e)^{-1}(\sqrt{-V}\psi)$ is an eigenfunction of H for the eigenvalue $-e$.

The operators B_e are non-negative Hilbert-Schmidt operators (if V decays sufficiently fast and $n \leq 3$), and in particular, they are compact. Their eigenvalues can be ordered: $\lambda_1(e) \geq \lambda_2(e) \geq \cdots \to 0$, and the eigenvalue curves are decreasing in e, in that $\tilde{e} \geq e$ implies that $\lambda_k(e) \leq \lambda_k(\tilde{e})$ for all k. This implies that the number of eigenvalues of

H less than or equal to $-e$ agrees with the number of eigenvalues of B_e greater than or equal to 1, counting multiplicities in both cases; cf. [53, Figure 4.1, p. 78] for an illustration. In this way, not only the number of eigenvalues of H can be bounded, but for instance also eigenvalue moments like $\sum_j |-e_j|^\gamma$, where the sum extends over all negative eigenvalues $-e_j$ of H. This fact lies at the heart of many important results in the field. Let us only mention here the Lieb-Thirring bound

$$\sum_j |-e_j| \le L_{1,3} \int_{\mathbb{R}^3} |V(x)|^{5/2} \, dx$$

in three dimensions for an absolute constant $L_{1,3} > 0$ and $V \in L^{5/2}(\mathbb{R}^3)$. It is used in those authors' proof of the stability of matter [54], which has found many generalizations [53] and which is much easier to follow than the original argument by Dyson and Lenard [16]. Good general textbooks that cover the Birman-Schwinger principle are [53, Section 4.3], [71, 82, 83] or [86, Section 7.9]. A classical reference is [46], and the papers [8, 34, 68] provide an extensive list of related literature. There is also a large number of further applications of the Birman-Schwinger principle in a variety of different contexts. For instance, complex-valued potentials are treated in [1, 20, 22, 23], Dirac operators in [12], the Bardeen-Cooper-Schrieffer model of superconductivity in [33] and the linearized 2D Euler equations in [47].

1.2 Non-Relativistic Galactic Dynamics and the Vlasov-Poisson System

In order to explain how the Birman-Schwinger principle will turn out to be useful in galactic dynamics, we are going to introduce the gravitational Vlasov-Poisson system. It is a standard PDE system to describe the time evolution of a self-gravitating system that consists of a large number of objects (like stars or galaxies), which interact via gravitational forces.

Galactic dynamics in general refers to the modeling of the time evolution of self-gravitating matter such as galaxies or, on an even larger scale, clusters of galaxies. One attempt to do so is to write down an N-body problem, with N quite large: $N \sim 10^6 - 10^{11}$ for galaxies and $N \sim 10^2 - 10^3$ for clusters of galaxies. This N-body problem consists of coupled Newtonian equations, one for each individual object (the 'objects' in a galaxy are stars, those in a cluster of galaxies are galaxies), to study the collective behavior of the system. While results may be obtainable numerically in this way, the mathematical complexity of even the three-body problem prevents one from rigorously addressing deeper questions (concerning for instance galaxy formation or stability) for such stellar systems.

Therefore, from the early days of the field, a statistical description of the evolution has been proposed by Vlasov [91] in 1938 for plasmas (in this case a related equation is satisfied) and by Jeans [41] in 1915 for gravitational systems; see [36] for an

interesting historical discussion of the origins of the equation. It is also known as the 'collisionless Boltzmann equation', which refers to the fact that collisions among the stars or galaxies are sufficiently rare to be neglected. A standard source of information on galactic dynamics is [9].

The time evolution of such a system is then governed by a distribution function $f = f(t, x, v)$ that depends on time $t \in \mathbb{R}$, position $x \in \mathbb{R}^3$ and velocity $v \in \mathbb{R}^3$. The quantity $\int_{\mathcal{X}} dx \int_{\mathcal{V}} dv \, f(t, x, v)$ should be thought of as the number of objects (henceforth called 'particles') at time t, which are located at some point $x \in \mathcal{X} \subset \mathbb{R}^3$ and which have velocities $v \in \mathcal{V} \subset \mathbb{R}^3$. Each individual particle follows a trajectory $(X(s), V(s))$ in phase space $\mathbb{R}^3 \times \mathbb{R}^3$ such that $(X(t), V(t)) = (x, v)$ at time t and

$$\dot{X}(s) = V(s), \quad \dot{V}(s) = F(s, X(s)), \tag{1.2}$$

where F denotes the force field that is collectively generated by all particles. The requirement that f be constant along the curves defined by (1.2) then leads to the relation

$$\begin{aligned}
0 &= \frac{d}{ds} [f(s, X(s), V(s))] \\
&= \partial_t f(s, X(s), V(s)) + V(s) \cdot \nabla_x f(s, X(s), V(s)) \\
&\quad + F(s, X(s)) \cdot \nabla_v f(s, X(s), V(s))
\end{aligned}$$

for all s. Evaluated at time t, this yields

$$\partial_t f(t, t, v) + v \cdot \nabla_x f(t, x, v) + F(t, x) \cdot \nabla_v f(t, x, v) = 0$$

for all (t, x, v), which is usually called the Vlasov equation (despite the historic inadequacy of this terminology). The next step is to express the force field F in terms of the distribution function f. Since we are aiming at describing gravitational binding, we need to have $F \sim -\nabla_x V_C$ for the Coulomb potential $V_C(x) = -\frac{1}{|x|}$ at large distances. This suggests to use the field $F = -\nabla_x U$ induced by the Poisson equation

$$\Delta_x U_f(t, x) = 4\pi \rho_f(t, x), \quad \lim_{|x| \to \infty} U_f(t, x) = 0,$$

$$\text{where} \quad \rho_f(t, x) = \int_{\mathbb{R}^3} f(t, x, v) \, dv \tag{1.3}$$

denotes the charge density induced by f. Observe that $\int_{\mathcal{X}} dx \, \rho_f(t, x)$ represents the number of particles at time t, of any velocity, which are located at some point $x \in \mathcal{X}$. Then

$$U_f(t, x) = -\int_{\mathbb{R}^3} \frac{\rho_f(t, y)}{|y - x|} \, dy \tag{1.4}$$

is Coulomb-like as $|x| \to \infty$.

To summarize, the Vlasov-Poisson system in the gravitational case is

$$\partial_t f(t, x, v) + v \cdot \nabla_x f(t, x, v) - \nabla_x U_f(t, x) \cdot \nabla_v f(t, x, v) = 0 \qquad (1.5)$$

together with (1.3), and the equations are supposed to hold for $(t, x, v) \in \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}^3$. Initial data $f(0, x, v) = f_0(x, v)$ at time $t = 0$ have to be specified for f only, since then (1.4) determines the initial data $U_f(0, x)$. We will exclusively be interested in classical solutions of (1.5) and (1.3), whose global-in-time existence is ensured, under reasonable assumptions on f_0, by [55, 67, 80]. For a mathematical overview of the system and more background material, the reader may wish to consult [27, 63, 73].

The gravitational Vlasov-Poisson system is widely used to describe non-relativistic galactic dynamics. When it comes to relativistic galactic dynamics, the appropriate model is the Einstein-Vlasov system [2]. In the present book, we will not be dealing with this more general system, but of course it will be tempting to determine which results could be transferred to the Einstein-Vlasov system; see [18, 19, 31, 32, 38–40] for work in this context that is related to the so-called Antonov bound.

1.3 Steady State Solutions

The Vlasov-Poisson system possesses an abundance of solutions $Q = Q(x, v)$ that are independent of time. It is therefore of interest to study the stability of those steady states and, more ambitiously, the dynamics close to a steady state. Let $e_Q(x, v) = \frac{1}{2}|v|^2 + U_Q(x)$ denote the particle energy and let $\ell^2 = |L|^2 = |x|^2|v|^2 - (x \cdot v)^2$ be the square of the angular momentum $L = x \wedge v$. Then both e_Q and ℓ^2 are conserved along solutions of the characteristic equations $\ddot{X}(s) = -\nabla U_Q(X(s))$, which result from (1.2) for $F = -\nabla U_Q$; note that also U_Q is independent of time. Next, recall that a function $g = g(x, v)$ is said to be spherically symmetric if $g(Ax, Av) = g(x, v)$ for all $A \in SO(3)$ and $x, v \in \mathbb{R}^3$. Expressed in more sophisticated terms, g needs to be equivariant w.r. to the group action $SO(3) \times (\mathbb{R}^3 \times \mathbb{R}^3) \to \mathbb{R}^3 \times \mathbb{R}^3$, $(A, x, v) \mapsto (Ax, Av)$. Now, it is the content of Jeans's theorem that the distribution function Q of every spherically symmetric steady state solution has to be of the form $Q = Q(e_Q, \ell^2)$; see [7, Section 2] for a precise formulation. Such steady state solutions are called non-isotropic, in contrast to the isotropic ones, which can be written as $Q = Q(e_Q)$; a solution of the latter form will necessarily be spherically symmetric [25, 72]. Observe that we are going to systematically abuse notation in that we consider $Q = Q(x, v)$ to be a function of (x, v) and at the same time write $Q = Q(e_Q, \ell^2)$ or $Q = Q(e_Q)$, which indicates that Q is a function of two or of one scalar variable(s); in general, no confusion will result from this simplification.

To precisely state our results later, we will focus on the isotropic case, and we need to introduce the following assumptions (Q1)–(Q4) that we are going to impose throughout the book on the profile function $Q : \mathbb{R} \to [0, \infty[$ and the (radial) density

$\rho_Q : [0, \infty[\rightarrow [0, \infty[$. The diligent reader is invited to check which parts of this work remain valid under less restrictive hypotheses (there are several) or for non-isotropic steady states.

(Q1) The support $K = \text{supp } Q$ of the steady state solution Q is compact and its mass $\|Q\|_{L^1(\mathbb{R}^6)}$ is finite.

(Q2) $Q \in L^\infty_{\text{loc}}(\mathbb{R})$ satisfies $Q \geq 0$, and there exists a cut-off energy $e_0 < 0$ such that $Q(e) = 0$ for $e \geq e_0$, $Q \in C^1(]-\infty, e_0[)$ and $Q > 0$ in some interval $[e_1, e_0[$, where $e_1 < e_0$. For $\hat{e} \in]U_Q(0), e_0[$, there exists $\varepsilon > 0$ such that

$$\inf\{|Q'(e)| : e \in [\hat{e} - \varepsilon, \hat{e} + \varepsilon]\} > 0.$$

(Q3) $Q' \in L^\infty_{\text{loc}}(\mathbb{R})$ and $Q'(e) \leq 0$ a.e.

(Q4) ρ_Q is continuous and has compact support $\text{supp } \rho_Q = [0, r_Q]$. In addition, $\rho_Q \in C^1([0, r_Q])$.

For one result (Corollary 4.17), we will need more precise information on the behavior of Q' close to $e = e_0$.

(Q5) There are constants $C > 0$ and $\alpha > 0$ such that

$$|Q'(e)| \leq C(e_0 - e)^\alpha, \quad e \in [U_Q(0), e_0[.$$

1.4 Examples

To illustrate that the general assumptions on Q as stated in Sect. 1.3 are verified in many cases, we consider the steady state solution class of the polytropes and the King models in some more detail. It should be remarked that many further examples could be given, for instance by using [69] or [74, Theorem 3.1(a)], which basically says that under mild technical assumptions on Q and if $Q(e) = C(e_0 - e)^k_+ + \mathcal{O}((e_0 - e)^{k+\delta}_+)$ as $e \to e_0-$ for some $e_0 < 0$, $k \in]-\frac{1}{2}, \frac{3}{2}[$, $C > 0$ and $\delta > 0$, then the resulting steady state solution will have a finite radius and finite mass.

1.4.1 Polytropes

We consider the polytropes

$$Q(e_Q) = (e_0 - e_Q)^k_+ \tag{1.6}$$

for a cut-off energy $e_0 < 0$ and $k \in]-\frac{1}{2}, \frac{7}{2}[$. Then

$$\rho_Q(r) = c_n(e_0 - U_Q(r))_+^n, \quad n = k + \frac{3}{2} \in]1,5[, \quad c_n = (2\pi)^{3/2} \frac{\Gamma(k+1)}{\Gamma(k+\frac{5}{2})};$$

see [7, Example 4.1]. All these steady state solutions do have finite radius r_Q (i.e., compact support) and finite mass $M_Q = \int_{\mathbb{R}^3} \rho_Q(x)\,dx = 4\pi \int_0^{r_Q} r^2 \rho_Q(r)\,dr = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} Q(x,v)\,dx\,dv$. The limiting case $k = 7/2$ is called the Plummer sphere, where M_Q is still finite, but $r_Q = \infty$. We have $Q'(e) = -k(e_0 - e)_+^{k-1} \le 0$ (outside of $e = e_0$ for $k \le 1$) and $\rho_Q \in C^1([0, r_Q])$. Thus, if we take $k > 1$ for simplicity, then assumptions (Q1)–(Q5) are satisfied.

1.4.2 King models

The ansatz function for the King model [9, pp. 307–311] is given by

$$Q(e_Q) = (\exp(e_0 - e_Q) - 1)_+$$

for some cut-off energy $e_0 < 0$. Then $Q \in C^1(]-\infty, e_0[)$ and $Q'(e) = -\exp(e_0 - e) \le 0$ for $e < e_0$. The associated steady state solution does exist and has finite radius and finite mass; see [74, Theorem 3.1(a) and Sect. 4]. The density is found to be

$$\begin{aligned}
\rho_Q(r) &= \int_{\mathbb{R}^3} Q(x,v)\,dv \\
&= \int_{\mathbb{R}^3} \left(\exp\left(e_0 - \frac{1}{2}|v|^2 - U_Q(r)\right) - 1 \right)_+ dv \\
&= (\sqrt{2\pi})^3 \left(e^s \operatorname{erf}(\sqrt{s}) - \sqrt{\frac{4s}{\pi}} \left(1 + \frac{2s}{3}\right) \right), \quad s = e_0 - U_Q(r),
\end{aligned}$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$ denotes the error function, which has the asymptotic expansion $\operatorname{erf}(x) = \frac{2x}{\sqrt{\pi}} - \frac{2x^3}{3\sqrt{\pi}} + \mathcal{O}(x^5)$ as $x \to 0$. For $\varphi(s) = e^s \operatorname{erf}(\sqrt{s}) - \sqrt{\frac{4s}{\pi}}(1 + \frac{2s}{3})$, this yields the asymptotic expansion $\varphi(s) = \frac{8s^{5/2}}{15\sqrt{\pi}} + \mathcal{O}(s^{7/2})$ as $s \to 0^+$. Since $U_Q \in C^2([0, \infty[)$, we infer that in particular $\rho_Q'(r_Q) = 0$ holds and it follows that assumptions (Q1)–(Q4) are satisfied. However, since $Q(e) = (e_0 - e) + \mathcal{O}((e_0 - e)^2)$ as $e \to e_0-$, assumption (Q5) does not hold for the King model.

1.5 Linearization and the Antonov Stability Estimate

Without being too precise about its properties, we consider an isotopic steady state solution $Q = Q(e_Q)$. To study the stability of Q, we will closely follow [30] and write $f(t) = Q + g(t)$ with g 'small'. The total energy

$$\mathcal{H}(f(t)) = \frac{1}{2} \int_{\mathbb{R}^3} |v|^2 \, f(t, x, v) \, dx \, dv - \frac{1}{8\pi} \int_{\mathbb{R}^3} |\nabla U_{f(t)}(t, x)|^2 \, dx$$

is conserved along solutions, so it could be suspected to be a Lyapunov function. The expansion about Q then yields

$$\mathcal{H}(f(t)) = \mathcal{H}(Q) + \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\frac{1}{2} |v|^2 + U_Q \right) g(t) \, dx \, dv$$
$$- \frac{1}{8\pi} \int_{\mathbb{R}^3} |\nabla U_{g(t)}|^2 \, dx + \mathcal{O}(g^3); \tag{1.7}$$

note that $f \mapsto U_f$ is linear. The linear term on the right-hand side of (1.7) does not vanish, i.e., Q is not a critical point of $\mathcal{H}$. However, this defect can be remedied by making use of the fact that every 'Casimir functional'

$$\mathcal{C}_\Phi(f(t)) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \Phi(f(t, x, v)) \, dx \, dv$$

is also conserved along solutions, provided that Φ is sufficiently well-behaved. Passing from $\mathcal{H}$ to

$$\mathcal{H}_\Phi = \mathcal{H} + \mathcal{C}_\Phi$$

and repeating the expansion, one arrives at

$$\mathcal{H}_\Phi(f(t)) = \mathcal{H}_\Phi(Q) + \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} (e_Q + \Phi'(Q)) \, g(t) \, dx \, dv$$
$$+ \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \Phi''(Q) \, g(t)^2 \, dx \, dv - \frac{1}{8\pi} \int_{\mathbb{R}^3} |\nabla U_{g(t)}|^2 \, dx + \mathcal{O}(g^3). \tag{1.8}$$

Writing $e = e_Q$, since $Q = Q(e)$, the equation $e + \Phi'(Q(e)) = 0$ can be (formally) solved by taking $\Phi'(\xi) = -Q^{-1}(\xi)$, at least if for instance $Q'(e) < 0$ is verified for the relevant e in the support of Q. Then Q becomes a critical point of this $\mathcal{H}_\Phi$, and due to $1 + \Phi''(Q(e))Q'(e) = 0$ and $Q'(e) < 0$, the expansion (1.8) simplifies to

$$\mathcal{H}_\Phi(f(t)) = \mathcal{H}_\Phi(Q) + \frac{1}{2} \mathcal{A}(g(t), g(t)) + \mathcal{O}(g^3),$$
$$\mathcal{A}(g, g) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{dx \, dv}{|Q'(e_Q)|} |g|^2 - \frac{1}{4\pi} \int_{\mathbb{R}^3} |\nabla_x U_g|^2 \, dx. \tag{1.9}$$

Thus, one can expect that the stability of Q will be determined by the properties of the quadratic (second variation) part $\mathcal{A} = 2 \, D^2 \mathcal{H}_\Phi(Q)$, which we will call the Antonov functional. It should also be noted that $\mathcal{A}(g(t), g(t))$ is conserved along solutions $g(t)$ of the system that is linearized about Q; see [63, Prop. 3.2] and (1.21) below.

If we now consider functions $u = u(x, v)$ that are spherically symmetric and odd in v, i.e., they satisfy $u(x, -v) = -u(x, v)$, then the celebrated Antonov stability estimate [4, 5] is

$$\mathcal{A}(\mathcal{T}u, \mathcal{T}u) \geq c \|u\|_Q^2 \tag{1.10}$$

for some constant $c > 0$ that only depends on Q, where

$$\mathcal{T}g = \{g, e_Q\} = v \cdot \nabla_x g - \nabla_v g \cdot \nabla_x U_Q \tag{1.11}$$

for the standard Poisson bracket $\{g, h\} = \nabla_x g \cdot \nabla_v h - \nabla_v g \cdot \nabla_x h$. The weighted inner product

$$(g, h)_Q = \iint\limits_K \frac{1}{|Q'(e_Q)|} \overline{g(x, v)} h(x, v) \, dx \, dv \tag{1.12}$$

induces the norm $\| \cdot \|_Q$, and $K = \operatorname{supp} Q \subset \mathbb{R}^6$ denotes the support of the steady state solution Q, which is compact, if (Q1) holds. Perturbations of the form $g = \mathcal{T}u$ are called 'dynamically accessible', for reasons explained in [62]; also see [66]. Antonov [4, 5] could prove that the positive definiteness (1.10) is equivalent to the linear stability of Q. Many works followed these pioneering observations, and until to date, almost all stability proofs, linear or nonlinear, use the Antonov stability estimate in one way or another. The bound (1.10), or variations thereof, is applied in a number of papers, both in the physics and in the mathematics community, to address a variety of stability issues; see [15, 26, 28, 30, 42, 43, 50, 51, 60, 89] and many further.

1.6 The Best Constant in the Antonov Stability Estimate

In this section, we will explain the connection of the functional $u \mapsto \mathcal{A}(\mathcal{T}u, \mathcal{T}u)$ from (1.10) to a certain self-adjoint operator L. Before doing so, we need to introduce some relevant notation, function spaces, etc. Since we restrict ourselves to isotropic steady states, the solutions will be spherically symmetric. Thus, we will consider (1.5) and (1.3) in the spherical symmetric framework only, and it is well-known [7] that then the system can be written as

$$\partial_t f(t, r, p_r, \ell^2) + p_r \, \partial_r f(t, r, p_r, \ell^2) + \left(\frac{\ell^2}{r^3} - \partial_r U_f(t, r) \right) \partial_{p_r} f(t, r, p_r, \ell^2) = 0$$

and

$$U_f''(t, r) + \frac{2}{r} U_f'(t, r) = 4\pi \rho_f(t, r), \quad \lim_{r \to \infty} U_f(t, r) = 0,$$
$$\rho_f(t, r) = \frac{2\pi}{r^2} \int_0^\infty d\ell \, \ell \int_{\mathbb{R}} dp_r \, f(t, r, p_r, \ell^2), \tag{1.13}$$

the $'$ indicating $\frac{d}{dr}$ or ∂_r, and $p_r = \frac{x \cdot v}{r}$. If $g = g(x, v)$ is spherically symmetric, then $\rho_g(x) = \rho_g(r)$ and $U_g(x) = U_g(r)$ are radially symmetric, and we will in general denote

$$\rho_g(x) = \int_{\mathbb{R}^3} g(x, v)\, dv, \quad U_g(x) = -\int_{\mathbb{R}^3} \frac{\rho_g(y)}{|y - x|}\, dy. \tag{1.14}$$

Also $g = g(x, v)$ can be identified with a function $g = g(r, p_r, \ell)$ or $g = g(r, p_r, \ell^2)$; see Appendix I, Section A.1.

Next, define the linear operator $\mathcal{K}$ by

$$\mathcal{K}g = \{Q, U_g\};$$

it should be mentioned that both $\mathcal{T}$ from (1.11) and $\mathcal{K}$ do arise naturally upon linearizing the Vlasov-Poisson system about Q; see (1.21) below. Since $U_g(x) = U_g(|x|) = U_g(r)$, we obtain

$$\mathcal{K}g = \{Q, U_g\} = -\nabla_v Q \cdot \nabla_x U_g = -Q'(e_Q)\, v \cdot \frac{x}{r}\, U_g'(r) = -Q'(e_Q)\, p_r\, U_g'(r). \tag{1.15}$$

The operator L is introduced as

$$Lu = -\mathcal{T}^2 u - \mathcal{K}\mathcal{T}u. \tag{1.16}$$

For what concerns the appropriate function spaces, we will pass to action-angle variables as follows. On $K = \operatorname{supp} Q$, we consider the equation

$$\ddot{r} = -U'_{\text{eff}}(r, \ell), \tag{1.17}$$

where $U_{\text{eff}}(r, \ell) = U_Q(r) + \frac{\ell^2}{2r^2}$ is the effective potential that occurs in the energy function

$$e_Q = e_Q(r, p_r, \ell) = \frac{1}{2}|v|^2 + U_Q(r) = \frac{1}{2}p_r^2 + U_{\text{eff}}(r, \ell),$$

where $p_r = \dot{r}$ is the radial velocity and ℓ should be thought of as fixed. By standard Hamiltonian system theory (see Section A.1 for details), it is then possible to write spherically symmetric functions $g = g(x, v) = g(r, p_r, \ell)$ in the form $g = g(\theta, I, \ell)$ if we apply a canonical transformation $(\theta, I) \mapsto (r, p_r)$ at fixed ℓ. Working in action-angle variables has many advantages. First of all, it turns out that e_Q becomes a function of (I, ℓ) alone, $e_Q = E(I, \ell)$. Secondly, the functions g are 2π-periodic in θ, so they can be conveniently represented as a Fourier series

$$g(\theta, I, \ell) = \sum_{k \in \mathbb{Z}} g_k(I, \ell)\, e^{ik\theta},$$

where

$$g_k(I, \ell) = \frac{1}{2\pi} \int_0^{2\pi} g(\theta, I, \ell)\, e^{-ik\theta}\, d\theta$$

are the Fourier coefficients. The spaces X_{odd}^α (cf. Appendix II, Sect. B.1) are defined in terms of this series representation by means of the norms

$$\|g\|_{X^\alpha}^2 \sim \sum_{k \in \mathbb{Z}} (1 + k^2)^\alpha \|g_k\|_{L^2_{\frac{1}{|Q'|}}(D)}^2,$$

where $L^2_{\frac{1}{|Q'|}}(D)$ is a weighted L^2-space on the domain D of the variables (I, ℓ). The subscript 'odd' in X_{odd}^α indicates that the functions are odd in v, which translates into the condition $g_{-k} = -g_k$ for $k \in \mathbb{Z}$ on the coefficients (so that in particular $g_0 = 0$).

Now we can give a precise meaning to the fact that $\mathcal{A}(\mathcal{T}u, \mathcal{T}u) = (Lu, u)_Q$ is the quadratic form associated with the operator L from (1.16). We have the following result.

Lemma 1.1 L *is self-adjoint on the domain* $\mathcal{D}(L) = X_{\text{odd}}^2$ *in* X_{odd}^0. *In addition,* $(Lu, u)_Q = \mathcal{A}(\mathcal{T}u, \mathcal{T}u)$ *holds for* $u \in X_{\text{odd}}^2$.

Proof Most of this will be shown later; see Corollary B.19 for the properties of L. At this point, let us just mention that by (B.44) in Corollary B.19 the term $(\mathcal{K}\mathcal{T}u, u)_Q$ can be written as $\frac{1}{4\pi} \int_{\mathbb{R}^3} |\nabla_x U_{\mathcal{T}u}|^2 \, dx$. Hence, we deduce that

$$\begin{aligned} (Lu, u)_Q &= (-\mathcal{T}^2 u, u)_Q - (\mathcal{K}\mathcal{T}u, u)_Q \\ &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{dx \, dv}{|Q'(e_Q)|} |\mathcal{T}u|^2 - \frac{1}{4\pi} \int_{\mathbb{R}^3} |\nabla_x U_{\mathcal{T}u}|^2 \, dx \\ &= \mathcal{A}(\mathcal{T}u, \mathcal{T}u); \end{aligned} \tag{1.18}$$

recall (1.9). $\qquad\square$

As a consequence, we can re-express (1.10) as follows.

Theorem 1.2 (Antonov stability estimate) *If* $u \in X_{\text{odd}}^2$, *then*

$$(Lu, u)_Q = \mathcal{A}(\mathcal{T}u, \mathcal{T}u) \geq c \|u\|_Q^2 \tag{1.19}$$

for $c = \frac{1}{r_Q^3} \|Q\|_{L^1(\mathbb{R}^6)} > 0$, *where* $\operatorname{supp} \rho_Q = [0, r_Q]$.

We will indicate a proof of Theorem 1.2 in Chapter 2. Therefore,

$$\lambda_* = \inf \{(Lu, u)_Q : u \in X_{\text{odd}}^2, \|u\|_Q = 1\} > 0 \tag{1.20}$$

is well-defined; it is the 'best constant' in the Antonov stability estimate and a main object of study in the present work. We will derive many results related to λ_*, as will be described in Section 1.8. In particular, we will be able to characterize the cases where λ_* is attained, in the sense that $\lambda_* = (Lu_*, u_*)_Q$ for some minimizing function $u_* \in X_{\text{odd}}^2$ such that $\|u_*\|_Q = 1$. It turns out that then u_* will be an eigenfunction of L corresponding to the eigenvalue λ_*, so that $Lu_* = \lambda_* u_*$. The quantity λ_* will

be of fundamental importance for the dynamics of the gravitational Vlasov-Poisson system.

Lemma 1.3 *Let $u_* \in X_{odd}^2$ be a minimizer and define*

$$g_*(t, x, v) = \cos(\sqrt{\lambda_*}t)\, u_*(x, v) - \frac{1}{\sqrt{\lambda_*}}\, \sin(\sqrt{\lambda_*}t)\, (\mathcal{T}u_*)(x, v).$$

Then g_* is a $\frac{2\pi}{\sqrt{\lambda_*}}$-periodic solution of the equation

$$\partial_t g + \mathcal{T}g + \mathcal{K}g = 0 \tag{1.21}$$

that is obtained by linearizing (1.5) and (1.3) about Q.

Proof To linearize the system about Q, let $f = Q + g$ as before. As a consequence of the fact that $v \cdot \nabla_x f - \nabla_x U_f \cdot \nabla_v f = \{f, e_f\}$ for $e_f(x, v) = \frac{1}{2}\,|v|^2 + U_f(x)$, we may write

$$
\begin{aligned}
0 = \partial_t f + \{f, e_f\} &= \partial_t g + \left\{Q + g, \frac{1}{2}\,|v|^2 + U_Q + U_g\right\} \\
&= \partial_t g - \nabla_v Q \cdot \nabla_x U_g + v \cdot \nabla_x g - \nabla_v g \cdot \nabla_x U_Q - \nabla_v g \cdot \nabla_x U_g,
\end{aligned}
$$

which is equivalent to

$$\partial_t g + \mathcal{T}g + \mathcal{K}g = \nabla_v g \cdot \nabla_x U_g. \tag{1.22}$$

Thus, (1.21) is indeed the linearization. Next, note that u_* is odd in v. Hence, $\rho_{u_*}(x) = \int_{\mathbb{R}^3} u_*(x, v)\, dv = 0$ implies that $U_{u_*} = 4\pi\Delta^{-1}\rho_{u_*} = 0$ and therefore $\mathcal{K}u_* = 0$ by (1.15). Consequently,

$$
\begin{aligned}
\partial_t g_* + \mathcal{T}g_* + \mathcal{K}g_* = &-\sqrt{\lambda_*}\sin(\sqrt{\lambda_*}t)\, u_* - \cos(\sqrt{\lambda_*}t)\, \mathcal{T}u_* \\
&+ \cos(\sqrt{\lambda_*}t)\, \mathcal{T}u_* - \frac{1}{\sqrt{\lambda_*}}\sin(\sqrt{\lambda_*}t)\, \mathcal{T}^2 u_* \\
&+ \cos(\sqrt{\lambda_*}t)\, \mathcal{K}u_* - \frac{1}{\sqrt{\lambda_*}}\sin(\sqrt{\lambda_*}t)\, \mathcal{K}\mathcal{T}u_* \\
= &-\sqrt{\lambda_*}\sin(\sqrt{\lambda_*}t)\, u_* + \frac{1}{\sqrt{\lambda_*}}\sin(\sqrt{\lambda_*}t)\, L u_* = 0,
\end{aligned}
$$

as claimed.   □

At present, it is not known if periodic solutions to (1.5) and (1.3) close to steady state solutions do exist; see [17, 56, 70]. However, in this case, $\frac{2\pi}{\sqrt{\lambda_*}}$ will conceivably be the limiting period of the oscillations, a fact for which there is some numerical evidence [70]. To give a heuristic argument, suppose that g_ε is an ε-small and T_ε-periodic solution to (1.22) such that $T_\varepsilon \to T_0$ as $\varepsilon \to 0$. Then, $\tilde{g}_\varepsilon = \varepsilon^{-1}g_\varepsilon$ will be of order one, T_ε-periodic and satisfies

$$\partial_t \tilde{g}_\varepsilon + \mathcal{T}\tilde{g}_\varepsilon + \mathcal{K}\tilde{g}_\varepsilon = \varepsilon \nabla_v \tilde{g}_\varepsilon \cdot \nabla_x U_{\tilde{g}_\varepsilon}.$$

Assuming now that $\tilde{g}_\varepsilon \to \tilde{g}_*$ in a suitable norm, $\tilde{g}_* \neq 0$ will be T_0-periodic and such that

$$\partial_t \tilde{g}_* + \mathcal{T}\tilde{g}_* + \mathcal{K}\tilde{g}_* = 0.$$

If $\partial_t + \mathcal{T} + \mathcal{K}$ does have a one-dimensional kernel, then $\tilde{g}_*$ is proportional to g_* from Lemma 1.3, and hence $T_0 = \frac{2\pi}{\sqrt{\lambda_*}}$.

1.7 Domains in Action-Angle Variables

Before we will be able to describe our main results and their connection to the Birman-Schwinger principle in Sect. 1.8, we have to take a closer look at the domains that occur as the supports of steady state solutions, expressed in action-angle variables (θ, I, ℓ); recall that the particle energy $e_Q = E(I, \ell)$ is a function of (I, ℓ) alone.

The frequency functions associated with the energy E are

$$\omega_1(I, \ell) = \frac{\partial E(I, \ell)}{\partial I}, \quad \omega_2(I, \ell) = \frac{\partial E(I, \ell)}{\partial L_3} = 0, \quad \omega_3(I, \ell) = \frac{\partial E(I, \ell)}{\partial \ell},$$

where (I, L_3, ℓ) are the action variables. We would like to emphasize that ω_1, together with the corresponding period function $T_1(I, \ell) = \frac{2\pi}{\omega_1(I, \ell)}$, will be a main player in the game, and understanding its properties will be of central importance. This is due to the fact that in action-angle variables the operator $\mathcal{T}$ from (1.11) is found to be very simple:

$$(\mathcal{T}g)(\theta, I, \ell) = \omega_1(I, \ell)\, \partial_\theta g(\theta, I, \ell),$$

or $g_k \mapsto ik\omega_1\, g_k$ in terms of the Fourier coefficients. Since ω_1 is independent of θ, this also yields $-\mathcal{T}^2 g = -\omega_1^2\, \partial_\theta^2 g$ or $g_k \mapsto k^2\omega_1^2\, g_k$. It will turn out (see Section 3.1) that ω_1 is strictly positive, so that, at fixed ℓ, the map $I \mapsto E(I, \ell)$ is strictly increasing. Therefore, it can be inverted as a map $E \mapsto I(E, \ell)$, and accordingly functions $g = g(\theta, I, \ell)$ can be viewed as functions $\tilde{g}(\theta, E, \ell) = g(\theta, I(E, \ell), \ell)$ and vice versa.

From (Q1)–(Q4) in Sect. 1.3, the following can be shown (cf. the argument in Sect. 1.7.1 below):

In action-angle variables, one has

$$K = \{(\theta, E, \beta) : \theta \in [0, 2\pi], \beta \in [0, \beta_*], E \in [e_{\min}(\beta), e_0]\},$$

where $E = E(I, \ell)$ and $I = I(E, \ell)$. Furthermore, $\beta = \ell^2$, $\beta_* > 0$, and

$$e_{\min}(\beta) = U_{\text{eff}}(r_0(\beta), \beta)$$

is the minimal energy of the effective potential $U_{\text{eff}}(\cdot, \beta)$, which is attained at the unique point $r_0(\beta)$; see Appendix I, Sect. A.1. Also, $e_{\min}(\cdot)$ is non-decreasing and

$$\min\{e_{\min}(\beta) : \beta \in [0, \beta_*]\} = U_Q(0) < e_0,$$
$$\max\{e_{\min}(\beta) : \beta \in [0, \beta_*]\} = e_0.$$

We will always denote

$$D = \{(E, \beta) : \beta \in [0, \beta_*], E \in [e_{\min}(\beta), e_0]\},$$

which at times will be expressed in terms of ℓ as

$$D = \{(E, \ell) : \ell \in [0, \ell_*], E \in [e_{\min}(\ell), e_0]\}, \tag{1.23}$$

and similarly, we will write

$$K = \{(\theta, E, \ell) : \theta \in [0, 2\pi], \ell \in [0, l_*], E \in [e_{\min}(\ell), e_0]\}. \tag{1.24}$$

It is also understood that K and D can be written in terms of the variables (θ, I, β), (θ, I, ℓ) and (I, β), (I, ℓ), respectively, without this being reflected by renaming the sets. In any case, we will always have $K = [0, 2\pi] \times D$.

For illustration, we are going to determine K and D for the polytropes and the King models, respectively. A general domain D is shown in Fig. 1.1.
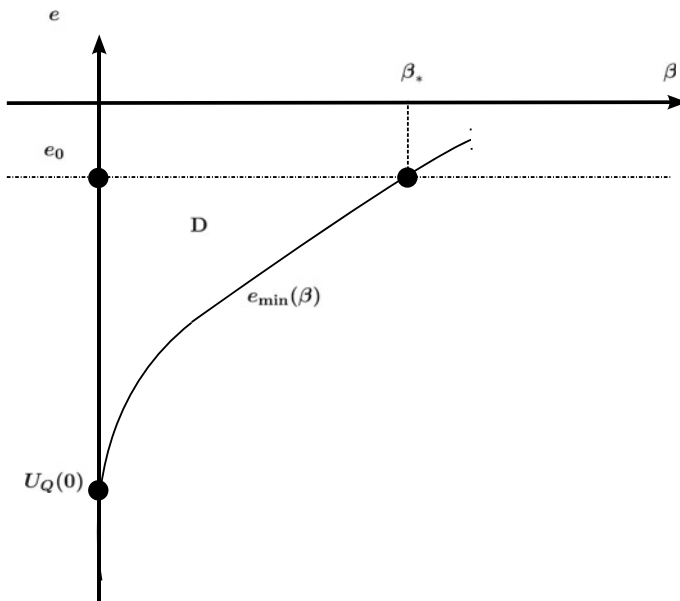

Fig. 1.1 The domain D in coordinates $(e, \beta) = (E, \beta)$

1.7.1 Polytropes Revisited

We wish to determine the support

$$K = \operatorname{supp} Q = \{e_0 - e_Q \geq 0\}$$

of the polytropes in terms of $\beta = \ell^2$ and $e = e_Q$. More precisely, since always $\theta \in [0, 2\pi]$ on K for the angular variable θ, we have to exhibit a set D of (e, β) such that $K = [0, 2\pi] \times D$. On this domain D, we need to have

$$e_0 \geq e \geq U_{\text{eff}}(r, \beta) \geq U_{\text{eff}}(r_0(\beta), \beta) = U_Q(r_0(\beta)) + \frac{\beta}{2r_0(\beta)^2}, \qquad (1.25)$$

with $r_0(\beta)$ denoting the unique point where the effective potential $U_{\text{eff}}(r, \beta) = U_Q(r) + \frac{\beta}{2r^2}$ attains its minimum value $e_{\min}(\beta) = U_{\text{eff}}(r_0(\beta), \beta)$. From (1.25), we get

$$2r_0(\beta)^2 (e_0 - U_Q(r_0(\beta))) \geq \beta.$$

Let

$$J = \{\beta \geq 0 : 2r_0(\beta)^2 (e_0 - U_Q(r_0(\beta))) \geq \beta\}.$$

First, we claim that J is an interval. To see this, note that

$$2r^2(e_0 - U_{\text{eff}}(r, \beta)) + \beta = 2r^2\left(e_0 - U_Q(r) - \frac{\beta}{2r^2}\right) + \beta = 2r^2(e_0 - U_Q(r)).$$

Therefore,

$$2r^2(e_0 - U_Q(r)) \geq \beta \quad \Longleftrightarrow \quad U_{\text{eff}}(r, \beta) \leq e_0,$$

which implies that

$$J = \{\beta \geq 0 : e_{\min}(\beta) \leq e_0\}. \qquad (1.26)$$

Now $\beta \mapsto e_{\min}(\beta)$ is increasing by Lemma A.7(c) below (which is a general result), and thus J has to be an interval.

The next aim is to show that $[0, \varepsilon] \subset J$ for some $\varepsilon > 0$ small enough. For, by Lemma A.7(f), we have

$$r_0(\beta)^4 = \frac{1}{A(0)} \beta + \mathcal{O}(\beta^2) \quad \text{and} \quad e_{\min}(\beta) = U_Q(0) + \mathcal{O}(\beta^{1/2})$$

as $\beta \to 0^+$. Since $U_Q(0) < e_0$ (the cut-off energy), the condition $e_{\min}(\beta) \leq e_0$ from the characterization of J in (1.26) is satisfied with strict inequality at $\beta = 0$. It follows that $[0, \varepsilon] \subset J$ if $\varepsilon > 0$ is sufficiently small.

Now, we are going to show that J is bounded. First, if $\beta \in J$, then $r_0(\beta) \leq r_Q$, where $\operatorname{supp} \rho_Q = [0, r_Q]$. Otherwise, we would have $r_0(\beta) > r_Q$ for some

$\beta \in J \setminus \{0\}$. Since r_Q is characterized by $U_Q(r_Q) = e_0$, this gives $U_Q(r_0(\beta)) > e_0$, and consequently $\beta \le 2r_0(\beta)^2 (e_0 - U_Q(r_0(\beta))) \le 0$, which is a contradiction. Then, $r_0(\beta) \le r_Q$ for $\beta \in J$ in turn leads to the boundedness of J, owing to

$$\beta \le 2r_0(\beta)^2 (e_0 - U_Q(r_0(\beta))) \le 2r_Q^2 (e_0 - U_Q(r_0(\beta))) \le 2r_Q^2 (e_0 - U_Q(0))$$

uniformly for $\beta \in J$.

Lastly, we will check that $\beta_* = \max J$ satisfies $e_{\min}(\beta_*) = e_0$. In fact, at β_*, we must have $2r_0(\beta_*)^2 (e_0 - U_Q(r_0(\beta_*))) = \beta_*$. Thus,

$$e_{\min}(\beta_*) = U_{\text{eff}}(r_0(\beta_*), \beta_*) = U_Q(r_0(\beta_*)) + \frac{\beta_*}{2r_0(\beta_*)^2} = e_0. \tag{1.27}$$

To summarize, since the condition on e is $e_0 \ge e \ge e_{\min}(\beta)$, we have shown that

$$D = \{(\beta, e) : \beta \in [0, \beta_*], e \in [e_{\min}(\beta), e_0]\}$$

and $K = [0, 2\pi] \times D$ for the support K of Q in terms of e and β, and the lower boundary curve $[0, \beta_*] \ni \beta \mapsto e_{\min}(\beta)$ strictly increases from $U_Q(0)$ to e_0.

We would also like to point out that $r_0(\beta_*) \in]0, r_Q[$. By construction, one has $r_0(\beta_*) \le r_Q$, so suppose that we had $r_0(\beta_*) = r_Q$. Since $r_0(\beta)^3 U_Q'(r_0(\beta)) = \beta$, (1.27) yields

$$e_0 = U_Q(r_0(\beta_*)) + \frac{\beta_*}{2r_0(\beta_*)^2} = U_Q(r_0(\beta_*)) + \frac{1}{2} r_0(\beta_*) U_Q'(r_0(\beta_*)). \tag{1.28}$$

But $U_Q(r_Q) = e_0$, whence $0 = U_Q'(r_Q) = \frac{4\pi}{r_Q^2} \int_0^{r_Q} s^2 \rho_Q(s) \, ds$, which is a contradiction. The relation (1.28) characterizes $r_0(\beta_*)$, since $\varphi(r) = U_Q(r) + \frac{1}{2} r U_Q'(r)$ satisfies $\varphi'(r) = \frac{1}{2} r B(r)$ for $B(r) = \frac{U_Q'(r)}{r} + 4\pi \rho_Q(r) > 0$ from Lemma A.6(b). In addition, $\varphi(0) = U_Q(0) < e_0$ and $\varphi(r_Q) = e_0 + \frac{1}{2} r_Q U_Q'(r_Q) > e_0$.

Finally, observe that the reasoning in this section did not depend on the specific form of the polytropic ansatz function (1.6), but only on the general properties of the functions $r_0(\beta)$ and $e_{\min}(\beta)$.

1.7.2 King Models Revisited

Exactly as in Sect. 1.7.1, here we also get

$$K = \operatorname{supp} Q = [0, 2\pi] \times D, \quad D = \{(\beta, e) : \beta \in [0, \beta_*], e \in [e_{\min}(\beta), e_0]\},$$

for the corresponding functions $r_0(\beta)$ and $e_{\min}(\beta) = U_{\text{eff}}(r_0(\beta), \beta)$. In addition, we have $r_0(\beta_*) \in]0, r_Q[$.

1.8 Summary of the Main Results

Now, we are in a position to outline the main results of this book. In Chap. 3, we will study the properties of ω_1 or equivalently of T_1 in some detail. First, it is shown (Theorem 3.2) that

$$\delta_1 = \inf \{ \omega_1(e, \ell) : (e, \ell) \in \mathring{D} \} > 0. \tag{1.29}$$

This fact has been mentioned above and it will be used many times. The number δ_1, or more precisely δ_1^2, is intimately related to the spectrum of L, since $\delta_1^2 = \min \sigma_{\mathrm{ess}}(L)$ is the minimum of the essential spectrum of L. In this connection, let us also mention that the essential spectrum of L can be determined explicitly, and it is large in the sense that $[\lambda_c, \infty[\subset \sigma_{\mathrm{ess}}(L)$ for some $\lambda_c > \delta_1^2$. Furthermore, $\lambda_* \le \delta_1^2$ is satisfied (Section 3.4). Along with (1.29), we will also prove that

$$\Delta_1 = \sup \{ \omega_1(e, \ell) : (e, \ell) \in \mathring{D} \} < \infty;$$

see Theorem 3.5. Concerning the regularity of ω_1 or T_1, it is not very difficult to see that $T_1 \in C^1(\mathring{D})$, as will be derived in Theorem 3.6. It is considerably harder to verify that $T_1 \in C(D)$, i.e., that T_1 can be continuously extended to the boundary ∂D of D. This will be done in a series of lemmas, and the results are summarized in Theorem 3.13; the most challenging part is to make sure that T_1 is continuous at $(e, \beta) = (U_Q(0), 0)$, which is the lower left corner of D. It will also turn out that T_1 is increasing on the lower boundary curve of D (Lemma 3.14) and on the left boundary part of D (Lemma 3.15).

In Chapter 4, we are going to make the connection of the spectral problem for L to the Birman-Schwinger principle. We will be using an approach to reformulate the problem that is inspired by the physics reference [61], although this paper does neither use the operator L nor realize the underlying Birman-Schwinger principle. Let L_r^2 denote the L^2-Lebesgue space of radially symmetric functions $\Psi(x) = \Psi(r)$ on $\mathbb{R}^3$ with inner product

$$\langle \Psi, \Phi \rangle = \int_{\mathbb{R}^3} \overline{\Psi(x)} \, \Phi(x) \, dx = 4\pi \int_0^\infty r^2 \, \overline{\Psi(r)} \, \Phi(r) \, dr.$$

It will be shown that one can define a family $\mathcal{Q}_\lambda$ of non-negative Hilbert-Schmidt operators on L_r^2 with the following properties for $\lambda < \delta_1^2$:

(a) λ is an eigenvalue of L if and only if 1 is an eigenvalue of $\mathcal{Q}_\lambda$.

This observation provides a natural way for showing that λ_* is an eigenvalue of L, provided that one has $\lambda_* < \delta_1^2$ (i.e., there is a spectral gap). The first eigenvalue function $\mu_1(\lambda)$ of $\mathcal{Q}_\lambda$ turns out to be increasing in λ, and one has to locate the value of λ, where μ_1 becomes 1; in this way, we will be able to show that λ_* is attained. Furthermore, we will also prove:

(b) if $u \in X_{\mathrm{odd}}^2$ is an eigenfunction of L for the eigenvalue λ, then $\Psi = 4\pi \int_{\mathbb{R}^3} p_r u$ $dv \in L_r^2$ is an eigenfunction of $\mathcal{Q}_\lambda$ for the eigenvalue 1;

(c) if $\Psi \in L_r^2$ is an eigenfunction of $\mathcal{Q}_\lambda$ for the eigenvalue 1, then $u = (-\mathcal{T}^2 - \lambda)^{-1}(|Q'(e_Q)| p_r \Psi) \in X_{\mathrm{odd}}^2$ is an eigenfunction of L for the eigenvalue λ.

Thus, if we compare (a)–(c) for our galactic dynamics setup to (a)–(c) from the Schrödinger case in Sect. 1.1, then we see that both are formally identical if we associate $p_r \sim \sqrt{-V}$ and $-\Delta \sim -\mathcal{T}^2$ and furthermore disregard the velocity average $\int_{\mathbb{R}^3} dv$; the appearance of $|Q'(e_Q)|$ in $|Q'(e_Q)| p_r \Psi$ is due to the $(\cdot, \cdot)_Q$ that is used. There is yet another fact that supports the analogy of both approaches. One of the ways to represent $\mathcal{Q}_\lambda$ is

$$\mathcal{Q}_\lambda \Psi = 4\pi \int_{\mathbb{R}^3} p_r \, (-\mathcal{T}^2 - \lambda)^{-1} \, (|Q'(e_Q)| \, p_r \Psi) \, dv. \qquad (1.30)$$

Comparing this relation to (1.1), it turns out that both relations do agree if we apply the same identifications as before.

Throughout the book, we are going to exploit this Birman-Schwinger principle in galactic dynamics to deal with the question in which cases λ_* from (1.20) is attained. However, there seems to be a wide range of further possible applications that could for instance be related to a limiting absorption principle or L^p–L^q-estimates on the 'free resolvent' $(-\mathcal{T}^2 - \lambda)^{-1}$, in the spirit of [45] for the Laplacian. One advantage when dealing with (1.1) is that, in three dimensions, the operator has the explicit integral kernel

$$B_e(x, y) = \sqrt{-V(x)} \, \frac{1}{4\pi|x - y|} \, \exp(-\sqrt{e} \, |x - y|) \sqrt{-V(y)},$$

which allows for hands-on estimates. It would also be desirable to obtain something similar for (1.30).

The explicit form of the operator Q_λ is

$$\mathcal{Q}_\lambda : L_r^2 \to L_r^2,$$

$$(\mathcal{Q}_\lambda \Psi)(r) = \frac{16\pi}{r^2} \sum_{k \neq 0} \int_0^\infty d\tilde{r} \, \Psi(\tilde{r}) \iint_D d\ell \, \ell \, de \, \mathbf{1}_{\{r_-(e, \ell) \leq r, \tilde{r} \leq r_+(e, \ell)\}} \frac{\omega_1(e, \ell) \, |Q'(e)|}{k^2 \omega_1^2(e, \ell) - \lambda}$$

$$\times \sin(k\theta(r, e, \ell)) \sin(k\theta(\tilde{r}, e, \ell)),$$

where $r_\pm(e, \ell)$ are the maximal resp. minimal value of r along the orbit of (1.17) that has energy e, and $\theta(r, e, \ell)$ is the associated angle. Note that $\lambda < \delta_1^2$ implies $k^2 \omega_1^2(e, \ell) - \lambda \geq \delta_1^2 - \lambda > 0$ for $k \neq 0$, so the denominators do not vanish. It turns out that the family $\mathcal{Q}_\lambda$ can be analytically continued to $\mathcal{Q}_z$ for $z \in \Omega = \mathbb{C} \setminus [\delta_1^2, \infty[$, by simply replacing λ with z. In addition, we can write $(\mathcal{Q}_z \Psi)(r) = \langle K_{\bar{z}}(r, \cdot), \Psi \rangle$ for some $L^2 \times L^2$-integral kernel K, which allows us to show that each $\mathcal{Q}_z$ is a Hilbert-Schmidt operator on L_r^2. Furthermore, $\langle \mathcal{Q}_\lambda \Psi, \Psi \rangle \geq 0$ and $\lambda \to \langle \mathcal{Q}_\lambda \Psi, \Psi \rangle$

are increasing for real λ. Then, the spectrum of $\mathcal{Q}_\lambda$ consists of $\mu_1(\lambda) \geq \mu_2(\lambda) \geq \ldots \to 0$ (the eigenvalues are listed according to their multiplicities). In addition,

$$\mu_1(\lambda) = \|\mathcal{Q}_\lambda\| = \sup \{\langle \mathcal{Q}_\lambda \Psi, \Psi \rangle : \|\Psi\|_{L_r^2} \leq 1\},$$

where $\| \cdot \| = \| \cdot \|_{\mathcal{B}(L_r^2)}$, and every function

$$\mu_k(\cdot) : \,] - \infty, \delta_1^2[\to \,]0, \infty[$$

for $k \in \mathbb{N}$ is monotone increasing and locally Lipschitz continuous. According to the Birman-Schwinger characterization of an eigenvalue λ for L, we have to determine those k and λ, where $\mu_k(\lambda) = 1$. Since we expect $\lambda_* \leq \delta_1^2$ to be the principal eigenvalue of L, more specifically we need to find λ such that $\mu_1(\lambda) = 1$. In this respect, the quantity

$$\mu_* = \lim_{\lambda \to \delta_1^2 -} \mu_1(\lambda) = \sup \{\mu_1(\lambda) : \lambda \in [0, \delta_1^2[\} \in [\mu_1(0), \infty]$$

will be important, and in what follows, we are going to outline our results, depending on μ_*.

Let us first recall that $\delta_1^2 = \min \sigma_{\mathrm{ess}}(L)$, and if $\lambda_* < \delta_1^2$ and λ_* were an eigenvalue of L, then there would exist a spectral gap. We are going to prove in Theorem 4.13 that the conditions $\lambda_* < \delta_1^2$ and $\mu_* > 1$ are equivalent, and in this case, $\mu_1(\lambda_*) = 1$ and λ_* is an eigenvalue of L. The difficult part of the argument is to show that a spectral gap $\lambda_* < \delta_1^2$ forces λ_* to be an eigenvalue. This is accomplished by studying (at great length in Appendix C) a certain evolution equation, for which $\lambda_* < \delta_1^2$ translates into a compactness condition; the argument is summarized in Section C.1.

Next, we turn to the case where $\mu_* < 1$. Then necessarily $\lambda_* = \delta_1^2$, so there is no spectral gap and we cannot use the Birman-Schwinger principle. Nevertheless, it is possible to prove (Theorem 4.14) that now $\lambda_* = \delta_1^2$ is not an eigenvalue, provided that the following condition is satisfied:

(ω_1-1) $\{(I, \ell) \in D : \omega_1(I, \ell) = \delta_1\}$ has the Lebesgue measure zero.

This excludes (Lemma B.12) that δ_1^2 is an eigenvalue of $-\mathcal{T}^2$. The proof works by deriving suitable estimates for the operators $\mathcal{Q}_{\delta_1^2 - \varepsilon + i\varepsilon^3}$ in the limit $\varepsilon \to 0^+$. We would not be surprised if the case $\mu_* < 1$ could not occur at all, but we were not able to verify this.

The most pathological case seems to be $\mu_* = 1$. Then once again $\lambda_* = \delta_1^2$, there is no spectral gap and the Birman-Schwinger principle does not apply. To see that here one needs to add another condition on ω_1, let us change the perspective and ask where, in D, $\delta_1 = \inf_{\mathring{D}} \omega_1 = \min_D \omega_1$ is attained. If this happens at an interior point $(\hat{e}, \hat{\beta}) \in \mathring{D}$, then $\nabla \omega_1(\hat{e}, \hat{\beta}) = (0, 0)$ and the following condition will be verified:

(ω_1-2) There are a point $(\hat{e}, \hat{\beta}) \in \mathring{D}$, a neighborhood U of $(\hat{e}, \hat{\beta})$ and a constant $C_1 > 0$ such that $\omega_1(\hat{e}, \hat{\beta}) = \delta_1$ and

$$|\omega_1(e, \beta) - \delta_1| \leq C_1 |(e, \beta) - (\hat{e}, \hat{\beta})|^2, \quad (e, \beta) \in U. \tag{1.31}$$

But then Corollary 4.16 implies that $\mu_* = \infty$, which is not compatible with $\mu_* = 1$. Hence, we can assume that the minimum is attained at some point $(\hat{e}, \hat{\beta}) \in \partial D$, the boundary of D. According to Corollary 3.16, then $(\hat{e}, \hat{\beta})$ lies on the 'upper line' $\{(e, \beta) : e = e_0, \beta \in [0, \beta_*]\}$ of the boundary and one needs to have more precise information on the behavior of ω_1 close to $(\hat{e}, \hat{\beta}) = (e_0, \hat{\beta})$. If $\nabla\omega_1(e_0, \hat{\beta}) \sim (0, 0)$ (the following motivation is not rigorous since we don't know that ω_1 is differentiable on ∂D), then we would be in a similar situation as what has been described before. Therefore, we can assume that $\nabla\omega_1(e_0, \hat{\beta}) \not\sim (0, 0)$ in the sense that at least one of the derivatives $\frac{\partial\omega_1}{\partial e}$ and $\frac{\partial\omega_1}{\partial\beta}$ does not vanish at $(e_0, \hat{\beta})$. If it is exactly one of the two derivatives that does not vanish, one could also derive a bunch of results, with techniques that are similar to the ones outlined below. Hence, we are going to assume that both derivatives do not vanish, in a weak sense that does not need the differentiability, as formulated in the following condition:

(ω_1-3) There are a point $(e_0, \hat{\beta}) \in D$ and a constant $c_1 > 0$ such that $\omega_1(e_0, \hat{\beta}) = \delta_1$
    and
$$|\omega_1(e, \beta) - \delta_1| \geq c_1|(e, \beta) - (e_0, \hat{\beta})|, \quad (e, \beta) \in D;$$

it would be sufficient to require (ω_1-3) only locally in a neighborhood of $(e_0, \hat{\beta})$. Supposing that (ω_1-3) holds, we can show in Theorem 4.15 for $\mu_* = 1$ that $\lambda_* = \delta_1^2$ is an eigenvalue of L if and only if

$$\|\mu_1'\|_{L^\infty(]-\infty, \delta_1^2[)} < \infty \tag{1.32}$$

is verified; since $\mu_1(\cdot)$ is differentiable a.e., this condition is meaningful. The proof works by first observing that, as a consequence of (ω_1-3), the operator $Q_{\delta_1^2} = \lim_{\lambda \to \delta_1^2-} Q_\lambda$ does exist in the Hilbert-Schmidt norm (Lemma 4.9) and hence is a Hilbert-Schmidt operator itself. In addition, $\mu_* = 1$ is its first eigenvalue $\mu_1(\delta_1^2)$. Due to the compactness of $Q_{\delta_1^2}$, if $\Psi_j \in L_r^2$ is a normalized eigenfunction of Q_{λ_j} for $\mu_1(\lambda_j)$ and $\lambda_j \to \delta_1^2-$, then a subsequence will converge to a normalized eigenfunction Ψ_* of $Q_{\delta_1^2}$ for the eigenvalue $\mu_* = 1$ (Corollary 4.11, no need to assume (1.32)). Once again, the situation is very much analogous to what is known for Schrödinger operators, cf. [82, pp. 83–85] and [84, Section 2] for instance: a threshold eigenvalue and eigenfunction of the Birman-Schwinger operator do not immediately give rise to a threshold eigenvalue and eigenfunction of the Schrödinger operator, but in fact the existence of the latter is characterized by an additional condition, which is (1.32) in our case. To understand its meaning, suppose for simplicity of the presentation that there is $\varepsilon > 0$ such that $]\delta_1^2 - \varepsilon, \delta_1^2[\ni \lambda \mapsto \mu_1(\lambda)$ is real analytic, and in addition that there are $\Psi_\lambda \in L_r^2$ satisfying $\|\Psi_\lambda\|_{L_r^2} = 1$, $Q_\lambda\Psi_\lambda = \mu_1(\lambda)\Psi_\lambda$, so that also $]\delta_1^2 - \varepsilon, \delta_1^2[\ni \lambda \mapsto \Psi_\lambda$ is real analytic. This will follow from the Kato-Rellich perturbation theory if μ_* is known to be a simple eigenvalue of $Q_{\delta_1^2}$. In the general case, which is much more technical, one needs to work with appropriate sequences

$\lambda_j \to \delta_1^2-$ that are constructed using an appropriate generalization of the standard Kato-Rellich perturbation theory (Appendix IV). In the real analytic case, define $\psi_\lambda(r, p_r, \ell) = |Q'(e_Q)| \, p_r \Psi_\lambda(r)$ and $g_\lambda = (-\mathcal{T}^2 - \lambda)^{-1} \psi_\lambda$. Then it is found that

$$\|g_\lambda\|_{X^0}^2 = \frac{1}{4\pi} \langle \mathcal{Q}_\lambda' \Psi_\lambda, \Psi_\lambda \rangle = \frac{1}{4\pi} \mu_1'(\lambda)$$

and μ_1' is increasing. Thus, (1.32) is equivalent to the condition $\sup \|g_\lambda\|_{X^0} < \infty$, i.e., to the boundedness of $(g_\lambda) \subset X^0$. In addition, one can prove that

$$Lg_\lambda = (1 - \mu_1(\lambda))\psi_\lambda + \lambda g_\lambda,$$

cf. Lemma 4.7(c). Since $\mu_1(\lambda) \to \mu_1(\delta_1^2) = \mu_* = 1$, the weak convergence $g_\lambda \rightharpoonup g_*$ is seen to be sufficient to ensure that $g_* \neq 0$ and $Lg_* = \delta_1^2 g_*$, i.e., g_* is the wanted eigenfunction of L. To establish the converse assertion, i.e., that the existence of an eigenfunction of L for $\lambda_* = \delta_1^2$ leads to (1.32), a different argument has to be used; see Theorem 4.15. Corollary 4.17 contains an example of a situation where (1.32) can be shown to hold. For this, we add (Q5) from Sect. 1.3 as an additional condition on Q. It should not be surprising that the regularity of Q' close to $e = e_0$ will become important in this respect since we are dealing with integrals of the form

$$\sum_{k \neq 0} \iint_D d\beta \, de \, \frac{\omega_1(e, \beta) \, |Q'(e)|}{k^2 \omega_1^2(e, \beta) - \lambda} \, (\ldots)$$

many times. If $\lambda \sim \delta_1^2$ and $k = \pm 1$, then the behavior of ω_1 close to $(e, \beta) = (e_0, \hat{\beta})$ gets important; this is addressed by condition $(\omega_1\text{-}3)$. On the other hand, there is an interplay with the term $|Q'(e)|$ for e close to e_0, which could compensate for possible losses (or it could be bad itself). Generally speaking, many different results could be derived for $\mu_* = 1$ by combining assumptions of ω_1 with assumptions on Q' close to e_0.

Let us remark that we don't see an immediate path to calculate μ_* for a given steady state solution Q. However, there might be a smart way to settle this question, and in any case μ_*, together with additional important quantities like λ_* and δ_1, for sure could be determined numerically. Another notable fact is as follows. The Vlasov-Poisson system (1.5) and (1.3) has many invariances; see Chap. 6; quantities that remain invariant under the scaling could be expected to be of 'fundamental' importance. It turns out that μ_* is such a quantity, but λ_* and δ_1 are not. On the contrary, the conditions $\lambda_* < \delta_1^2$ and $\lambda_* = \delta_1^2$ are both invariant. We will deduce several other invariants in Chap. 6, among them the "Eddington-Ritter relation", which says that

$$\frac{2\pi}{\sqrt{\lambda_*}} \sqrt{\rho_Q(0)}$$

is invariant; note that $\frac{2\pi}{\sqrt{\lambda_*}}$ is the "linear period" from Lemma 1.3.

There are several other operators around that are used to assist (by means of their coercivity) stability proofs for stellar systems, among them the "Hartree-Fock exchange operator" by Lynden-Bell [57, 58] and the "Guo-Lin operator" [29, 88]. Concerning the latter, we are able to make a connection to the operators $\mathcal{Q}_\lambda$ that we are using, more precisely to $\mathcal{Q}_0$. Let $\lambda_{\mathrm{GL}} > 0$ denote the best constant for the Guo-Lin operator; see (5.2). Then we have

$$\lambda_{\mathrm{GL}} + \mu_1(0) = 1$$

by Lemma 5.1, and $0 < \mu_1(0) < 1$ implies that $\lambda_{\mathrm{GL}} > 0$ will always be attained (Corollary 5.2). Of course, the clear advantage of the operators $\mathcal{Q}_\lambda$ is the underlying Birman-Schwinger principle, as they can be used to detect the λ_* that will be the eigenvalue.

Finally, there are four appendices. Appendix I and Appendix II contain the necessary background material for what concerns the change of coordinates to action-angle variables, function spaces and operators. Appendix III is independent and provides a proof (using a new evolution equation) of the fact that $\lambda_* < \delta_1^2$ implies that λ_* is an eigenvalue of L; this will enter into the theorems obtained in Sect. 4.2. Lastly, Appendix IV concerns some specifics of the Kato-Rellich perturbation theory that are also used to study the properties of $\mathcal{Q}_\lambda$ as $\lambda \to \delta_1^2-$.