

# The Dirac point electron in zero-gravity Kerr–Newman spacetime

M. K.-H. KIESSLING AND A. S. TAHVILDAR-ZADEH

Department of Mathematics  
Rutgers, The State University of New Jersey  
110 Frelinghuysen Rd., Piscataway, NJ 08854

Orig. submission: Nov. 13, 2014; Revised version: March 21, 2015

## Abstract

Dirac’s wave equation for a point electron in the topologically nontrivial maximal analytically extended electromagnetic Kerr–Newman spacetime is studied in a limit  $G \rightarrow 0$ , where  $G$  is Newton’s constant of universal gravitation. The following results are obtained: the formal Dirac Hamiltonian on the static spacelike slices is essentially self-adjoint; the spectrum of the self-adjoint extension is symmetric about zero, featuring a continuum with a gap about zero that, under two smallness conditions, contains a point spectrum. The symmetry result extends to the Dirac operator on a generalization of the zero- $G$  Kerr–Newman spacetime with different electric-monopole / magnetic-dipole-moment ratio.

arXiv:1410.0419v3 [math-ph] 20 May 2015

# 1 Introduction

In 1976 Chandrasekhar [13, 12], Page [36], and Toop [47] showed that Dirac’s equation for a point electron in the Kerr–Newman spacetime [33] separates essentially completely<sup>1</sup> in oblate spheroidal coordinates. Although this remarkable discovery enabled detailed mathematical studies of the behavior of a Dirac electron in a charged, rotating black hole spacetime [27, 5, 17, 18, 19, 20, 3, 50, 51, 4] (see also [14] for neutral rotating black holes), there are perplexing conceptual issues which await clarification. Beside those that hark back to the enigmatic quantum-mechanical meaning of Dirac’s equation in Minkowski spacetime, see [45] and [29], serious new issues arise because of the physically somewhat questionable character of the Kerr–Newman solution, unveiled by Carter [10]; see also [34, 24].

Namely, the maximal analytical extension of the stationary axisymmetric Kerr–Newman spacetime has a very strong curvature singularity on a timelike<sup>2</sup> cylindrical surface whose cross-section with constant- $t$  hypersurfaces is a circle; here,  $t$  is a coordinate pertinent to the *asymptotically (at spacelike  $\infty$ ) timelike* Killing field that encodes the stationarity of the “outer regions” of the Kerr–Newman spacetime. This circle is commonly referred to as *the “ring” singularity*. The region near the ring is especially pathological since it harbors closed timelike loops.<sup>3</sup> Carter [10] also showed that the maximal analytically extended Kerr–Newman manifold is “cross-linked through the ring.”<sup>4</sup> Interestingly, this non-trivial topology was discovered a few years earlier in a family of static vacuum spacetimes by Zipoy [52], who completely described their maximal analytical extension. Since Zipoy seems to have been the first to discover this non-trivial topology in exact spacetime solutions to Einstein’s vacuum equations, we henceforth will refer to it as the *Zipoy topology*. Carter showed that the Zipoy topology survives the vanishing-charge limit of the Kerr–Newman manifold, which yields the maximal analytic extension [7] of Kerr’s solution [28] to Einstein’s vacuum equations ( $R_{\mu\nu} = 0$ ), cf. [23]. He furthermore showed that this topology also survives the vanishing-mass limit of the Kerr manifold, which yields an otherwise flat vacuum spacetime consisting of two static spacetime ends which are cross-linked through the ring. This vanishing-mass limit of the Kerr manifold coincides with the vanishing-mass limit of Zipoy’s oblate spheroidal family of static vacuum spacetimes<sup>5</sup>

In the black-hole sector of their parameter space the Kerr–Newman spacetimes also have a Cauchy horizon, an event horizon, and an ergosphere horizon; see [23, 34, 24]. From the “safe perspective of an observer at spatial infinity” the ring singularity, the acausal region, and the Cauchy horizon are invisible, being “hidden” behind the event horizon, and no exotic or even objectionable physics would ever seem to happen: a Dirac spinor wavefunction initially supported outside the event horizon will either keep spreading within the outer region or eventually (as  $t \rightarrow \infty$ ) accumulate (in parts or wholly) at the event horizon, see [19, 20]. Yet, an inquisitive physicist may also want to study the spinor wave function in other coordinates designed to “follow it across

---

<sup>1</sup>In contrast to the familiar separation-of-variables results for, say, the Laplacian in a rectangular box or a cylinder or a sphere, Chandrasekhar, Page, and Toop obtained a system of ODEs for functions of only one variable each which is *not* of triangular structure, and so cannot be solved one equation at a time.

<sup>2</sup>In a limiting sense of course, since the metric is singular on this surface.

<sup>3</sup>The timelike ring singularity of the Kerr–Newman manifold is itself the limit of closed timelike loops, for which reason it is not possible to interpret this singular source of the stationary and axisymmetric Kerr–Newman electromagnetic fields outside of the outer ergosphere horizon as a “rotating charged ring.” For a careful analysis of the ring sources of the electromagnetic  $z$  GKN fields, see [43].

<sup>4</sup>The complement of a wedding ring in ordinary three-dimensional Euclidean space is topologically non-trivial, too, but “looping through the ring once brings you back to where you began;” in a spacelike slice of the maximal analytically extended Kerr–Newman spacetime “you need to loop through the ring twice to get back to square one.”

<sup>5</sup>In the same paper Zipoy also described another, prolate spheroidal family of static vacuum spacetimes, whose metric is nowadays known as “Zipoy–Voorhees metric.”

the event horizon,”<sup>6</sup> however, it is neither clear how to continue in a “physically correct” manner beyond the Cauchy horizon<sup>7</sup> nor what to make of the acausal region of closed timelike loops, nor how to correctly handle the timelike singularity. Moreover, if one inquires into the “physics beyond the event horizon,” one has the option of allowing the support of the initial spinor wave function to be spread over both asymptotically flat ends, or some other parts of the maximal analytically extended Kerr–Newman spacetime. It is not so clear which options (if any) are physically reasonable and which ones are science fiction, although astrophysicists can argue for the “physical black hole spacetime,” i.e. the part of the maximal analytically extended Kerr–Newman spacetime which is the asymptotic limit of the topologically simple spacetime of a charged, rotating star collapsing into a black hole [41].

The horizons are absent in the hyper-extremal parameter regime. Even though the absence of the Cauchy horizon is a welcome simplification, this regime is rarely studied because the absence of the event horizon renders the singularity “naked,” and the (weak) cosmic censorship hypothesis, according to which “nature abhors naked singularities” [37, 38], has (unfortunately) discouraged physicists from investigating spacetimes with naked singularities. Yet once a piece of a Dirac spinor wave function has crossed the event horizon of a Kerr–Newman black hole it is no longer shielded from possible harm done by the spacetime singularity — viz., inside the event horizon the singularity *is* naked —, and so one may as well study the effects of naked singularities directly. Be that as it may, the hyper-extremal regime retains the closed timelike loops which according to the standard interpretation of general relativity turn the entire manifold into a *causally vicious set*, something that many physicists (including the authors) would regard as physically suspicious.

A strategy to rid the Kerr–Newman manifold from its Cauchy horizon, and all its other acausal aspects, is to take a *zero-gravity limit*  $G \rightarrow 0$ , where  $G$  is Newton’s constant of universal gravitation. This would be quite uninteresting if the zero- $G$  limit of the Kerr–Newman manifold would simply yield a Minkowski spacetime decorated with the electric field of a point charge and the magnetic field of a point dipole, as one might be tempted to guess from the asymptotically flat ends of the Kerr–Newman spacetime. However, as shown by one of us in the accompanying paper [43], the zero-gravity limit of the maximal analytically extended Kerr–Newman spacetime yields a static, flat, yet *two-leafed*, *cross-linked* spacetime which is decorated with Appell–Sommerfeld [2, 40] electromagnetic fields<sup>8</sup>  $\mathbf{F} = d\mathbf{A}$  with  $\mathbf{A}$  as in (5), whose sources are certain “finite charge and current distributions” supported by the one-dimensional ring.<sup>9</sup> Although the gravitational (viz.: curvature) aspects of the Kerr–Newman manifold, its event horizon included, vanish in this limit too, one does retain the topological, the singular, and all the electromagnetic aspects of the spacetime. Studies of the Dirac equation for a point electron in this zero- $G$  Kerr–Newman (*zGKN*) spacetime will therefore illuminate the role played by the topological and electromagnetic aspects of the Kerr–

---

<sup>6</sup>When the analogous study was carried out by Oppenheimer and Snyder [35] for classical gaseous matter undergoing gravitational collapse it leveled the ground for building our modern understanding of the physics of gravitational collapse, involving the formation of black holes and their singularities. Poetically speaking their work revealed that there is more physics in general relativity than meets the (distant observer’s) eye.

<sup>7</sup>If instead of the Cauchy problem one studies  $t$ -periodic solutions, then one can continue across the Cauchy and the event horizons using a weak matching procedure [17, 18]. However, Finster et al. [17, 18] found that no  $t$ -periodic solutions exist which are normalized over a constant- $t$  slice of the “physical black hole spacetime;” see main text.

<sup>8</sup>The zero- $G$  limit of the electromagnetic Kerr–Newman fields yields fields originally discovered by Appell [2], who obtained them from the Coulomb potential of a point charge by a complex translation of the charge’s position. Appell noticed that the fields change sign when looping once through the ring, but did not conclude — apparently — that they live naturally on a topologically nontrivial space. Sommerfeld [40] seems to have been first to introduce “branched Riemann spaces,” three-dimensional analogues of topologically non-trivial Riemann surfaces, to which we will refer as *Sommerfeld spaces*, and to construct electromagnetic fields (harmonic functions) on them, which in general we will call *Sommerfeld fields*. Eventually Evans [16] and his students [32], [1] laid their rigorous foundations.

<sup>9</sup>Strictly speaking, the ring singularity is not part of the manifold; it’s rather a ring “defect.”

Newman manifold in the relativistic quantum mechanics of the electron.

In this paper we study the Dirac equation for a point electron in static, electromagnetic, flat spacetimes with Zipoy topology which include the zGKN spacetimes as special case, but which in general can sport Sommerfeld fields  $\mathbf{F} = d\mathbf{A}$  with  $\mathbf{A}$  as in (6), which differ from the Appell–Sommerfeld fields only in a single number, the ratio  $I\pi a^2/Qa$  of their magnetic dipole moment to the magnetic dipole moment of the Appell–Sommerfeld fields of same charge  $Q$ ; here,  $|a|$  is the radius of the ring singularity. By constructing an operator that anti-commutes with the pertinent Dirac Hamiltonian we show that the spectrum of any of its self-adjoint extensions is symmetric about zero; this result holds for arbitrary  $(Q, I)$ . All other results are obtained for Dirac’s electron in the zGKN spacetime ( $I\pi a^2/Qa = 1$ ): by adapting an argument of Winklmeier–Yamada for the Dirac equation of a point electron in the outer region of the Kerr–Newman black hole spacetime, we show that our formal Dirac Hamiltonian is essentially self-adjoint on a spacelike slice of the maximal analytically extended, static zGKN spacetime. Then we exploit the Chandrasekhar–Page–Toop separation-of-variables method for Dirac’s equation on a general Kerr–Newman spacetime, and the Prüfer transform, to show that the self-adjoint Dirac operator has a continuous spectrum with a gap about zero which, under two smallness conditions, contains a pure point spectrum associated with time-periodic  $L^2$  spinor fields, representing bound states of Dirac’s point electron in the electromagnetic field of the ring singularity of the zGKN spacetime.

In the next section we formulate our main results about the Dirac equation for a point electron in the zGKN spacetime; one result is valid also for a Dirac electron in static, flat spacetimes having Zipoy topology featuring electromagnetic Sommerfeld fields of arbitrary  $I\pi a/Q$ -ratio. In sections 3, 4, 5, 6, and 7 we prove our main theorems about the spectrum. In section 8 we conclude with a list of interesting questions left unanswered by this work.

## 2 Formulation of the main results

We begin by formulating the Dirac equation for a point electron in electromagnetic, static, flat spacetimes with Zipoy topology which generalize zero- $G$  Kerr–Newman spacetimes to zero- $G$  Kerr spacetimes equipped with Sommerfeld fields of arbitrary  $I\pi a/Q$ -ratio. We then state our main theorems about the spectrum of the pertinent Dirac operators.

### 2.1 Dirac’s equation for a point electron on zero- $G$ Kerr spacetimes equipped with electromagnetic Sommerfeld fields of arbitrary $I\pi a/Q$ -ratio

#### 2.1.1 Zero- $G$ Kerr spacetimes

Our limit  $G \rightarrow 0$  of the maximal analytic extension of the well-known *Kerr* family of stationary, axisymmetric spacetime solutions of Einstein’s vacuum equations yields a one-parameter family of static, flat, but topologically nontrivial spacetimes<sup>10</sup>  $(\mathcal{M}, \mathbf{g})$  which consist of two “cross-linked leaves.” Explicitly, let  $\mathcal{C} \equiv \{(t, r, \theta, \varphi) : t \in \mathbb{R}, r \in \mathbb{R}, \theta \in [0, \pi], \varphi \in [0, 2\pi)\}$  denote a rectangular “four-dimensional cylinder,” and let  $\mathcal{S} \equiv \{(t, r, \theta, \varphi) : t \in \mathbb{R}, r = 0, \theta = \pi/2, \varphi \in [0, 2\pi)\} \subset \mathcal{C}$  denote a rectangular “two-dimensional slab” in  $\mathcal{C}$ . Then  $\mathcal{C} \setminus \mathcal{S}$  is a covering chart of oblate spheroidal

---

<sup>10</sup> We emphasize that Zipoy [52] found a large class of static, axisymmetric, flat but topologically nontrivial solutions to the Einstein vacuum equations which in their zero- $G$  limit coincide with the members of the zero- $G$  Kerr spacetime family.

(Boyer–Lindquist, or BL) coordinates<sup>11</sup> for this spacetime, with line element

$$ds_{\mathbf{g}}^2 = dt^2 - (r^2 + a^2) \sin^2 \theta d\varphi^2 - \frac{r^2 + a^2 \cos^2 \theta}{r^2 + a^2} (dr^2 + (r^2 + a^2) d\theta^2); \quad (1)$$

here,  $a^2 > 0$  is the only parameter of these spacetimes, and we have set the speed of light  $c = 1$ . Our sign convention of  $(+, -, -, -)$  for the metric follows [18].

The static, axisymmetric character of these zero- $G$  Kerr spacetimes is manifest in (1). Also, since  $r \in \mathbb{R}$  occurs strictly quadratically in (1), it is clear that the manifold consists of two “conjoined identical twins.” To exhibit their flatness, and in the process also the topological nontrivial juncture, we introduce cylindrical coordinates  $(t, \varrho, z, \varphi)$  on Minkowski spacetime  $\mathbb{R}^{1,3}$ , with the same  $(t, \varphi)$  as in Boyer–Lindquist coordinates, and with the cylindrical coordinates  $(\varrho, z)$  related to the elliptical coordinates  $(r, \theta)$  by

$$\varrho = \sqrt{r^2 + a^2} \sin \theta, \quad z = r \cos \theta. \quad (2)$$

In cylindrical  $(t, \varrho, z, \varphi)$  coordinates the metric takes the familiar form for flat Minkowski spacetime

$$ds_{\mathbf{g}}^2 = dt^2 - d\varrho^2 - \varrho^2 d\varphi^2 - dz^2, \quad (3)$$

except that the map (2) makes it plain that the chart  $\mathcal{C} \setminus \mathcal{S}$  will be mapped into *two* copies of Minkowski spacetime which are “doubly conjoined,” in a smooth yet crossing manner, at the set spanned by  $\mathcal{S}$ . The metric  $\mathbf{g}$  given by the line element (1) has a singularity at  $\mathcal{S}$ , which is the singularity *of the spacetime*, not *in the spacetime*. The set  $\mathcal{S}$  is the boundary of a timelike open solid cylinder in the zGKN manifold; the cross section at any instant  $t$  of this cylindrical surface is a translate of the *ring*  $\mathcal{R}_0 \equiv \{(t, r, \theta, \varphi) : t = 0, r = 0, \theta = \pi/2, \varphi \in [0, 2\pi)\}$  of Euclidean radius  $\sqrt{a^2} = |a| > 0$ , for which reason one speaks of a “ring singularity;” we write  $|a|$  because we need to allow  $a \in \mathbb{R} \setminus \{0\}$  for reasons that become clear in the next subsection. The points on the ring are *conical singularities* for the metric, meaning that the limit as the radius goes to zero of the ratio of the circumference to radius of a small circle centered at a point of the ring and lying in a meridional plane  $\varphi = \text{const.}$  is not  $2\pi$ ; instead, here it is  $4\pi$ . See [43] for details.

The key topological features of this manifold can easily be visualized. Namely, although the fixed timelike planes  $\{(t, r, \theta, \varphi) : t = t_0, \varphi = \varphi_0\}$  cannot be embedded into  $\mathbb{R}^3$ , each such plane can be immersed in it, the immersion consisting of two Euclidean half planes, stacked up upon each other, then cut along a line segment of length  $|a|$  orthogonal to the planes’ boundaries “with scissors,” then smoothly “cross-glued” at the cut such that the “upper” and “lower” sheets are cross-linked like an  $\times$  along the cut, while remaining like  $\parallel$  beyond the cut; the singular endpoint of the  $\times$ -line is not part of the two-sheeted manifold. Shown in Fig.1 are the ring singularity and the part  $\{r \in (-1, 1), \theta \in (0, \pi)\}$  of a constant-azimuth section (slightly curved, for the purpose of visualization) of the two-sheeted static spacelike slice of the zero- $G$  Kerr spacetime. The coordinate grid shows a few of the curves  $\theta = \text{const.}$  (hyperbolas, transiting through the ring from one sheet to the other) and  $r = \text{const.}$  (oblate semi-ellipses, remaining outside the ring on a single sheet).

We can view  $\mathcal{M}$  as a *branched covering* over the base manifold  $\mathbb{R}^{1,3} \setminus \mathcal{S}$  (mildly abusing notation), with the projection map  $\Pi : \mathcal{M} \rightarrow \mathbb{R}^{1,3} \setminus \mathcal{S}$  being  $\Pi(t, r, \theta, \varphi) = (t, \varrho, z, \varphi)$ . The pre-image of a point in the base consists of two points, degenerating into one point at each “ring”

$$\mathcal{R}_{t_0} = \{(t, r, \theta, \varphi) \mid t = t_0, r = 0, \theta = \pi/2, 0 \leq \varphi \leq 2\pi\}. \quad (4)$$

---

<sup>11</sup>The notation  $(t, r, \theta, \varphi)$  for the BL coordinates is standard in the relativity literature, and should not be confused for instance with the Schwarzschild coordinates on the outer region of that spacetime, or with just standard spherical coordinates of Minkowski spacetime. All standard non-flat  $(t, r, \theta, \varphi)$  coordinate systems reduce to standard spherical coordinates of the flat Minkowski spacetime near “ $r = \infty$ .” Note that in BL coordinates  $r$  takes any real value.

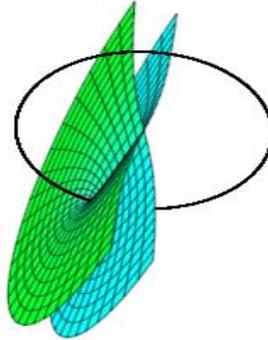


Figure 1: An illustration of the Zipoy topology.

The pullback of the Minkowski metric  $\eta$  under  $\Pi$  endows  $\mathcal{M}$  with a flat Lorentzian metric  $\mathbf{g} = \Pi^*\eta$ , whose line element is given in (1).

### 2.1.2 Zero-G Kerr-Newman spacetimes

The spacetime  $(\mathcal{M}, \mathbf{g})$  introduced above can be decorated with any static electromagnetic Sommerfeld field  $\mathbf{F} = d\mathbf{A}$ , satisfying the flat space Maxwell equations locally but respecting the topologically nontrivial character of the spacetime. The zero- $G$  limit of the maximal analytical extension of the *Kerr–Newman* family<sup>12</sup> of stationary axisymmetric solutions to the Einstein–Maxwell equations, written in BL coordinates, yields precisely the zero- $G$  Kerr (zGK) spacetime decorated with a particular electromagnetic Sommerfeld field, the *Appell–Sommerfeld field*, whose four-potential one-form reads

$$\mathbf{A} = -\frac{r}{r^2 + a^2 \cos^2 \theta} (Qdt - Qa \sin^2 \theta d\varphi). \quad (5)$$

Here,  $Q$  is the total charge “seen from infinity” in the  $r > 0$  sheet, defined by computing the electric outward flux through a spherical surface surrounding the ring singularity in that sheet, and we need to allow  $a \in \mathbb{R} \setminus \{0\}$  to accommodate a magnetic dipole moment vector pointing either “up” or “down” w.r.t. the  $z$  axis defined in (2).

The field  $\mathbf{F}$  is singular on the same ring  $\mathcal{R}_t$  as is the metric, while for  $r$  very large positive its electric and magnetic components approach, respectively, the asymptotics of an “electric monopole field in  $\mathbb{R}^3$  of a charge  $Q$ ” and a “magnetic dipole field in  $\mathbb{R}^3$  of dipole moment  $Qa$ ,” aligned parallel or anti-parallel to the “ $z$  direction,” for  $-r$  very large positive, i.e. in the other sheet, these electric monopole and magnetic dipole fields correspond to a charge  $-Q$  and magnetic dipole moment  $-Qa$ .

**Remark 2.1.** *Note that in contrast to the metric (1), the electromagnetic potential (5) is not invariant under a change of sign of  $a \in \mathbb{R} \setminus \{0\}$ ; however, the difference is merely in the direction of the magnetic moment vector that corresponds to the above potential, which either points along or opposite to the “ $z$  direction” defined by (2). Since a 180-degree rotation around a diametrical axis through the ring transforms its magnetic moment into the negative thereof, and since all objectively physical quantities are invariant under such a rotation, the choices  $a > 0$  and  $a < 0$  are physically*

<sup>12</sup>We recall that for fixed  $G > 0$  (and speed of light  $c$ ) the Kerr–Newman family is a three-parameter family of electrovac spacetimes, the parameters being ADM mass (energy)  $M(> 0)$ , ADM angular momentum  $J = Ma \in \mathbb{R}$ , and total charge  $Q \in \mathbb{R}$ , all defined in a single asymptotic end. Note that in units where  $c = 1$  the angular momentum per unit mass  $a$  has physical dimension of length; equivalently, of time.

equivalent, given  $Q$  and  $a^2 > 0$ . Therefore, without loss of generality, we can choose  $a > 0$  in all ensuing calculations involving (5).

**Remark 2.2.** *It may be tempting to speculate whether the magnetic dipole moment  $Qa$  can be interpreted as due to a “gyrating charged ring,” with “ $a$ ” the angular momentum per unit mass “of the singularity,” as has been attempted for Kerr–Newman spacetimes. Moreover, since Kerr–Newman spacetimes have a gyromagnetic ratio  $Q/M =: g_{\text{KN}}Q/2M$  (in units with  $c = 1$ ), amounting to a  $g$ -factor  $g_{\text{KN}} = 2$  (see [10]), and since the KN parameters  $(M, Q, a)$  are independent of  $G$ , and so is the KN gyromagnetic ratio, one could be tempted to assign the  $z\text{GKN}$  spacetime the same gyromagnetic ratio of  $Q/M$  and  $g$ -factor of 2. However, since  $M$  does not show in the  $z\text{GKN}$  metric, such an assignment would be reasonable only if there were no other way to construct  $z\text{GKN}$  than taking the zero- $G$  limit of KN. Yet this is not the case: as already pointed out in footnote 10, the underlying spacetime manifold of  $z\text{GKN}$  can be obtained as zero- $G$  limit of either, the stationary family of Kerr spacetimes — having both an ADM mass  $M_K$  and ADM angular momentum  $J = M_K a$  —, or a static family of Zipoy spacetimes — having an ADM mass  $M_z$  but zero ADM angular momentum; note also that  $M_z \neq M_K$  in general. This (say)  $z\text{GZ}$  spacetime can now be equipped with an arbitrary Sommerfeld field, in particular: the Appell–Sommerfeld field of  $z\text{GKN}$ , without being logical compelled to interpret its magnetic moment  $Qa$  as being due to a “gyrating ring of charge  $Q$ ” with angular momentum per unit mass “ $a$ ,” although this is logically possible; in any event, it’s better to refrain from assigning  $z\text{GKN}$  any spacetime  $g$ -factor.*

### 2.1.3 Generalizations of $z\text{GKN}$ spacetimes to arbitrary charge and current.

Since the electric and magnetic components of Maxwell’s vacuum equations decouple in the zero- $G$  limit, to decorate the zero- $G$  Kerr spacetime with a generalization of the electromagnetic Appell–Sommerfeld field  $\mathbf{F} = d\mathbf{A}$  having electric charge  $Q \in \mathbb{R}$  and current  $I \in \mathbb{R}$ , all that needs to be done is to replace the magnetic dipole moment  $Qa$  in formula (5) by  $I\pi a^2$ , thus

$$\mathbf{A} = -\frac{r}{r^2 + a^2 \cos^2 \theta} (Qdt - I\pi a^2 \sin^2 \theta d\varphi). \quad (6)$$

Again, electric charge  $Q$  and magnetic dipole moment  $I\pi a^2$  are as “seen” from spacelike infinity in the  $r > 0$  sheet; viewing from spacelike infinity in the other sheet one “sees”  $-Q$  and  $-I\pi a^2$ .

**Remark 2.3.** *The Sommerfeld field (6) makes it plain that the  $z\text{GKN}$  spacetimes are but a special one-parameter subfamily in a two-parameter family of qualitatively similar electromagnetic spacetimes with arbitrary charge  $Q$  and magnetic moment  $I\pi a^2$  (given  $a^2$ ). The ease with which this result was accomplished stands in stark contrast to the difficulties in generalizing the Kerr–Newman family to electromagnetic spacetimes with a magnetic moment different from  $Qa$ .*

**Remark 2.4.** *As does the metric (1), the generalized Sommerfeld field (6) depends on “ $a$ ” only through  $a^2$ , hence it is invariant under a sign change of  $a \in \mathbb{R} \setminus \{0\}$ . So when (1) is combined with (6) we can choose  $a > 0$  without any further ado. Of course, in (6) there is now the new parameter  $I \in \mathbb{R}$ , but by the same rotation argument as given for the  $z\text{GKN}$  fields the choices  $I > 0$  and  $I < 0$  are physically equivalent, given  $Q$  and  $a^2 > 0$ . Therefore, without loss of generality, from here on we choose  $\text{sign}(I) = \text{sign}(Q)$ , and  $a > 0$ .*

### 2.1.4 The Dirac equation on electromagnetic spacetimes: Cartan’s frame method

In arbitrary coordinates  $(x^\mu)$  (with  $c = 1$  and  $\hbar = 1$ ), the Dirac equation for a spin-1/2 electron of empirical rest mass  $m$  and charge  $-e < 0$  interacting (through minimal coupling) with an

electromagnetic field  $\mathbf{F} = d\mathbf{A}$  in a spacetime  $(\mathcal{M}, \mathbf{g})$  reads

$$\tilde{\gamma}^\mu (-i\nabla_\mu + eA_\mu)\Psi + m\Psi = 0; \quad (7)$$

here  $\nabla$  is the covariant derivative (on bi-spinors) associated to the spacetime metric  $\mathbf{g}$ , and the  $(\tilde{\gamma}^\mu)_{\mu=0}^3$  are Dirac matrices associated to this metric, i.e. satisfying

$$\tilde{\gamma}^\mu \tilde{\gamma}^\nu + \tilde{\gamma}^\nu \tilde{\gamma}^\mu = 2g^{\mu\nu} \mathbf{1}_{4 \times 4}, \quad (8)$$

while the  $A_\mu$  are the pertinent components of the electromagnetic potential,  $\mathbf{A} = A_\mu dx^\mu$ .

Using Cartan's frame method (see [8] and refs. therein) one can express the above covariant derivative on spinors in terms of standard derivatives:

$$\tilde{\gamma}^\mu \nabla_\mu = \gamma^\mu \mathbf{e}_\mu + \frac{1}{4} \Omega_{\mu\nu\lambda} \gamma^\lambda \gamma^\mu \gamma^\nu, \quad (9)$$

where the  $\gamma^\nu$  are Dirac gamma matrices for the Minkowski spacetime, satisfying

$$\gamma^\nu \gamma^\mu + \gamma^\mu \gamma^\nu = 2\eta^{\mu\nu} \mathbf{1}_{4 \times 4}, \quad (10)$$

with

$$(\boldsymbol{\eta}) = \text{diag}(1, -1, -1, -1) \quad (11)$$

being the matrix of the Minkowski metric in rectangular coordinates; and  $\{\mathbf{e}_\mu\}_{\mu=0}^3$  is a *Cartan frame*, i.e. an orthonormal frame of vectors spanning the tangent space at each point of the spacetime manifold. We thus have

$$(\mathbf{e}_\mu)^\nu (\mathbf{e}_\lambda)^\kappa g_{\nu\kappa} = \eta_{\mu\lambda}. \quad (12)$$

On the one hand, it follows that

$$\tilde{\gamma}^\mu = (\mathbf{e}_\nu)^\mu \gamma^\nu. \quad (13)$$

On the other hand, let  $\{\boldsymbol{\omega}^\mu\}_{\mu=0}^3$  denote the *dual* frame to  $\{\mathbf{e}_\mu\}$ , i.e. the orthonormal basis for the cotangent space at each point of the manifold that is dual to the basis for the tangent space:

$$\boldsymbol{\omega}_\mu(\mathbf{e}^\nu) = \mathbf{e}^\nu(\boldsymbol{\omega}_\mu) = \delta_\mu^\nu. \quad (14)$$

Then the  $\Omega_{\mu\nu\lambda}$  are by definition the *Ricci rotation coefficients* of the frame  $\{\boldsymbol{\omega}^\mu\}_{\mu=0}^3$ , defined in the following way: Let the one-forms  $\boldsymbol{\Omega}_\nu^\mu$  satisfy

$$d\boldsymbol{\omega}^\mu + \boldsymbol{\Omega}_\nu^\mu \wedge \boldsymbol{\omega}^\nu = 0. \quad (15)$$

This does not uniquely define the  $\boldsymbol{\Omega}_\nu^\mu$ . However, there exists a unique set of such 1-forms satisfying the extra condition

$$\boldsymbol{\Omega}_{\mu\nu} = -\boldsymbol{\Omega}_{\nu\mu}, \quad (16)$$

where the first index is lowered by the Minkowski metric:  $\boldsymbol{\Omega}_{\mu\nu} := \eta_{\mu\lambda} \boldsymbol{\Omega}_\nu^\lambda$ . Since  $\{\boldsymbol{\omega}^\mu\}$  forms a basis for the space of 1-forms, we then have  $\boldsymbol{\Omega}_{\mu\nu} = \Omega_{\mu\nu\lambda} \boldsymbol{\omega}^\lambda$ , which defines the rotation coefficients  $\Omega_{\mu\nu\lambda}$ .

The Dirac equation (7) on a spacetime  $(\mathcal{M}, \mathbf{g})$  with an electromagnetic 4-potential  $\mathbf{A}$  can thus be written in the following form:

$$\gamma^\mu \left( \mathbf{e}_\mu + \Gamma_\mu + ie\tilde{A}_\mu \right) \Psi + im\Psi = 0; \quad (17)$$

here, the  $\Gamma_\mu$  are connection coefficients,

$$\Gamma_\mu := \frac{1}{4} \Omega_{\nu\lambda\mu} \gamma^\nu \gamma^\lambda = \frac{1}{8} \Omega_{\nu\lambda\mu} [\gamma^\nu, \gamma^\lambda], \quad (18)$$

and the  $\tilde{A}_\mu$  are the components of the potential  $\mathbf{A}$  in the  $\boldsymbol{\omega}^\mu$  basis, i.e.  $\mathbf{A} = \tilde{A}_\mu \boldsymbol{\omega}^\mu$ , or,

$$\tilde{A}_\mu := (\mathbf{e}_\mu)^\nu A_\nu. \quad (19)$$

### 2.1.5 Frame formulation of the Dirac equation on zGK spacetimes featuring generalized electromagnetic Sommerfeld fields with arbitrary $I\pi a/Q$ -ratio

As explained in section 2.1.1, the single chart  $\mathcal{C} \setminus \mathcal{S}$  of oblate spheroidal coordinates  $(t, r, \theta, \varphi)$  covers the whole zero- $G$  Kerr spacetime  $(\mathcal{M}, \mathbf{g})$ , and in section 2.1.2 we saw that in these coordinates the generalized electromagnetic Sommerfeld one-form  $\mathbf{A}$  is everywhere on  $(\mathcal{M}, \mathbf{g})$  given by the simple formula (6). It is therefore natural that one would like to write Dirac's equation (7) in these coordinates as well, in the hope of achieving at least some partial separation of variables.<sup>13</sup>

However, unlike Cartesian coordinates  $(x^\mu)$  in Minkowski spacetime, oblate spheroidal coordinate derivatives do not give rise to an orthonormal basis for the tangent space at each point of a zero- $G$  Kerr spacetime. Thus, to bring (7) into the Cartan form (17) using oblate spheroidal coordinates, one also needs to construct a suitable Cartan frame. Following Chandrasekhar [13, 12], Page [36], Toop [47] (see also Carter-McLenaghan [11]), we introduce a special orthonormal frame  $\{\mathbf{e}_\mu\}_{\mu=0}^3$  on the tangent bundle  $T\mathcal{M}$  which is adapted to the oblate spheroidal coordinates, such that the Dirac equation takes a comparatively simple form.

We begin by introducing a Cartan (co-)frame  $\{\omega^\mu\}_{\mu=0}^3$  for the cotangent bundle<sup>14</sup>

$$\omega^0 := \frac{\varpi}{|\rho|}(dt - a \sin^2 \theta d\varphi), \quad \omega^1 := |\rho|d\theta, \quad \omega^2 := \frac{\sin \theta}{|\rho|}(-adt + \varpi^2 d\varphi), \quad \omega^3 := \frac{|\rho|}{\varpi}dr, \quad (20)$$

with the abbreviations

$$\varpi := \sqrt{r^2 + a^2}, \quad \rho := r + ia \cos \theta. \quad (21)$$

Let us denote the oblate spheroidal coordinates  $(t, r, \theta, \varphi)$  collectively by  $(y^\nu)$ . Let  $g_{\mu\nu}$  denote the coefficients of the spacetime metric (1) in oblate spheroidal coordinates, i.e.  $g_{\mu\nu} = \mathbf{g}\left(\frac{\partial}{\partial y^\mu}, \frac{\partial}{\partial y^\nu}\right)$ . One easily checks that written in the  $\{\omega^\mu\}$  frame, the spacetime line element is

$$ds_{\mathbf{g}}^2 = g_{\mu\nu} dy^\mu dy^\nu = \eta_{\alpha\beta} \omega^\alpha \omega^\beta. \quad (22)$$

This shows that the frame  $\{\omega^\mu\}_{\mu=0}^3$  is indeed orthonormal. With respect to this frame the electromagnetic Sommerfeld potential (6) becomes  $\mathbf{A} = \tilde{A}_\mu \omega^\mu$ , with

$$\tilde{A}_0 = -Q \frac{r}{|\rho|\varpi} - (Q - I\pi a) \frac{a^2 r \sin^2 \theta}{\varpi |\rho|^3}, \quad \tilde{A}_1 = 0, \quad \tilde{A}_2 = -(Q - I\pi a) \frac{ar \sin \theta}{|\rho|^3}, \quad \tilde{A}_3 = 0. \quad (23)$$

**Remark 2.5.** *We observe that for  $Q = I\pi a$ , all but one of the quantities  $\tilde{A}_\mu$  vanish, and the non-vanishing one,  $\tilde{A}_0$ , reduces to  $-Qr/|\rho|\varpi$ .*

**Remark 2.6.** *Clearly, the Cartan co-frame (20) is not invariant under the replacement  $a \rightarrow -a$ ; it is adapted to (5). If we replace  $a \rightarrow -a$  in (5), we also need to replace  $a \rightarrow -a$  in (20); in the same vein, replacing  $a \rightarrow -a$  in (20) corresponds to replacing  $I \rightarrow -I$  in (6). Therefore, in the following, a change  $a \rightarrow -a$  needs to be simultaneously accompanied by the change  $I \rightarrow -I$ .*

Next, let the frame of vector fields  $\{\mathbf{e}_\mu\}$  be the *dual* frame to  $\{\omega^\mu\}$ . Thus  $\{\mathbf{e}_\mu\}$  yields an orthonormal basis for the tangent space at each point in the manifold:

$$\mathbf{e}_0 = \frac{\varpi}{|\rho|}\partial_t + \frac{a}{\varpi|\rho|}\partial_\varphi, \quad \mathbf{e}_1 = \frac{1}{|\rho|}\partial_\theta, \quad \mathbf{e}_2 = \frac{a \sin \theta}{|\rho|}\partial_t + \frac{1}{|\rho| \sin \theta}\partial_\varphi, \quad \mathbf{e}_3 = \frac{\varpi}{|\rho|}\partial_r. \quad (24)$$

<sup>13</sup>The idea of using special frames adapted to a coordinate system in order to separate spinorial wave equations in those coordinates goes back to Kinnersley [30] and Teukolsky [44].

<sup>14</sup>This particular frame is called a *canonical symmetric tetrad* in [11].

Next, the anti-symmetric matrix  $(\Omega_{\mu\nu}) = (\eta_{\mu\lambda}\Omega_\nu^\lambda)$  is computed to be

$$(\Omega_{\mu\nu}) = \begin{pmatrix} 0 & -C\omega^0 - D\omega^2 & D\omega^1 - B\omega^3 & -A\omega^0 - B\omega^2 \\ & 0 & D\omega^0 + F\omega^2 & -E\omega^1 - C\omega^3 \\ & (\text{anti-sym}) & 0 & -B\omega^0 - E\omega^2 \\ & & & 0 \end{pmatrix}, \quad (25)$$

with

$$A := \frac{a^2 r \sin^2 \theta}{\varpi |\rho|^3}, \quad B := \frac{ar \sin \theta}{|\rho|^3}, \quad C := \frac{a^2 \sin \theta \cos \theta}{|\rho|^3}, \quad D := \frac{a \cos \theta \varpi}{|\rho|^3}, \quad E := \frac{r \varpi}{|\rho|^3}, \quad F := \frac{\varpi^2 \cos \theta}{|\rho|^3 \sin \theta}. \quad (26)$$

With respect to this frame on a zero- $G$  Kerr spacetime, and picking the *Weyl* representation<sup>15</sup> for the Dirac matrices  $\gamma^\mu$ , the covariant derivative part of the Dirac operator (7) can be expressed with the help of the operator

$$\mathfrak{D} := \tilde{\gamma}^\mu \nabla_\mu = \begin{pmatrix} 0 & \mathfrak{l}' + \mathfrak{m}' \\ \mathfrak{l} + \mathfrak{m} & 0 \end{pmatrix}, \quad (27)$$

where

$$\mathfrak{l} := \frac{1}{|\rho|} \begin{pmatrix} D_+ & L_- \\ L_+ & D_- \end{pmatrix} \quad (28)$$

and

$$\mathfrak{l}' := \frac{1}{|\rho|} \begin{pmatrix} D_- & -L_- \\ -L_+ & D_+ \end{pmatrix}, \quad (29)$$

with

$$D_\pm := \pm \varpi \partial_r + \left( \varpi \partial_t + \frac{a}{\varpi} \partial_\varphi \right), \quad L_\pm := \partial_\theta \pm i (a \sin \theta \partial_t + \csc \theta \partial_\varphi), \quad (30)$$

while

$$\begin{aligned} \mathfrak{m} &:= \frac{1}{2} [(-2C + F + iB)\sigma_1 + (-A + 2E + iD)\sigma_3] \\ &= \frac{1}{2|\rho|} \begin{pmatrix} \frac{r}{\varpi} + \frac{\varpi}{\rho} & \cot \theta + \frac{ia \sin \theta}{\rho} \\ \cot \theta + \frac{ia \sin \theta}{\rho} & -\frac{r}{\varpi} - \frac{\varpi}{\rho} \end{pmatrix} \end{aligned} \quad (31)$$

and

$$\mathfrak{m}' := \frac{1}{2} [(2C - F + iB)\sigma_1 + (A - 2E + iD)\sigma_3] = -\mathfrak{m}^*, \quad (32)$$

where the  $\sigma_k$  are Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (33)$$

We note that the principal part of  $|\rho|\mathfrak{D}$  has an additive separation property:

$$|\rho| \begin{pmatrix} 0 & \mathfrak{l}' \\ \mathfrak{l} & 0 \end{pmatrix} = \left[ \gamma^3 \varpi \partial_r + \gamma^0 \left( \varpi \partial_t + \frac{a}{\varpi} \partial_\varphi \right) \right] + \left[ \gamma^1 \partial_\theta + \gamma^2 (a \sin \theta \partial_t + \csc \theta \partial_\varphi) \right], \quad (34)$$

<sup>15</sup>Here and throughout this paper we use the *Weyl* (spinor) representation for the gamma matrices, see [46] for details.

where the coefficients of the two square-bracketed operators are functions of only  $r$ , respectively only  $\theta$ . Moreover, it is possible to transform away the lower order term in  $\mathfrak{D}$ , so that exact separation can be achieved for  $|\rho|\mathfrak{D}$ . Namely, let

$$\chi(r, \theta) := \frac{1}{2} \log(\varpi \bar{\rho} \sin \theta). \quad (35)$$

It is easy to see that

$$\mathfrak{m} = \mathfrak{l}\chi, \quad \mathfrak{m}' = \mathfrak{l}'\bar{\chi}. \quad (36)$$

Let us therefore define the diagonal matrix

$$\mathfrak{D} := \text{diag}(e^{-\chi}, e^{-\chi}, e^{-\bar{\chi}}, e^{-\bar{\chi}}) \quad (37)$$

and a new bispinor  $\hat{\Psi}$  related to the original  $\Psi$  by

$$\Psi = \mathfrak{D}\hat{\Psi}. \quad (38)$$

Denoting the upper and lower components of a bispinor  $\Psi$  by  $\psi_1$  and  $\psi_2$  respectively, it then follows that

$$(\mathfrak{l} + \mathfrak{m})\psi_1 = (\mathfrak{l} + \mathfrak{m})(e^{-\chi}\hat{\psi}_1) = e^{-\chi} [\mathfrak{l} - \mathfrak{l}\chi + \mathfrak{m}] \hat{\psi}_1 = e^{-\chi} \mathfrak{l}\hat{\psi}_1, \quad (39)$$

and similarly

$$(\mathfrak{l}' + \mathfrak{m}')\psi_2 = e^{-\bar{\chi}} \mathfrak{l}'\hat{\psi}_2. \quad (40)$$

We now put it all together. We set

$$\mathfrak{R} := \text{diag}(\rho, \rho, \bar{\rho}, \bar{\rho}) \quad (41)$$

and note that  $|\rho|\mathfrak{D}^{-*}\mathfrak{D} = \mathfrak{R}$  while  $\mathfrak{D}^{-*}\gamma^\mu\mathfrak{D} = \gamma^\mu$ . Thus, setting  $\Psi = \mathfrak{D}\hat{\Psi}$  in (7) and left-multiplying the equation by the diagonal matrix  $\mathfrak{D}' := |\rho|\mathfrak{D}^{-*}$  we conclude that  $\hat{\Psi}$  solves a new Dirac equation

$$\left( |\rho|\gamma^\mu(\mathbf{e}_\mu + ie\tilde{A}_\mu) + im\mathfrak{R} \right) \hat{\Psi} = 0. \quad (42)$$

Finally, let us compute the Hamiltonian form of (42). Let matrices  $M^\mu$  be defined by

$$|\rho|\gamma^\mu\mathbf{e}_\mu = M^\mu\partial_\mu. \quad (43)$$

Thus in particular

$$M^0 = \varpi\gamma^0 + a \sin \theta \gamma^2. \quad (44)$$

We may thus rewrite (42) as

$$M^0\partial_t\hat{\Psi} = - \left( M^k\partial_k + ie|\rho|\gamma^\mu\tilde{A}_\mu + im\mathfrak{R} \right) \hat{\Psi}, \quad (45)$$

so that, defining

$$\hat{H} := -i(M^0)^{-1} \left( M^k\partial_k + ie|\rho|\gamma^\mu\tilde{A}_\mu + im\mathfrak{R} \right), \quad (46)$$

we can now rewrite the Dirac equation (42) in Hamiltonian form:

$$i\partial_t\hat{\Psi} = \hat{H}\hat{\Psi}. \quad (47)$$

**Remark 2.7.** We note that for  $Q \neq \mathcal{I}\pi a$  the quantity  $|\rho|\gamma^\mu\tilde{A}_\mu$  in (46) is a function of both  $r$  and  $\theta$ , and unlike the other terms in the Dirac equation (42) it does not separate into a sum of two terms each depending only on one of these variables. It follows that the Dirac equation will not be exactly separable in its four spacetime variables on  $zGK$  spacetimes decorated with a generalized Sommerfeld field having a magnetic moment different from  $Qa$ .

Even when  $Q = \mathcal{I}\pi a$ , so that  $|\rho|\gamma^\mu\tilde{A}_\mu$  reduces to  $|\rho|\gamma^0\tilde{A}_0 = -(Qr/\varpi)\gamma^0$ , which is a function of only  $r$ , the separation of variables Ansatz does not yield a system of ordinary differential equations which can be solved one at a time, unlike the situation for the familiar Dirac equation for the spectrum of Hydrogen in Minkowski spacetime.

### 2.1.6 A Hilbert space for $\hat{H}$

In order to decide what is the correct inner product to use for the space of bispinor fields defined on the zGKN spacetime, we pause to consider the action for the original Dirac equation (7), which should be obtainable from this equation upon left-multiplying it by the *conjugate bispinor*  $\bar{\Psi}$ , defined as

$$\bar{\Psi} := \Psi^\dagger \gamma^0, \quad (48)$$

and integrating the result on the spacetime. Note that the dagger in the above formula is the usual notation for “conjugate-transpose”, i.e.  $\Psi^\dagger = \Psi^{*t}$ , and that  $\gamma^0$  is the zero-th Dirac gamma matrix for the Minkowski space, defined by (10).<sup>16</sup> Thus, using oblate spheroidal coordinates,

$$\mathcal{S}[\Psi] = \int dt \int_{\Sigma_t} \Psi^\dagger \gamma^0 [\tilde{\gamma}^\mu \nabla_\mu \Psi + \dots] d\mu_{\Sigma_t}, \quad (49)$$

where

$$d\mu_{\Sigma_t} = |\rho|^2 \sin \theta d\theta d\varphi dr \quad (50)$$

is the volume element of  $\Sigma_t$ , the spacelike  $t = \text{constant}$  slice of zGKN. It follows that the natural inner product for bispinors on  $\Sigma_t$  needs to be

$$\langle \Psi, \Phi \rangle = \int_{\Sigma} \Psi^\dagger \gamma^0 \tilde{\gamma}^0 \Phi d\mu_{\Sigma} = \int_0^{2\pi} \int_0^\pi \int_{-\infty}^\infty \Psi^\dagger M \Phi |\rho|^2 \sin \theta d\theta d\varphi dr, \quad (51)$$

with

$$M := \gamma^0 \tilde{\gamma}^0 = \gamma^0 \mathbf{e}_\nu^0 \gamma^\nu = \frac{\varpi}{|\rho|} \alpha^0 + \frac{a \sin \theta}{|\rho|} \alpha^2. \quad (52)$$

Here,  $\alpha^2$  is the second one of the three Dirac alpha matrices in the Weyl (spinor) representation, viz.

$$\alpha^k = \gamma^0 \gamma^k = \begin{pmatrix} \sigma_k & 0 \\ 0 & -\sigma_k \end{pmatrix}, \quad k = 1, 2, 3; \quad (53)$$

for notational convenience, we have also set

$$\alpha^0 = \begin{pmatrix} \mathbf{1}_{2 \times 2} & 0 \\ 0 & \mathbf{1}_{2 \times 2} \end{pmatrix} \quad (54)$$

for the  $4 \times 4$  identity matrix.

Now, let  $\Psi = \mathfrak{D} \hat{\Psi}$  and  $\Phi = \mathfrak{D} \hat{\Phi}$ , with  $\mathfrak{D}$  as in (37). Then we have

$$\langle \Psi, \Phi \rangle = \int_{-\infty}^\infty \int_0^{2\pi} \int_0^\pi \hat{\Psi}^\dagger \hat{M} \hat{\Phi} d\theta d\varphi dr, \quad (55)$$

where

$$\hat{M} := \alpha^0 + \frac{a \sin \theta}{\varpi} \alpha^2. \quad (56)$$

The eigenvalues of  $\hat{M}$  are  $\lambda_\pm = 1 \pm \frac{a \sin \theta}{\varpi}$ , both of which are positive everywhere on this space with Zipoy topology. (Note that  $\lambda_- \rightarrow 0$  on the ring, which is not part of the space time but at its

<sup>16</sup>On the Minkowski space, the  $\gamma^0$  matrix plays a double role: In addition to being one of the four Dirac gamma matrices, it is also the matrix of the Hermitian quadratic form of signature (2,2) defined on the bispinor space, which is one of the key geometric structures needed in order to define the Dirac operator on a general four-dimensional Lorentzian manifold, see [9] for details. In our context, the first role is played by  $\tilde{\gamma}^0$ , and the second one by  $\gamma^0$ .

boundary.) We may thus take the above as the definition of a positive definite inner product given by the matrix  $\hat{M}$  for bispinors defined on the rectangular cylinder  $\mathcal{Z} := \mathbb{R} \times [0, \pi] \times [0, 2\pi]$  (which is the  $t = \text{const.}$  section of  $\mathcal{C}$ ) with its natural measure:

$$\langle \hat{\Psi}, \hat{\Phi} \rangle_{\hat{M}} := \int_{\mathcal{Z}} \hat{\Psi}^\dagger \hat{M} \hat{\Phi} d\theta d\varphi dr. \quad (57)$$

An alternative way of arriving at this inner product is to define the conserved Dirac current

$$j^\mu = \bar{\Psi} \gamma^\mu \Psi = \Psi^\dagger \gamma^0 \gamma^\mu \Psi, \quad (58)$$

and consider the integral of its time component  $j^0$  on the Cauchy hypersurface  $\Sigma_t$  with its induced measure (50):

$$\int_{\Sigma_t} j^0 d\mu_\Sigma = \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^\pi \Psi^\dagger \gamma^0 \gamma^0 \Psi |\rho|^2 \sin \theta d\theta d\varphi dr = \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^\pi \hat{\Psi}^\dagger \hat{M} \hat{\Psi} d\theta d\varphi dr. \quad (59)$$

The corresponding Hilbert space is denoted by  $\mathbf{H}$ , thus

$$\mathbf{H} := \left\{ \hat{\Psi} : \mathcal{Z} \rightarrow \mathbb{C}^4 \mid \|\hat{\Psi}\|_{\hat{M}}^2 := \langle \hat{\Psi}, \hat{\Psi} \rangle_{\hat{M}} < \infty \right\}. \quad (60)$$

Note that  $\mathbf{H}$  is *not equivalent* to standard  $L^2(\mathcal{Z})$  whose inner product has the identity matrix in place of  $\hat{M}$ .

After these preparations we are now ready to state our main results.

## 2.2 Statement of the Main Theorems

Our results about the symmetry of the spectrum are valid for the Dirac Hamiltonian on a static spacelike slice of the zero- $G$  Kerr spacetime decorated with Sommerfeld fields of arbitrary charge  $Q$  and current  $I$ . The essential self-adjointness, and location of essential and point spectra, are stated only for the Dirac Hamiltonian on a static spacelike slice of the  $z\text{GKN}$  spacetime; however, we conjecture that these results also hold for the more general Hamiltonian as long as the coupling constant  $(Q - I\pi a)e$  is sufficiently small.

In the ensuing four sections we will prove the following Theorems about  $\hat{H}$ .

### 2.2.1 Symmetry of the spectrum of the Dirac Hamiltonians

We shall find an operator which anti-commutes with any self-adjoint extension of the formal Dirac operator  $\hat{H}$  on  $\mathbf{H}$ , with the help of which we prove:

**THEOREM 2.8.** *Let any self-adjoint extension of the formal Dirac operator  $\hat{H}$  on  $\mathbf{H}$  be denoted by the same letter. Suppose  $E \in \text{spec } \hat{H}$ . Then  $-E \in \text{spec } \hat{H}$ .*

Note that the above result holds for *any* self-adjoint extension of  $\hat{H}$ , whatever  $Q$  and  $I$  are.

### 2.2.2 Essential self-adjointness of the Dirac Hamiltonian on $z\text{GKN}$

Let  $\mathcal{Z}^*$  denote  $\mathcal{Z}$  with the ring singularity removed. By adapting an argument of Winklmeier–Yamada [51], we shall prove:

**THEOREM 2.9.** *For  $Q = I\pi a$ , i.e. for  $z\text{GKN}$ , the operator  $\hat{H}$  with domain  $C_c^\infty(\mathcal{Z}^*, \mathbb{C}^4)$  is e.s.a. in  $\mathbf{H}$ .*

### 2.2.3 The continuous spectrum of the Dirac Hamiltonians on zGKN

By adapting an argument of Weidmann [48], we shall prove:

**THEOREM 2.10.** *For  $Q = I\pi a$  the continuous spectrum of  $\hat{H}$  on  $\mathbb{H}$  is  $\mathbb{R} \setminus (-m, m)$ .*

### 2.2.4 The point spectrum of the Dirac Hamiltonian on zGKN

With the help of the Chandrasekhar–Page–Toop formalism to separate variables, and the Prüfer transform, we will be able to control the point spectrum for the zGKN Dirac Hamiltonian:

**THEOREM 2.11.** *Suppose  $Q = I\pi a$ . Then, if  $2m|a| < 1$  and  $|eQ| < \sqrt{2m|a|(1 - 2m|a|)}$ , the point spectrum of  $\hat{H}$  on  $\mathbb{H}$  is nonempty and located in  $(-m, m)$ ; the end points are not included.*

This completes the formulation of our main results. We next turn to their proofs. The proofs of our main theorems are distributed over four sections corresponding to the various aspects of the spectrum, i.e. symmetry, essential self-adjointness, continuous spectrum, and point spectrum.

## 3 Proof of Theorem 2.8 (Symmetry of the energy spectrum)

Suppose  $E \in \mathbb{R}$  is an eigenvalue of  $\hat{H}$ . Then there exists  $\hat{\Psi} \in \mathbb{H}$  such that

$$\hat{H}\hat{\Psi} = E\hat{\Psi}. \quad (61)$$

Suppose one can find a bounded linear, or conjugate-linear, operator  $\hat{C} : \mathbb{H} \rightarrow \mathbb{H}$  that anti-commutes with  $\hat{H}$ , i.e.

$$[\hat{C}, \hat{H}]_+ = \hat{C}\hat{H} + \hat{H}\hat{C} = 0. \quad (62)$$

It is then easy to see that  $-E$  must also be an eigenvalue of  $\hat{H}$ , since

$$\hat{H}\hat{C}\hat{\Psi} = -\hat{C}\hat{H}\hat{\Psi} = -\hat{C}E\hat{\Psi} = -E\hat{C}\hat{\Psi}. \quad (63)$$

This argument can be extended to show the symmetry of other parts of the spectrum. (See e.g. Glazman [22], p. 205.)

Let  $\hat{K} : \mathbb{H} \rightarrow \mathbb{H}$  denote the complex conjugation operator  $\hat{K}\hat{\Psi}(x) = \hat{\Psi}^*(x)$ , and let  $\hat{S} : \mathbb{H} \rightarrow \mathbb{H}$  denote the operator  $(\hat{S}\hat{\Psi})(x) = \hat{\Psi}(\varsigma(x))$  where  $\varsigma : \mathcal{Z} \rightarrow \mathcal{Z}$  is the sheet swapping map,

$$\varsigma(r, \theta, \varphi) = (-r, \pi - \theta, \varphi). \quad (64)$$

We claim that the operator  $\hat{C} : \mathbb{H} \rightarrow \mathbb{H}$  given (in Weyl representation) by  $\hat{C} := \gamma^0 \hat{K} \hat{S}$ , viz.

$$(\hat{C}\hat{\Psi})(x) = \gamma^0 \hat{\Psi}^*(\varsigma(x)), \quad (65)$$

anti-commutes with  $\hat{H}$ . Note that the double-sheetedness of the underlying space plays an essential role in the definition of this operator.

**Remark 3.1.** *The operator  $\hat{C}$  should not be confused with the operator  $\tilde{C}$  given in Weyl representation by*

$$\tilde{C} := i\gamma^2 \hat{K}.$$

*One easily checks that if  $\hat{\Psi}$  solves  $i\hbar\partial_t \hat{\Psi} = (\hat{H}_0 + e\mathcal{A})\hat{\Psi}$ , then  $\tilde{C}\hat{\Psi}$  solves  $i\hbar\partial_t (\tilde{C}\hat{\Psi}) = (\hat{H}_0 - e\mathcal{A})(\tilde{C}\hat{\Psi})$ . In particular, if  $\hat{\Psi}$  is an eigen-bi-spinor of  $\hat{H}_0 + e\mathcal{A}$  with eigenvalue  $E$ , then  $\tilde{C}\hat{\Psi}$  is an eigen-bi-spinor of  $\hat{H}_0 - e\mathcal{A}$  with eigenvalue  $-E$  (note that the two Hamiltonians here are different!). For this reason  $\tilde{C}$  is called the charge conjugation operator.*

To prove the claim, first note that  $\gamma^0 = \beta$  anti-commutes with all three  $\alpha^k$  matrices. Recall that

$$\hat{H}(x) = \hat{M}^{-1} \mathfrak{H} \quad (66)$$

and

$$\mathfrak{H} := -i\alpha^3 \partial_r + \frac{1}{\varpi} \left( -i\alpha^1 \partial_\theta - i\alpha^2 \csc \theta \partial_\varphi \right) - \frac{ia}{\varpi^2} \alpha^0 \partial_\varphi + \frac{m}{\varpi} \gamma^0 \mathfrak{R} + \frac{e|\rho|}{\varpi} \left( \tilde{A}^0(x) \alpha^0 + \tilde{A}^2(x) \alpha^2 \right). \quad (67)$$

Now

$$\hat{M}^{-1} = \frac{\varpi^2}{|\rho|^2} \left( \alpha^0 - \frac{a \sin \theta}{\varpi} \alpha^2 \right). \quad (68)$$

Thus, keeping in mind that  $\overline{\alpha^2} = -\alpha^2$ , we find that

$$\hat{C} \hat{M}^{-1} = \gamma^0 \overline{\hat{M}^{-1} \circ \varsigma} \hat{K} \hat{S} = \frac{\varpi^2}{|\rho|^2} \gamma^0 \left( \alpha^0 + \frac{a \sin \theta}{\varpi} \alpha^2 \right) \hat{K} \hat{S} = \frac{\varpi^2}{|\rho|^2} \left( \alpha^0 - \frac{a \sin \theta}{\varpi} \alpha^2 \right) \gamma^0 \hat{K} \hat{S} = \hat{M}^{-1} \hat{C}. \quad (69)$$

So we only need to check that  $\hat{C}$  anti-commutes with  $\mathfrak{H}$ .

It is enough to check that each term in  $\mathfrak{H}$  goes through an odd number of sign changes (either one or three) as the three operators  $\hat{K}$ ,  $\hat{S}$ , and multiplication by  $\gamma^0$ , filter through that term. Recalling that the potential  $\mathbf{A}$  is anti-symmetric with respect to sheet swap:  $\tilde{A}_\mu \circ \varsigma = -\tilde{A}_\mu$ , this becomes obvious for most terms in  $\mathfrak{H}$ . Only the term involving  $\mathfrak{R}$  requires some care. We first check that  $\mathfrak{R} \circ \varsigma = -\mathfrak{R}$  and that  $\gamma^0 \overline{\mathfrak{R}} = \mathfrak{R} \gamma^0$ . Then

$$\hat{C} \gamma^0 \mathfrak{R} = \gamma^0 \gamma^0 \overline{\mathfrak{R} \circ \varsigma} \hat{K} \hat{S} = -\gamma^0 \gamma^0 \overline{\mathfrak{R}} \hat{K} \hat{S} = -\gamma^0 \mathfrak{R} \gamma^0 \hat{K} \hat{S} = -\gamma^0 \mathfrak{R} \hat{C}, \quad (70)$$

establishing the anti-commutation property. The proof of Theorem 2.8 is complete.

Before moving on to the proof of the next theorem on the list, we pause briefly to recall our earlier discussion that showed that the physics does not change if in the zGKN electromagnetic spacetime solution one changes  $a \rightarrow -a$ , respectively changes  $I \rightarrow -I$  or  $a \rightarrow -a$  in its generalization involving  $Qa \rightarrow I\pi a^2$ . This suggests that the spectrum of our Dirac Hamiltonian must be invariant under these transformations. However, recall that our co-frame is adapted to the electromagnetic fields written as in (5), respectively (6), so that a sign change  $a \rightarrow -a$  needs to be accompanied by a sign change  $I \rightarrow -I$  in order for the physics (here: the spectrum of the Hamiltonian) to remain unchanged. We now use a variant of the strategy of proof of Theorem 2.8 to prove exactly this. Thus we write the Hamiltonian defined by (66) as  $\hat{H} = \hat{H}_{a,I}$  to emphasize the dependence on the two parameters  $a$  and  $I$ .

**PROPOSITION 3.2.** *There exists an involutive isometry  $C : \mathbb{H} \rightarrow \mathbb{H}$  such that*

$$C \hat{H}_{a,I} = \hat{H}_{-a,-I} C. \quad (71)$$

*Thus, the spectral properties of the two Hamiltonians are identical.*

*Proof.* Let  $T : \mathbb{H} \rightarrow \mathbb{H}$  be defined by

$$T \hat{\Psi}(r, \theta, \varphi) = \hat{\Psi}(r, \pi - \theta, -\varphi),$$

and let  $\vec{S}$  denote the ‘‘classical spin operator,’’ in Weyl representation given by:

$$S_k := \begin{pmatrix} \sigma_k & \\ & \sigma_k \end{pmatrix}, \quad k = 1, 2, 3.$$

It can then be readily checked from (66), (67), (68), and (23) that

$$C := S_3 T$$

will do the job. □

## 4 Proof of Theorem 2.9 (Essential self-adjointness ( $Q = I\pi a$ ))

We now show that the Dirac Hamiltonian  $\hat{H}$  is essentially self-adjoint on  $\mathbf{H}$ ; recall that  $\mathbf{H}$  is equipped with the inner product<sup>17</sup> (57).

We observe that  $M^0 = \varpi\gamma^0\hat{M}$ , so we may rewrite (46) as

$$\hat{H} = \hat{M}^{-1}\gamma^0 \left( \frac{-i}{\varpi} M^k \partial_k + e \frac{|\rho|}{\varpi} \gamma^\mu \tilde{A}_\mu + \frac{m}{\varpi} \mathfrak{R} \right) = \hat{M}^{-1} \mathfrak{H}, \quad (72)$$

where

$$\mathfrak{H} := \mathfrak{M} + \frac{1}{\varpi} \mathfrak{N} + \mathfrak{P} + \mathfrak{Q}, \quad (73)$$

with

$$\mathfrak{M} := -i\alpha^3 \partial_r \quad (74)$$

$$\mathfrak{N} := -i\alpha^1 \partial_\theta - i\alpha^2 \csc \theta \partial_\varphi \quad (75)$$

$$\mathfrak{P} := -i \frac{a}{\varpi^2} \alpha^0 \partial_\varphi + \frac{m}{\varpi} \gamma^0 \mathfrak{R} \quad (76)$$

$$\mathfrak{Q} := e \frac{|\rho|}{\varpi} \gamma^0 \gamma^\mu \tilde{A}_\mu. \quad (77)$$

Thus,

$$\langle \hat{\Psi}, \hat{H} \hat{\Phi} \rangle_{\hat{M}} = \int_{\mathcal{Z}} \hat{\Psi} \mathfrak{H} \hat{\Phi} d\theta d\varphi dr. \quad (78)$$

Evidently,  $\mathfrak{H}$  is Hermitian symmetric on the Hilbert space  $L^2(\mathcal{Z}; \mathbb{C}^4)$  with its natural inner product

$$(\hat{\Phi}, \hat{\Psi}) = \int_{\mathcal{Z}} \hat{\Phi}^\dagger \hat{\Psi} d\theta d\varphi dr. \quad (79)$$

It is furthermore easy to see that  $\hat{H}$  is e.s.a. on  $\mathbf{H}$  if and only if  $\mathfrak{H}$  is e.s.a. on  $L^2(\mathcal{Z}; \mathbb{C}^4)$ .

We shall prove that  $\mathfrak{H}$  is e.s.a. on  $L^2(\mathcal{Z}; \mathbb{C}^4)$  when  $Q = I\pi a$ , i.e. for a Dirac point electron in zGKN.

**THEOREM 4.1.** *For  $Q = I\pi a$  the operator  $\mathfrak{H}$  with domain  $C_c^\infty(\mathcal{Z}^*, \mathbb{C}^4)$  is e.s.a. in  $L^2(\mathcal{Z}, \mathbb{C}^4)$ .*

*Proof.* Let us write

$$\mathfrak{H} = \mathfrak{H}^0 + \mathfrak{Q}. \quad (80)$$

Here,  $\mathfrak{H}^0$  is the free Hamiltonian. We will first show that  $\mathfrak{H}^0$  is essentially self-adjoint; this proof is an easy adaptation of the method first employed by Winklmeier and Yamada [51]. We will then conclude essential self-adjointness of  $\mathfrak{H}$  by using a perturbation argument.

To this end, let us consider the decomposition with respect to the azimuthal angle  $\varphi$  of the Hilbert space  $L^2(\mathcal{Z}; \mathbb{C}^4)$  into partial wave subspaces  $L^2([0, \pi] \times \mathbb{R}, d\theta dr)$ :

$$L^2(\mathcal{Z}; \mathbb{C}^4) = \bigoplus_{\kappa \in \mathbb{Z} + \frac{1}{2}} \left( L^2_\kappa([0, \pi] \times \mathbb{R}, d\theta dr) \right)^4 \quad (81)$$

<sup>17</sup>As pointed out by the anonymous referee, an alternate proof of essential self-adjointness might be possible following the strategy of Chernoff [15], who proved essential self-adjointness of the Dirac operator on certain *complete* spacetimes. Such a proof would be highly welcome, indeed, for it would avoid a partial wave decomposition and be more direct. However, due to the presence of the naked ring singularity in the zGKN spacetime, the underlying manifold is not complete and thus it is far from obvious how to generalize Chernoff's result.

corresponding to the expansion of a bispinor field  $\hat{\Psi} \in L^2(\mathcal{Z}; \mathbb{C}^4)$  given by

$$\hat{\Psi}(r, \theta, \varphi) = \sum_{\kappa \in \mathbb{Z} + \frac{1}{2}} e^{i\kappa\varphi} \hat{\Psi}_\kappa(r, \theta). \quad (82)$$

For a discussion of why  $\kappa$  needs to be a half-integer, see [18, 17].

Let  $\mathfrak{H}_\kappa^0 := \mathfrak{H}^0|_{L_\kappa^2}$ . Then  $\mathfrak{H}_\kappa^0 = \mathfrak{S}_\kappa + \varpi^{-1}\mathfrak{T}_\kappa + \mathfrak{B}_\kappa$ , with

$$\mathfrak{S}_\kappa = -i\alpha^3\partial_r, \quad \mathfrak{T}_\kappa = -i\alpha^1\partial_\theta + \alpha^2\kappa \csc\theta = \begin{pmatrix} \mathfrak{t}_\kappa & 0 \\ 0 & -\mathfrak{t}_\kappa \end{pmatrix}, \quad \mathfrak{B}_\kappa = \frac{a\kappa}{\varpi^2}\alpha^0 + \frac{m}{\varpi}\gamma^0\mathfrak{A}. \quad (83)$$

We note that  $\mathfrak{B}_\kappa$  is a symmetric *bounded* multiplication operator on  $L_\kappa^2$ ; in fact,

$$\|\mathfrak{B}_\kappa\|_{L^\infty} \leq |\kappa/a| + m, \quad (84)$$

so that the task of showing e.s.a.-ness of  $\mathfrak{H}_\kappa^0$  reduces to showing e.s.a.-ness of  $\mathfrak{H}'_\kappa := \mathfrak{S}_\kappa + \varpi^{-1}\mathfrak{T}_\kappa$ .

Now  $\mathfrak{H}'_\kappa$  is block-diagonal:

$$\mathfrak{H}'_\kappa = \begin{pmatrix} \mathfrak{h}'_\kappa & 0 \\ 0 & -\mathfrak{h}'_\kappa \end{pmatrix}, \quad \mathfrak{h}'_\kappa := -i\sigma_3\partial_r + \varpi^{-1}\mathfrak{t}_\kappa, \quad \mathfrak{t}_\kappa := -i\sigma_1\partial_\theta + \sigma_2\kappa \csc\theta. \quad (85)$$

Thus it is enough to show  $\mathfrak{h}'_\kappa$  is e.s.a. We do so by showing that  $\ker(\mathfrak{h}'_\kappa \pm i) = \{0\}$ : Suppose  $\hat{\psi}_\kappa \in (L_\kappa^2([0, \pi] \times \mathbb{R}, d\theta dr))^2$  satisfies

$$\mathfrak{h}'_\kappa \hat{\psi}_\kappa = \pm i \hat{\psi}_\kappa. \quad (86)$$

As observed in [51], it is possible to decompose (86) with respect to the eigenspaces of the operator

$$\mathfrak{a}_\kappa := W\mathfrak{t}_\kappa W^{-1} = -i\sigma_2\partial_\theta + \kappa \csc\theta\sigma_1, \quad (87)$$

where

$$W := \begin{pmatrix} 0 & 1 \\ i & 0 \end{pmatrix}. \quad (88)$$

The operator  $\mathfrak{a}_\kappa$  has pure point spectrum and a complete set of eigenfunctions. More precisely, one has the following result [49] (here quoted from [51]):

**THEOREM 4.2.** (*Winklmeier, 2006*) *For all  $\kappa \in \mathbb{Z} + \frac{1}{2}$  the operator  $\mathfrak{a}_\kappa$  with domain  $(C_c^\infty((0, \pi)))^2$  is essentially self-adjoint in  $(L^2((0, \pi), d\theta))^2$ . Its closure (denoted again by  $\mathfrak{a}_\kappa$ ) is compactly invertible and its spectrum consists of simple eigenvalues only, given by*

$$\lambda_n^\kappa := \operatorname{sgn}(n) \left( |\kappa| - \frac{1}{2} + |n| \right), \quad n \in \mathbb{Z}^* = \mathbb{Z} \setminus \{0\}, \quad (89)$$

with corresponding normalized eigenfunctions  $\{g_n^\kappa\}_{n \in \mathbb{Z}^*}$  forming a complete orthonormal set in  $(L^2((0, \pi), d\theta))^2$ . Moreover,

$$\lambda_{-n}^\kappa = -\lambda_n^\kappa, \quad g_{-n}^\kappa = -\sigma_3 g_n^\kappa. \quad (90)$$

We can therefore write

$$\hat{\psi}_\kappa(r, \theta) = \sum_{n \in \mathbb{Z}^*} \xi_n(r) g_n^\kappa(\theta), \quad (91)$$

with functions  $\xi_n \in L^2(\mathbb{R}, dr)$ . Hence, performing a similarity transform on (86) with  $W$  and projecting on the spans of  $g_n^\kappa$  and  $g_{-n}^\kappa$  we obtain the following system (see [51] for details):

$$\begin{pmatrix} \frac{\lambda_n^\kappa}{\varpi} & i\partial_r \\ i\partial_r & -\frac{\lambda_n^\kappa}{\varpi} \end{pmatrix} \begin{pmatrix} \xi_n \\ \xi_{-n} \end{pmatrix} = \pm i \begin{pmatrix} \xi_n \\ \xi_{-n} \end{pmatrix} \quad (92)$$

However the operator in the above eigenvalue problem  $\mathcal{C}_\kappa = i\sigma_1\partial_r + \frac{\lambda_n^\kappa}{\varpi}\sigma_3$  is clearly e.s.a., since  $\frac{\lambda_n^\kappa}{\varpi}$  is bounded, hence  $\xi_n = \xi_{-n} = 0$ .

This completes the proof of essential self-adjointness of  $\mathfrak{H}^0$ .

Consider now the term  $\mathfrak{Q} = e\frac{|\rho|}{\varpi}\gamma^0\gamma^\mu\tilde{A}_\mu$  coming from the electromagnetic potential. It can be rewritten as  $\mathfrak{Q} = -eQ\mathfrak{V}_1 - e(Q - I\pi a)\mathfrak{V}_2$ , where

$$\mathfrak{V}_1 := \frac{r}{\varpi^2}\alpha^0, \quad \mathfrak{V}_2 := \frac{ar\sin\theta}{\varpi|\rho|^2}\hat{M}. \quad (93)$$

The first term,  $\mathfrak{V}_1$ , is clearly bounded, whereas the second one,  $\mathfrak{V}_2$ , blows up on the ring. However, since by hypothesis we restrict ourselves to the case  $Q = I\pi a$ , the  $\mathfrak{V}_2$  term is absent from  $\mathfrak{Q}$ , and essential self-adjointness of  $\mathfrak{H}$  follows easily from that of  $\mathfrak{H}^0$  and the boundedness of  $eQ\mathfrak{V}_1$ .

The proof of Theorem 4.1 is complete.  $\square$

For the proof the remaining statements in this paper we rely on the fact that the Dirac equation of a point electron in zGKN separates into four (coupled) ordinary differential equations, each of which depends on only one of the four oblate spheroidal coordinates, with the coupling being effected through shared parameters in the equations. This is carried out in the next section before we resume with proving our claims.

## 5 Chandrasekhar–Page–Toop separation-of-variables ( $Q = I\pi a$ )

When  $Q = I\pi a$  the Dirac equation (47) for the bispinor  $\hat{\Psi}$  allows a clear separation also for the remaining  $r$  and  $\theta$  derivatives (commonly referred to in the literature as “radial” and “angular” derivatives, even though  $r$  is not a radial distance and  $\theta$  is not an angle, except at infinity). Thus, when  $Q = I\pi a$  the Dirac equation (47) becomes

$$(\hat{R} + \hat{A})\hat{\Psi} = 0, \quad (94)$$

where

$$\hat{R} := \begin{pmatrix} imr & 0 & D_- + ieQ\frac{r}{\varpi} & 0 \\ 0 & imr & 0 & D_+ + ieQ\frac{r}{\varpi} \\ D_+ + ieQ\frac{r}{\varpi} & 0 & imr & 0 \\ 0 & D_- + ieQ\frac{r}{\varpi} & 0 & imr \end{pmatrix}, \quad (95)$$

$$\hat{A} := \begin{pmatrix} -ma\cos\theta & 0 & 0 & -L_- \\ 0 & -ma\cos\theta & -L_+ & 0 \\ 0 & L_- & ma\cos\theta & 0 \\ L_+ & 0 & 0 & ma\cos\theta \end{pmatrix}, \quad (96)$$

where  $D_\pm$  and  $L_\pm$  have been given in (30). Once a solution  $\hat{\Psi}$  to (94) is found, the bispinor  $\Psi := \mathfrak{D}\hat{\Psi}$  solves the original Dirac equation (7).

### 5.0.5 The Chandrasekhar Ansatz

Assume now that a solution  $\hat{\Psi}$  of (94) is of the form

$$\hat{\Psi} = e^{-i(Et - \kappa\varphi)} \begin{pmatrix} R_1S_1 \\ R_2S_2 \\ R_2S_1 \\ R_1S_2 \end{pmatrix}, \quad (97)$$

with  $R_k$  being complex-valued functions of  $r$  alone, and  $S_k$  real-valued functions of  $\theta$  alone. Let

$$\vec{R} := \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, \quad \vec{S} := \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}. \quad (98)$$

Plugging the Chandrasekhar Ansatz (97) into (94) one easily finds that there must be  $\lambda \in \mathbb{C}$  such that

$$T_{rad}\vec{R} = E\vec{R}, \quad (99)$$

$$T_{ang}\vec{S} = \lambda\vec{S}, \quad (100)$$

where

$$T_{rad} := \begin{pmatrix} d_- & -m\frac{r}{\varpi} - i\frac{\lambda}{\varpi} \\ -m\frac{r}{\varpi} + i\frac{\lambda}{\varpi} & -d_+ \end{pmatrix} \quad (101)$$

$$T_{ang} := \begin{pmatrix} -ma \cos \theta & -l_- \\ l_+ & ma \cos \theta \end{pmatrix} \quad (102)$$

The operators  $d_{\pm}$  and  $l_{\pm}$  are now ordinary differential operators in  $r$  and  $\theta$  respectively, with coefficients that depend on the unknown  $E$ , and parameters  $a$ ,  $\kappa$ , and  $eQ$ :

$$d_{\pm} := i\frac{d}{dr} \pm \frac{-a\kappa + eQr}{\varpi^2} \quad (103)$$

$$l_{\pm} := \frac{d}{d\theta} \mp (aE \sin \theta - \kappa \csc \theta) \quad (104)$$

The angular operator  $T_{ang}$  in (100) is easily seen to be essentially self-adjoint on  $(C_c^\infty((0, \pi), \sin \theta d\theta))^2$  and in fact is self-adjoint on its domain inside  $(L^2((0, \pi), \sin \theta d\theta))^2$  (e.g. [42, 3]) with purely point spectrum  $\lambda = \lambda_n(am, aE, \kappa)$ ,  $n \in \mathbb{Z} \setminus 0$ . Thus in particular  $\lambda \in \mathbb{R}$ . It then follows that the radial operator  $T_{rad}$  is also essentially self-adjoint on  $(C_c^\infty(\mathbb{R}, dr))^2$  and in fact self-adjoint on its domain inside  $(L^2(\mathbb{R}, dr))^2$ .

Suppose  $\vec{R} = (R_1, R_2)^T \in (L^2(\mathbb{R}))^2$  is a nontrivial solution to  $T_{rad}\vec{R} = E\vec{R}$ , with  $E \in \mathbb{R}$ . Then

$$\begin{aligned} \frac{dR_1}{dr} - i \left( E - \frac{a\kappa - eQr}{\varpi^2} \right) R_1 + \frac{1}{\varpi} (imr - \lambda) R_2 &= 0 \\ -\frac{dR_2}{dr} - i \left( E - \frac{a\kappa - eQr}{\varpi^2} \right) R_2 + \frac{1}{\varpi} (imr + \lambda) R_1 &= 0. \end{aligned}$$

Multiply the first equation by  $\bar{R}_1$  and the second equation by  $\bar{R}_2$ , add them and take the real part, to obtain

$$\frac{d}{dr} (|R_1|^2 - |R_2|^2) = 0. \quad (105)$$

Thus the difference of the moduli squared of  $R_1$  and  $R_2$  is constant, hence zero since they need to be integrable at infinity. I.e.,

$$|R_1| = |R_2| := R. \quad (106)$$

Let  $R_j = Re^{i\Phi_j}$  for  $j = 1, 2$ . Multiply the first equation by  $\bar{R}_2$ , multiply the complex conjugate of the second equation by  $R_1$ , and add them to obtain

$$\frac{d}{dr} \left( \frac{R_1}{\bar{R}_2} \right) = 0. \quad (107)$$

Thus the ratio  $R_1/\bar{R}_2$ , and hence the sum of the arguments  $\Phi_1 + \Phi_2$  must be a constant, say  $\delta$ . Thus  $R_1 = \bar{R}_2 e^{i\delta}$ . Since multiplication by a constant phase factor is a gauge transformation for Dirac bispinors, we can replace  $\hat{\Psi}$  with  $\hat{\Psi}' = e^{-i\delta/2} \hat{\Psi}$  without changing anything. The spinor thus obtained has the same form as (97), now with  $R'_1 = \bar{R}'_2$ . Thus without loss of generality we can assume  $\delta = 0$  and  $R_1 = \bar{R}_2$ .

This motivates us to set

$$R_1 = \frac{1}{\sqrt{2}}(v - iu), \quad R_2 = \frac{1}{\sqrt{2}}(v + iu) \quad (108)$$

for real functions  $u$  and  $v$ . Consider the unitary matrix

$$U := \frac{1}{\sqrt{2}} \begin{pmatrix} -i & 1 \\ i & 1 \end{pmatrix}. \quad (109)$$

A change of basis using  $U$  brings the radial system (99) into the following standard (Hamiltonian) form

$$(H_{rad} - E) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (110)$$

where

$$H_{rad} := \begin{pmatrix} m \frac{r}{\varpi} + \frac{\gamma r + a\kappa}{\varpi^2} & -\partial_r + \frac{\lambda}{\varpi} \\ \partial_r + \frac{\lambda}{\varpi} & -m \frac{r}{\varpi} + \frac{\gamma r + a\kappa}{\varpi^2} \end{pmatrix}, \quad (111)$$

(cf. [46], eq (7.105)) with

$$\gamma := -eQ < 0. \quad (112)$$

## 6 Proof of Theorem 2.10 (Continuous spectrum of $\hat{H}$ on $zGKN$ )

Following Weidmann [48] we now prove the theorem about the continuous spectrum of  $\hat{H}$ . Recall the partial wave decomposition (81). Let  $\hat{H}_\kappa$  denote the restriction of  $\hat{H}$  to  $L^2_\kappa$ . The Chandrasekhar separation (97) and equation (99) yield that the spectrum of  $\hat{H}_\kappa$  coincides with that of  $T_{rad}$ , which coincides with that of  $H_{rad}$  since these last two are unitarily equivalent. Furthermore, the spectrum of  $\hat{H}$  equals the union of the spectra of  $\hat{H}_\kappa$ . Thus in order to prove the claim about the essential spectrum, it suffices to show that it holds for  $H_{rad}$  regardless of the values of  $\kappa$  and  $\lambda$ .

Since  $H_{rad}$  is a radial Dirac operator, one can then use results that are particular to one dimension. One such result is due to Weidmann [48]:

**THEOREM.** *Let  $P$  and  $J$  be matrices such that  $H_{rad} = J\partial_r + P$ . Suppose  $P$  can be written as  $P_1 + P_2$  in such a way that each component of  $P_1$  is integrable in  $[R, \infty)$  for some  $R > 0$ ,  $P_2$  is of bounded variation on  $[R, \infty)$  and*

$$\lim_{r \rightarrow \infty} P(r) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad a > b.$$

*Then each self-adjoint extension of  $h$  has a purely absolutely continuous spectrum in  $(-\infty, b] \cup [a, \infty)$ .*

Using this result, our claim follows by noticing that the hypotheses on  $P$  are satisfied, and

$$\lim_{r \rightarrow \infty} P(r) = \begin{pmatrix} m & 0 \\ 0 & -m \end{pmatrix}.$$

Proof of Theorem 2.10 is complete.

## 7 Proof of Theorem 2.11 (Point spectrum of $\hat{H}$ on zGKN)

By the remarks at the beginning of Section 6, we are interested in the eigenvalues  $E$  and square-integrable eigenfunctions in  $L^2(\mathbb{R}, dr)^2$  of the operator  $H_{\text{rad}}$ . One complication is that in our case the radial Hamiltonian  $H_{\text{rad}}$  depends on the unknown eigenvalues  $\lambda$  of the angular operator  $T_{\text{ang}}$  in (100), which in turn depend on the energy  $E$ . Since the angular operator is the same as the one on Kerr and Kerr-Newman spacetime studied in [42, 3], and since it is known that for a given value of  $E$  there is a largest negative eigenvalue  $\lambda = \Lambda(E)$ , our strategy is to show the existence, for a given value of  $\lambda < 0$ , of a smallest positive eigenvalue  $E = \mathcal{E}(\lambda)$  for  $H_{\text{rad}}$ , and then set up an iteration that converges to a pair  $(E, \lambda)$  for which the radial (99) and the angular (100) equations jointly have  $L^2$  solutions, thereby establishing the existence of a “positive-energy eigenstate” for the full Dirac Hamiltonian; note that by the symmetry of the spectrum there also exists a “negative-energy eigenstate.”

### 7.1 The Prüfer transform

Consider the equations (110) and (100) for unknowns  $(u, v)$  and  $(S_1, S_2)$ . Let us define new unknowns  $(R, \Omega)$  and  $(S, \Theta)$  via the Prüfer transform [39]

$$u = \sqrt{2}R \cos \frac{\Omega}{2}, \quad v = \sqrt{2}R \sin \frac{\Omega}{2}, \quad S_1 = S \cos \frac{\Theta}{2}, \quad S_2 = S \sin \frac{\Theta}{2}. \quad (113)$$

Thus

$$R = \frac{1}{2}\sqrt{u^2 + v^2}, \quad \Omega = 2 \tan^{-1} \frac{v}{u}, \quad S = \sqrt{S_1^2 + S_2^2}, \quad \Theta = 2 \tan^{-1} \frac{S_2}{S_1}. \quad (114)$$

As a result,  $R_1 = -iRe^{i\Omega/2}$  and  $R_2 = iRe^{-i\Omega/2}$ . Hence  $\hat{\Psi}$  can be re-expressed in terms of the Prüfer variables, thus

$$\hat{\Psi}(t, r, \theta, \varphi) = R(r)S(\theta)e^{-i(Et - \kappa\varphi)} \begin{pmatrix} -i \cos(\Theta(\theta)/2)e^{+i\Omega(r)/2} \\ +i \sin(\Theta(\theta)/2)e^{-i\Omega(r)/2} \\ +i \cos(\Theta(\theta)/2)e^{-i\Omega(r)/2} \\ -i \sin(\Theta(\theta)/2)e^{+i\Omega(r)/2} \end{pmatrix}, \quad (115)$$

and we obtain the following equations for the new unknowns

$$\frac{d}{dr}\Omega = 2\frac{mr}{\varpi} \cos \Omega + 2\frac{\lambda}{\varpi} \sin \Omega + 2\frac{a\kappa + \gamma r}{\varpi^2} - 2E, \quad (116)$$

$$\frac{d}{dr} \ln R = \frac{mr}{\varpi} \sin \Omega - \frac{\lambda}{\varpi} \cos \Omega. \quad (117)$$

Similarly,

$$\frac{d}{d\theta}\Theta = -2ma \cos \theta \cos \Theta + 2 \left( aE \sin \theta - \frac{\kappa}{\sin \theta} \right) \sin \Theta + 2\lambda, \quad (118)$$

$$\frac{d}{d\theta} \ln S = -ma \cos \theta \sin \Theta - \left( aE \sin \theta - \frac{\kappa}{\sin \theta} \right) \cos \Theta. \quad (119)$$

To simplify the analysis of these systems and reduce the number of parameters involved, we will henceforth set  $m = 1$ . Note that this is always possible by defining the constants  $a' = ma$ ,  $E' = E/m$ , and a change of variable  $r' = mr$ .

**Remark 7.1.** Equations (116)–(119) have certain symmetries that are connected with the action of operators  $\hat{C}$  and  $\tilde{C}$  introduced in Section 3: It is easy to see that these four equations are preserved under each of the following two transformations:

$$r \rightarrow -r, \quad \theta \rightarrow \pi - \theta, \quad \lambda \rightarrow -\lambda, \quad \kappa \rightarrow -\kappa, \quad E \rightarrow -E, \quad (120)$$

and

$$\Omega \rightarrow \pi - \Omega, \quad \Theta \rightarrow \pi - \Theta, \quad \lambda \rightarrow -\lambda, \quad \kappa \rightarrow -\kappa, \quad E \rightarrow -E, \quad \gamma \rightarrow -\gamma. \quad (121)$$

The map (120) corresponds to the action of  $\hat{C}$  and the map (121) to the action of  $\tilde{C}$ . Note that the latter does not preserve the Hamiltonian, since the sign of  $\gamma = -eQ$  is changed.

## 7.2 The realm of $L^2$ solutions

We note that in both of the above systems (116,117) and (118,119), when a solution to the first equation is known, the second equation in the system can be solved by quadrature. Moreover, the requirement that  $R$  and  $S$  be  $L^2$  functions of their argument determines what boundary values the solutions to the  $\Omega$  and  $\Theta$  equations should have. More precisely,

**PROPOSITION 7.2.** Any bispinor  $\hat{\Psi}$  of the form (115) constructed from solutions of (116), (117), (118), (119), with  $|\kappa| \geq \frac{1}{2}$ ,  $E > 0$  and  $\lambda < 0$ , belongs to the Hilbert space  $\mathbf{H}$  provided

$$\lim_{r \rightarrow -\infty} \Omega(r) = -\pi + \cos^{-1}(E), \quad \lim_{r \rightarrow \infty} \Omega(r) = -\cos^{-1}(E), \quad (122)$$

and

$$\Theta(0) = 0, \quad \Theta(\pi) = -\pi. \quad (123)$$

*Proof.* It is straightforward to compute that for a  $\hat{\Psi}$  of the form (115),

$$\begin{aligned} \|\hat{\Psi}\|_{\hat{M}}^2 &= 2 \int_0^{2\pi} \int_0^\pi \int_{-\infty}^\infty R^2(r) S^2(\theta) \left( 1 + \frac{a \sin \theta}{\varpi} \sin \Theta(\theta) \sin \Omega(r) \right) dr d\theta d\varphi \\ &= 4\pi \left[ \int_{-\infty}^\infty R^2 dr \int_0^\pi S^2 d\theta + a \int_{-\infty}^\infty R^2 \sin \Omega \frac{dr}{\varpi} \int_0^\pi S^2 \sin \Theta \sin \theta d\theta \right] \\ &\leq 8\pi \|R\|_{L^2}^2 \|S\|_{L^2}^2, \end{aligned}$$

and thus  $\hat{\Psi} \in \mathbf{H}$  provided  $R \in L^2(\mathbb{R}, dr)$  and  $S \in L^2((0, \pi), d\theta)$ .

Now (118) can be written as a smooth dynamical system in the  $(\theta, \Theta)$  plane by introducing a new independent variable  $\tau$  such that  $\frac{d\theta}{d\tau} = \sin \theta$ . Then, with dot representing differentiation in  $\tau$ , we have,

$$\begin{cases} \dot{\theta} &= \sin \theta \\ \dot{\Theta} &= -2a \sin \theta \cos \theta \cos \Theta + 2aE \sin^2 \theta \sin \Theta - 2\kappa \sin \Theta + 2\lambda \sin \theta \end{cases} \quad (124)$$

Identifying the line  $\Theta = \pi$  with  $\Theta = -\pi$ , this becomes a dynamical system on a closed finite cylinder  $\mathcal{C}_1 = [0, \pi] \times \mathbb{S}^1$ . The only equilibrium points of the flow are on the two circular boundaries: Two on the left boundary:  $S_- = (0, 0)$ ,  $N_- = (0, \pi)$ ; two on the right:  $S_+ = (\pi, -\pi)$  and  $N_+ = (\pi, 0)$ .

For  $\kappa > 0$ , the linearization of the flow at the equilibrium points reveals that  $S_-$  and  $S_+$  are hyperbolic saddle points (with eigenvalues  $\{1, -2\kappa\}$  and  $\{-1, 2\kappa\}$  respectively), while  $N_-$  is a source node (with eigenvalues 1 and  $2\kappa$ ) and  $N_+$  is a sink node (with eigenvalues  $-1$  and  $-2\kappa$ ). Note that the situation with  $\kappa < 0$  is entirely analogous, with the critical points switching their roles. For the remainder of this section therefore, we will assume  $\kappa > 0$ .

The  $\alpha$ -limit set of the orbit of any point in the interior of the cylinder must necessarily be either  $S_-$  or  $N_-$ , and likewise its  $\omega$ -limit set can only be either  $N_+$  or  $S_+$ . The only possible boundary values for  $\Theta(\theta)$  are therefore 0 and  $\pm\pi$  at each endpoint of the interval  $0 \leq \theta \leq \pi$ .

The boundary values (123) correspond to a heteroclinic orbit connecting the two saddles  $S_-$  and  $S_+$ . Suppose such a saddles connection exists. Since the eigendirection corresponding to the unstable manifold of  $S_-$  is  $v_1 = (\kappa + \frac{1}{2}, \lambda - a)^T$  and the stable manifold of  $S_+$  has the same eigendirection, it follows that for the said saddles connection, we have

$$\frac{d\Theta}{d\theta}|_{\theta=0} = \frac{d\Theta}{d\theta}|_{\theta=\pi} = \frac{\lambda - a}{\kappa + \frac{1}{2}} =: \delta < 0. \quad (125)$$

Thus  $\Theta = \delta\theta + o(1)$  as  $\theta \rightarrow 0$  and  $\Theta = -\pi + \delta(\theta - \pi) + o(1)$  as  $\theta \rightarrow \pi$ . Consider now the  $S$  equation (119). By the above,

$$\frac{d}{d\theta} \ln S = \begin{cases} \frac{\kappa}{\theta} + o(1) & \text{as } \theta \rightarrow 0 \\ \frac{\kappa}{\theta - \pi} + o(1) & \text{as } \theta \rightarrow \pi \end{cases} \quad (126)$$

Integrating in  $\theta$  we thereby conclude that  $S \sim |\theta|^\kappa$  for  $\theta$  small and  $S \sim |\pi - \theta|^\kappa$  for  $\theta$  near  $\pi$ . Therefore  $S$  is integrable on  $(0, \pi)$  and indeed it belongs to  $L^p((0, \pi), d\theta)$  for any  $p \geq 1$ .

In an analogous manner one shows that  $S \notin L^2$  when  $\Theta$  takes boundary values corresponding to either of the nodes  $N_-$  or  $N_+$ .

Consider next the  $\Omega$  equation (116). It can also be rewritten as a smooth dynamical system on a cylinder, in this case by setting  $\tau := \frac{r}{a}$  as new independent variable, as well as introducing a new dependent variable

$$\xi := \tan^{-1} \frac{r}{a} = \tan^{-1} \tau \quad (127)$$

Then, with dot again representing differentiation in  $\tau$ , (116) is equivalent to

$$\begin{cases} \dot{\xi} &= \cos^2 \xi \\ \dot{\Omega} &= 2a \sin \xi \cos \Omega + 2\lambda \cos \xi \sin \Omega + 2\gamma \sin \xi \cos \xi + 2\kappa \cos^2 \xi - 2aE \end{cases} \quad (128)$$

Once again, identifying  $\Omega = -\pi$  with  $\Omega = \pi$  turns this into a smooth flow on the closed finite cylinder  $\mathcal{C}_2 := [-\frac{\pi}{2}, \frac{\pi}{2}] \times \mathbb{S}^1$ . The only equilibrium points of the flow are on the two circular boundaries. For  $E \in [0, 1)$  there are two equilibria on each:  $S_- = (-\frac{\pi}{2}, -\pi + \cos^{-1} E)$  and  $N_- = (-\frac{\pi}{2}, \pi - \cos^{-1} E)$  on the left boundary, and  $S_+ = (\frac{\pi}{2}, -\cos^{-1} E)$  and  $N_+ = (\frac{\pi}{2}, \cos^{-1} E)$  on the right boundary.  $S_\mp$  are non-hyperbolic (degenerate) saddle-nodes, with eigenvalues 0 and  $\pm 2a\sqrt{1 - E^2}$ , while  $N_-$  is a degenerate source-node and  $N_+$  a degenerate sink-node.<sup>18</sup>

The boundary values (122) correspond to a heteroclinic orbit connecting  $S_-$  and  $S_+$ . Suppose such a saddles connection exists, and consider the  $R$  equation (117). As  $r \rightarrow \pm\infty$ , we will then have

$$\frac{d}{dr} \ln R \sim -\text{sgn}(r)\sqrt{1 - E^2} \quad (129)$$

so that integrating in  $r$  we will obtain

$$R(r) \sim e^{-|r|\sqrt{1 - E^2}} \quad \text{as } r \rightarrow \pm\infty \quad (130)$$

which ensures that  $R$  is integrable at infinity. Since the right-hand-side of the  $R$  equation is smooth in  $\Omega$  and  $r$ , and  $\Omega$  itself is smooth, it follows that  $R \in L^p(\mathbb{R}, dr)$  for all  $p \geq 1$ .

In an analogous manner one shows that  $R \notin L^2$  when  $\Omega$  takes boundary values corresponding to either of the nodes  $N_-$  or  $N_+$ .  $\square$

<sup>18</sup>For  $E = 1$  each  $S, N$  pair coalesces into one degenerate equilibrium:  $N_-^1 = (-\frac{\pi}{2}, \pm\pi)$  and  $N_+^1 = (\frac{\pi}{2}, 0)$  with both eigenvalues being zero.

### 7.3 Existence of heteroclinic orbits connecting the two saddles

From the proof of Proposition 7.2 it is evident that in order to establish the existence of an eigenfunction for the Dirac Hamiltonian of a point electron in the zGKN spacetime, we need to show that there exists a pair  $(E, \lambda)$  such that both dynamical systems (124) and (128) have a *saddle-saddle connecting orbit* for those values of  $E$  and  $\lambda$ . We call this type of orbit a *saddles connector* for the corresponding flow. We pave the road for our proof by recalling some general facts of flow on a cylinder.

#### 7.3.1 Flow on a finite cylinder

Let  $\mathcal{C} := [x_-, x_+] \times \mathbb{S}^1$  be a finite cylinder. We denote its universal cover by  $\bar{\mathcal{C}} := [x_-, x_+] \times \mathbb{R}$ , with coordinates  $(x, y)$ , and fix a fundamental domain  $\tilde{\mathcal{C}} := [x_-, x_+] \times [-\pi, \pi)$  in  $\bar{\mathcal{C}}$ . Consider the flow  $\Phi_t$  on  $\bar{\mathcal{C}}$  given by the dynamical system

$$\begin{cases} \dot{x} &= f(x) \\ \dot{y} &= g(x, y) \end{cases} \quad (131)$$

where the dot represents differentiation with respect to a formal “time” parameter  $\tau$ , the functions  $f$  and  $g$  are smooth, and  $g$  is  $2\pi$ -periodic in  $y$ :  $g(x, y) = g(x, y + 2\pi)$ . Let us moreover assume that  $f$  satisfies

$$f(x_-) = f(x_+) = 0, \quad f(x) > 0 \quad \forall x \in (x_-, x_+) \quad (132)$$

while  $g$  satisfies

$$g(x_-, y) = 0 \implies y \in \{n_-, s_-\}, \quad g(x_+, y) = 0 \implies y \in \{n_+, s_+\} \quad (133)$$

where  $-\pi \leq s_- < n_- \leq \pi$  and  $-\pi \leq s_+ < n_+ \leq \pi$ . These assumptions imply that the following four distinct points in  $\mathcal{C}$  are equilibrium points for the flow:

$$N_{\pm} := (x_{\pm}, n_{\pm}), \quad S_{\pm} := (x_{\pm}, s_{\pm}). \quad (134)$$

We shall further assume that the flow does not have any non-wandering points other than the above four equilibria.

The following assumptions fix the character of the four equilibrium points:

$$f'(x_-) \geq 0, \quad f'(x_+) \leq 0, \quad f''(x_{\pm}) \neq 0 \quad (135)$$

(where by  $f'(x_{\pm})$  we mean the left derivate at  $x_+$  and the right derivative at  $x_-$ ), and

$$D_y g(x_-, n_-) > 0, \quad D_y g(x_-, s_-) < 0, \quad D_y g(x_+, n_+) < 0, \quad D_y g(x_+, s_+) > 0, \quad (136)$$

where  $D_y g$  is the  $y$ -derivative of  $g(x, y)$ . Thus  $N_-$  is a (source) node,  $N_+$  a (sink) node, and  $S_{\pm}$  are saddle points. These will be hyperbolic if  $f'(x_{\pm}) \neq 0$ , and non-hyperbolic (degenerate) otherwise.

Later on, in order to have a well-defined notion of index for certain distinguished orbits on  $\mathcal{C}$ , we will also assume that the locations of the equilibria on the boundary of the cylinder are not arbitrary, but are subject to the single condition

$$s_- - n_- = n_+ - s_+ \pmod{2\pi} \quad (137)$$

(This is a condition on  $g(x, y)$ . Although we will not pursue this approach here, under this condition (131) can be viewed as a flow with two equilibrium points on a *2-torus*.)

For a point  $p \in \mathcal{C}$ , let  $\mathcal{O}(p)$  denote the flow orbit through  $p$ . Since  $\mathcal{C}$  is compact, all orbits are complete, meaning they exist for all  $s \in \mathbb{R}$ , and since the flow is autonomous, two orbits are either disjoint or they coincide. The orbit of an equilibrium point consists of only one point, namely the equilibrium itself. The  $\omega$ -limit of any other orbit in  $\mathcal{C}$  can be either  $N_+$  or  $S_+$ , and the  $\alpha$ -limit likewise can only be either  $N_-$  or  $S_-$ . All these facts are easy consequences of the existence and uniqueness theorem for ODEs.

### 7.3.2 Connecting orbits and corridors

Given a flow on  $\mathcal{C}$  as in the above, there are two distinguished orbits in the interior of the cylinder: Let  $\mathcal{W}^-$  denote the unique orbit of the flow whose  $\alpha$ -limit point is the saddle  $S_-$ , and let  $\mathcal{W}^+$  denote the unique orbit whose  $\omega$ -limit point is the saddle  $S_+$ . In the hyperbolic case ( $f'(x_{\pm}) \neq 0$ ) the uniqueness is immediate because  $\mathcal{W}^-$  is the unstable manifold of  $S_-$  and  $\mathcal{W}^+$  is the stable manifold of  $S_+$ . In the non-hyperbolic case  $f'(x_{\pm}) = 0$  the orbits  $\mathcal{W}^{\pm}$  are *center* manifolds for the corresponding saddle-nodes  $S_{\pm}$ . Recall that center manifolds may be non-unique, but in our case the uniqueness is assured because the equilibrium points are on the boundary of the domain, so the relevant part of the center manifolds are on the “saddle side” of the equilibrium, and not on the “node side” (see Figure 2). If  $\mathcal{W}^+$  and  $\mathcal{W}^-$  intersect, they must coincide, and the resulting

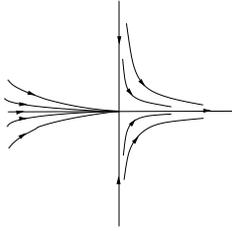


Figure 2: Flow near a saddle-node. The node part lies outside of the domain of concern.

orbit will connect the two saddle points, i.e. it will be the saddles connector we are after. Let us therefore assume that they are disjoint. The  $\omega$ -limit of  $\mathcal{W}^-$  must then necessarily be  $N_+$ , and the  $\alpha$ -limit of  $\mathcal{W}^+$  must be  $N_-$ .

On the other hand the assumptions we have made about the flow imply that there are also two orbits of the flow on the left boundary of the cylinder connecting  $N_-$  with  $S_-$ , call them  $(N_-S_-)_{\pm}$ , with  $+$  denoting the counterclockwise one (when viewed from a point on the cylinder’s axis and to the left of the cylinder), and similarly two joining  $S_+$  with  $N_+$ , called  $(S_+N_+)_{\pm}$ . Consider therefore the following collection of six heteroclinic orbits

$$\mathcal{H} := \{(N_-S_-)_{\pm}, \mathcal{W}^{\pm}, (S_+N_+)_{\pm}\}. \quad (138)$$

The cylinder  $\mathcal{C}$  is divided into two invariant regions  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , called *corridors*, by these orbits:  $\mathcal{C} = \mathcal{K}_1 \cup \mathcal{H} \cup \mathcal{K}_2$ . We would like to distinguish one of these two corridors. We do so as follows: Consider the lifting of the flow to the universal cover  $\tilde{\mathcal{C}}$ . Let  $\tilde{S}_-$  denote the unique copy of the node  $S_-$  that lies in the fundamental domain  $\tilde{\mathcal{C}}$ , and let  $\tilde{\mathcal{W}}^-$  be the unique orbit in  $\tilde{\mathcal{C}}$  whose  $\alpha$ -limit point is  $\tilde{S}_-$ . The  $\omega$ -limit point of this orbit is thus some copy of the node  $N_+$ , call it  $\tilde{N}_+$ , which has coordinates  $(x_+, n_+ - 2\pi k_+)$  for some  $k_+ \in \mathbb{Z}$ . Similarly, let  $\tilde{S}_+$  denote the unique point in the preimage of  $S_+$  under the covering map that lies in the fundamental domain  $\tilde{\mathcal{C}}$  and let  $\tilde{\mathcal{W}}^+$  denote the unique orbit whose  $\omega$ -limit point is this  $\tilde{S}_+$ . Let  $\tilde{N}_- = (x_-, n_- + 2\pi k_-)$ ,  $k_- \in \mathbb{Z}$  be the  $\alpha$ -limit

point of  $\widetilde{\mathcal{W}}^+$ . By definition the corridor  $\mathcal{K}_1$  is the open domain in  $\bar{\mathcal{C}}$  whose boundary contains the two orbits  $\widetilde{\mathcal{W}}^-$  and  $\widetilde{\mathcal{W}}^+$ .

We note that in  $\bar{\mathcal{C}}$  only one of the two corridors will have *both* of these orbits on its boundary, so this is the distinguishing feature of  $\mathcal{K}_1$ .

We orient the boundary of  $\mathcal{K}_1$  (which is a closed simple curve) in such a way that the orientation induced on  $\widetilde{\mathcal{W}}^-$  coincides with the direction of the flow on that orbit.

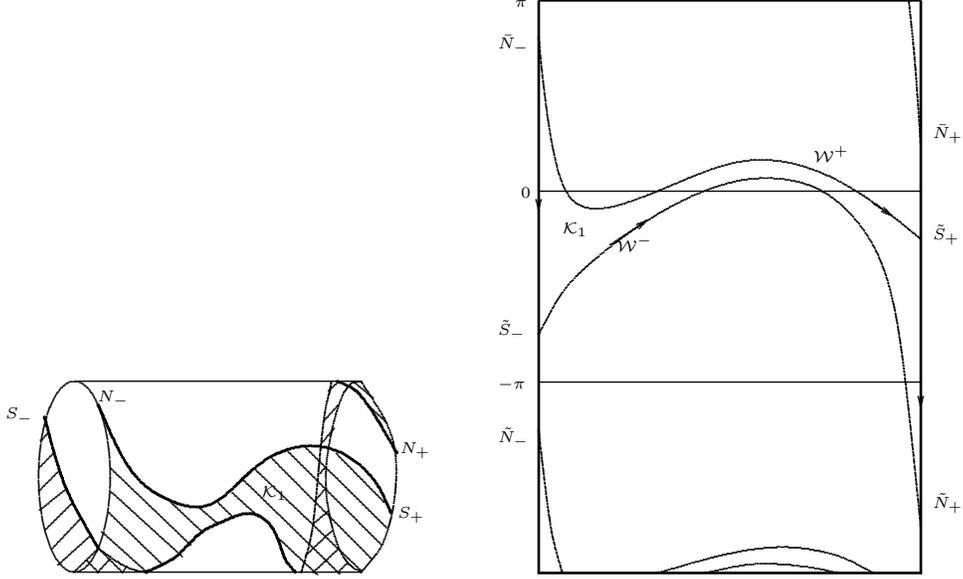


Figure 3: The corridor  $\mathcal{K}_1$  in  $\mathcal{C}$  and in  $\bar{\mathcal{C}}$ .

Furthermore, in the universal cover  $\bar{\mathcal{C}}$ , using the well-orderedness of  $\mathbb{R}$  it is possible to speak of  $\widetilde{\mathcal{W}}^+$  as being situated “above” or “below”  $\widetilde{\mathcal{W}}^-$ . It is evident that the boundary of the corridor  $\mathcal{K}_1$  is oriented clockwise if  $\widetilde{\mathcal{W}}^+$  is below  $\widetilde{\mathcal{W}}^-$ , and counterclockwise if it is the other way around. see Figure 3.

### 7.3.3 Parameter-dependent flows

Suppose that the function  $g$  in (131) depends smoothly on a parameter  $\mu \in I$  where  $I$  is an open interval in  $\mathbb{R}$  (It’s enough for the  $\mu$ -dependence to be  $C^1$ ). Thus the flow is now

$$\begin{cases} \dot{x} &= f(x) \\ \dot{y} &= g_\mu(x, y). \end{cases} \quad (139)$$

By the implicit function theorem, the locations of the equilibria  $S_\pm, N_\pm$  also depend –in a  $C^1$  fashion– on  $\mu$ , so long as the non-degeneracy conditions (136) are satisfied.

We need to make certain assumptions about the  $\mu$ -dependence of the flow regarding its monotonicity, and the topology of its nullclines:

**Monotonicity** We will only consider parameter-dependent flows for which the function  $g_\mu(x, y)$  is *monotone non-increasing* in  $\mu$ .

**Assumption (M):** For all  $(x, y) \in \bar{\mathcal{C}}$ , we have

$$\frac{\partial}{\partial \mu} g_\mu(x, y) \leq 0. \quad (140)$$

This assumption in particular implies a corresponding monotonicity for the distinguished orbits of the flow (139):

**LEMMA 7.3.** *Let  $\tilde{\mathcal{W}}_\mu^\pm = \{(x(\tau), y_\mu^\pm(\tau))\}_{\tau \in \mathbb{R}}$  be the distinguished orbits of (139). Then  $y_\mu^\pm$  are monotone in  $\mu$ , i.e.*

$$\mu_1 < \mu_2 \implies y_{\mu_1}^-(\tau) \geq y_{\mu_2}^-(\tau), \quad y_{\mu_1}^+(\tau) \leq y_{\mu_2}^+(\tau) \text{ for all } \tau \in \mathbb{R}. \quad (141)$$

*Proof.* Let

$$z(\tau) := \frac{\partial}{\partial \mu} y_\mu^-(\tau).$$

Then  $z$  satisfies the ODE

$$\dot{z} = \frac{\partial g_\mu}{\partial y}(x, y)z + \frac{\partial g_\mu}{\partial \mu}(x, y). \quad (142)$$

By the Implicit Function Theorem and the hypotheses above,

$$z(-\infty) = \frac{ds^-(\mu)}{d\mu} = - \left. \frac{\partial g / \partial y}{\partial g / \partial \mu} \right|_{x=x_-, y=s_-} < 0$$

and

$$\dot{z} - \frac{\partial g_\mu}{\partial y}(x, y)z \leq 0. \quad (143)$$

Let

$$U(\tau_0, \tau) := \exp \left( - \int_{\tau_0}^{\tau} \frac{\partial g_\mu}{\partial y}(x(\tau'), y(\tau')) d\tau' \right) \geq 0.$$

Integrating (143) on  $[\tau_0, \tau]$  we then obtain

$$U(\tau_0, \tau)z(\tau) \leq z(\tau_0),$$

and taking the limit  $\tau_0 \rightarrow -\infty$  we conclude  $z(\tau) \leq 0$  for all  $\tau \in \mathbb{R}$ . Therefore  $y_\mu^-$  is monotone non-increasing in  $\mu$ . The proof of monotonicity of  $y_\mu^+$  is completely analogous.  $\square$

**Topology of nullclines** Consider the subset of the cylinder  $\mathcal{C}$  defined by

$$\Gamma := \{(x, y) \in \mathcal{C} \mid g_\mu(x, y) = 0\}. \quad (144)$$

Thus  $\Gamma$  is the zero level-set of  $g_\mu$ . Since  $g_\mu$  is smooth,  $\Gamma$  is a curve (or collection of curves) in  $\mathcal{C}$ . These curves are referred to as the *y-nullclines* of the flow (131). They have the property that any orbit of the flow that crosses them must have a horizontal tangent at the crossing point. Moreover,  $\Gamma$  divides  $\mathcal{C}$  into two regions, thus  $\mathcal{C} = \Gamma \cup \mathcal{N} \cup \mathcal{P}$ , where

$$\mathcal{N} := \{(x, y) \in \mathcal{C} \mid g_\mu(x, y) < 0\}, \quad \mathcal{P} := \{(x, y) \in \mathcal{C} \mid g_\mu(x, y) > 0\}. \quad (145)$$

Thus the  $y$  coordinate of (the lift to the universal cover of) any orbit must decrease in  $\mathcal{N}$  and must increase in  $\mathcal{P}$ .

Evidently,  $\Gamma$  must include all the singular points of the flow:  $S_\pm \in \Gamma$ ,  $N_\pm \in \Gamma$ .

It may happen that as  $\mu$  crosses a critical value  $\mu_c \in I$ , the topology of the nullclines undergoes a dramatic change. This is indeed the case for the flows that we are studying in this paper. We introduce an assumption that amounts to having some control on this change in nullcline topology:

**Assumption (A):** There exists a  $\mu_c \in I$  such that  $\Gamma$ , the zero level-set of  $g_{\mu_c}$  has a *saddle point* at some interior point of  $\mathcal{C}$ . In particular, for  $\mu < \mu_c$ , the sets  $\mathcal{N}$  and  $\mathcal{P}$  are both connected, and  $\Gamma$  is the union of two disjoint curves  $\Gamma = \Gamma_u \cup \Gamma_d$ , with  $N_-, S_+ \in \overline{\Gamma_d}$  and  $S_-, N_+ \in \overline{\Gamma_u}$ . On the other hand for  $\mu > \mu_c$ ,  $\mathcal{N}$  is connected, while  $\mathcal{P}$  has two connected components  $\mathcal{P}_l$  and  $\mathcal{P}_r$ , each being a convex subset of  $\mathcal{C}$ . Moreover,  $\Gamma$  is the union of two disjoint curves  $\Gamma = \Gamma_l \cup \Gamma_r$ , with  $N_-, S_- \in \overline{\Gamma_l}$  and  $S_+, N_+ \in \overline{\Gamma_r}$ ; see Figure 4.

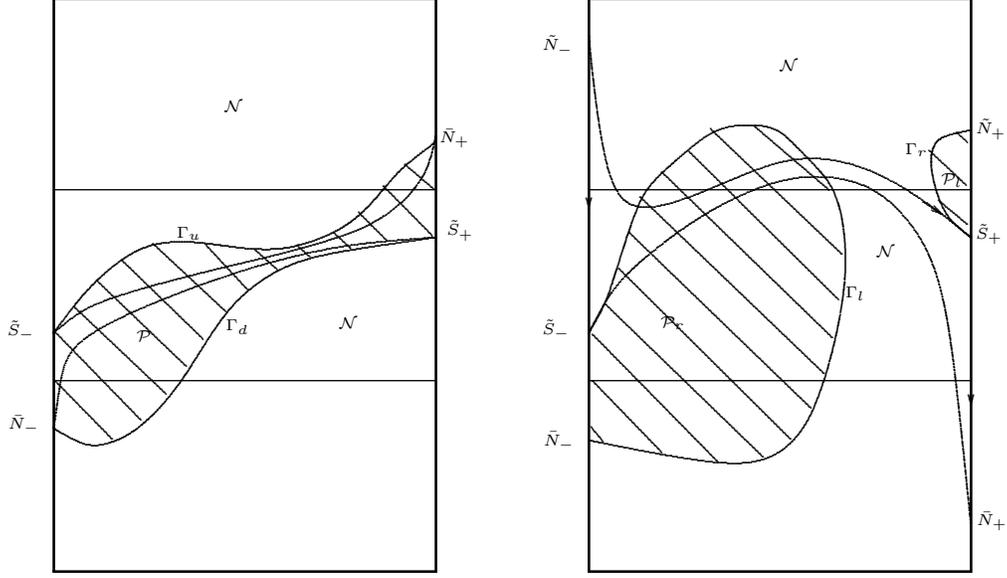


Figure 4: Change in the topology of nullclines:  $\mu < \mu_c$  (left) and  $\mu > \mu_c$  (right).

### 7.3.4 Winding number of orbits and corridors

Assuming (137), let

$$w_0 := \frac{s_+ - n_-}{2\pi} = \frac{n_+ - s_-}{2\pi}. \quad (146)$$

Given any orbit  $\mathcal{O} = \{(x_o(\tau), y_o(\tau)) \mid \tau \in \mathbb{R}\}$  of the flow (131), the following quantity is well-defined:

$$w(\mathcal{O}) = w_0 - \frac{1}{2\pi} (y_o(\infty) - y_o(-\infty)) = w_0 - \frac{1}{2\pi} \int_{-\infty}^{\infty} g_\mu(x_o(\tau), y_o(\tau)) d\tau \quad (147)$$

In particular for the two distinguished orbits  $\widetilde{\mathcal{W}}^\pm$  this is easily calculated to be

$$w(\widetilde{\mathcal{W}}^-) = w_0 - \frac{1}{2\pi} (n_+ - 2\pi k_+ - s_-) = k_+ \quad w(\widetilde{\mathcal{W}}^+) = w_0 - \frac{1}{2\pi} (s_+ - (n_- + 2\pi k_-)) = k_- \quad (148)$$

and is therefore an integer. We call these the *winding numbers*, of  $\mathcal{W}^-$  and  $\mathcal{W}^+$ , respectively.

On the other hand it is easy to see that these two must in fact be equal. The reason is that, since  $\widetilde{\mathcal{W}}^-$  goes from  $\widetilde{S}^- = (x_-, s_-)$  to  $\widetilde{N}^+ = (x_+, n_+ - 2\pi k_+)$ , there are two copies of it in the universal cover that go from  $(x_-, s_- + 2\pi k_+)$  to  $(x_+, n_+)$  and from  $(x_-, s_- + 2\pi(k_+ - 1))$  to  $(x_+, n_+ - 2\pi)$ . Since  $n_+ - 2\pi < s_+ < n_+$ , these two copies of  $\widetilde{\mathcal{W}}^-$  must sandwich any orbit that goes into  $\widetilde{S}_+$ , in particular  $\widetilde{\mathcal{W}}^+$ . Thus the  $\alpha$ -limit point of  $\widetilde{\mathcal{W}}^+$  has to be  $(x_-, n_- + 2\pi k_+)$ , and therefore  $k_+ = k_-$ .

**Definition 7.4.** The *winding number of the corridor*  $\mathcal{K}_1$  is the common value of the winding numbers of  $\widetilde{\mathcal{W}}^\pm$ .

Figure 4 shows a corridor of winding number zero when  $\mu < \mu_c$  (left) and one of winding number equal to one for  $\mu > \mu_c$  (right).

### 7.3.5 Continuity argument for existence of saddles connectors

As the parameter  $\mu$  varies, the two distinguished orbits  $\widetilde{\mathcal{W}}^\pm$ , and hence the corridor  $\mathcal{K}_1$  that they form will also vary. Let  $\mathcal{K}_1(\mu)$  denote the corresponding corridor for parameter value  $\mu$  (if it exists). Since the winding number  $w(\mathcal{K}_1(\mu))$  is integer-valued, if  $\mathcal{K}_1$  varies continuously with respect to  $\mu$ , its winding number would have to remain constant. It is however possible that for some value of  $\mu$  the two orbits  $\widetilde{\mathcal{W}}^\pm$  coincide and the corridor  $\mathcal{K}_1(\mu)$  disappears, leaving a saddles connector behind (for which the winding number will not be an integer). We would like to show that this is the only way for the winding number of  $\mathcal{K}_1(\mu)$  to be different for two different values of the parameter  $\mu$ . In other words:

**PROPOSITION 7.5.** *Suppose there exist two values  $\mu_0 < \mu_1$  in the interval  $I$  for each of which a non-empty corridor  $\mathcal{K}_1$  of finite winding number exists, and such that*

$$w(\mathcal{K}_1(\mu_0)) \leq 0 \quad \text{and} \quad w(\mathcal{K}_1(\mu_1)) \geq 1. \quad (149)$$

*Then there exists  $\mu_s \in (\mu_0, \mu_1)$  such that the flow (139) with  $\mu = \mu_s$  has a saddles connector  $\mathcal{S}(\mu_s)$ , whose lift to the universal cover  $\bar{\mathcal{C}}$  connects the saddle-node  $\tilde{S}_-$  to the saddle-node  $\tilde{S}_+$ .*

*Proof.* Let  $\mathbf{a}(\mu)$  denote the signed area of  $\mathcal{K}_1(\mu)$ , defined via Green's theorem:

$$\mathbf{a}(\mu) = \oint_{\partial\mathcal{K}_1(\mu)} (-y)dx = \int_{x_-}^{x_+} (y_\mu^+ - y_\mu^-) dx. \quad (150)$$

Here  $y_\mu^\pm$  are the  $y$ -components of the orbits  $\widetilde{\mathcal{W}}^\pm$ , thought of as functions of  $x$  (which is always possible since  $x(\tau)$  is a monotone increasing function of flow parameter  $\tau$ .) The following facts about  $\mathbf{a}(\mu)$  are easily verified:

1.  $\mathbf{a}(\mu) > 0$  if and only if  $w(\mathcal{K}_1(\mu)) \geq 1$ .

*Proof:* Since  $\widetilde{\mathcal{W}}^+$  and  $\widetilde{\mathcal{W}}^-$  cannot intersect without coinciding, it is clear from (150) that  $\mathbf{a}(\mu) > 0$  if and only if  $\widetilde{\mathcal{W}}^+$  is above  $\widetilde{\mathcal{W}}^-$ , which is equivalent to  $\tilde{S}_+ = (x_+, s_+)$  being above  $\tilde{N}_+ = (x_+, n_+ - 2\pi k)$  where  $k = w(\mathcal{K}_1(\mu))$ . Since  $n_+ > s_+$  we must have  $k \geq 1$ .

2.  $\mathbf{a}(\mu) < 0$  if and only if  $w(\mathcal{K}_1(\mu)) \leq 0$ .

*Proof:* Similar to above.

3.  $\mathbf{a}(\mu) = 0$  if and only if  $\mathcal{K}_1(\mu) = \emptyset$ .

*Proof:* From the definition (150) it is clear that the only way for  $\mathbf{a}(\mu)$  to be zero is for the two orbits  $\widetilde{\mathcal{W}}^+$  and  $\widetilde{\mathcal{W}}^-$  to coincide.

4.  $\mathbf{a}(\mu)$  is a continuous function of  $\mu$  for all  $\mu \in [\mu_0, \mu_1]$ .

*Proof:* Let  $\epsilon > 0$  be given. For  $\mu, \mu' \in [\mu_0, \mu_1]$  and  $\xi > 0$  small enough, we have

$$\begin{aligned} |\mathbf{a}(\mu') - \mathbf{a}(\mu)| &= \left| \int_{x_-}^{x_+} (y_{\mu'}^+ - y_\mu^+) - (y_{\mu'}^- - y_\mu^-) dx \right| \\ &\leq \int_{x_-}^{x_- + \xi} |y_{\mu'}^+| + |y_\mu^+| dx + \int_{x_- + \xi}^{x_+ - \xi} |y_{\mu'}^+ - y_\mu^+| dx + \int_{x_+ - \xi}^{x_+} |y_{\mu'}^+| + |y_\mu^+| dx \\ &\quad + \int_{x_-}^{x_- + \xi} |y_{\mu'}^-| + |y_\mu^-| dx + \int_{x_- + \xi}^{x_+ - \xi} |y_{\mu'}^- - y_\mu^-| dx + \int_{x_+ - \xi}^{x_+} |y_{\mu'}^-| + |y_\mu^-| dx \\ &= \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}. \end{aligned}$$

By Lemma 7.3 the functions  $y_\mu^\pm$  are monotone in  $\mu$ , thus for all  $\tau \in \mathbb{R}$  we have

$$|y_\mu^\pm(\tau)| \leq \max\{|y_{\mu_0}^\pm(\tau)|, |y_{\mu_1}^\pm(\tau)|\}$$

Thus in particular, using the continuity of orbits of (139) in the flow parameter  $\tau$ , for  $x$  near the boundary points  $x_\pm$  and all  $\mu \in [\mu_0, \mu_1]$  we have  $|y_\mu^\pm(x)| \leq C$ , where  $C > 0$  is a constant depending only on  $y_{\mu_0}^\pm(x_\pm)$  and  $y_{\mu_1}^\pm(x_\pm)$ , or in other words, on the finite winding numbers of the corridors  $\mathcal{K}_1(\mu_0)$  and  $\mathcal{K}_1(\mu_1)$ . Thus, given  $\epsilon$ , we may choose  $\xi$  small enough (depending on  $\epsilon$ ) such that I, III, IV and VI are all less than  $\epsilon/6$ .

Fixing  $\xi$  in this way, by continuous dependence of orbits of the flow (139) on the parameter  $\mu$  (and since we are on a compact interval in the flow parameter  $\tau$ ), it is possible for  $\delta > 0$  to be chosen small enough (depending on  $\epsilon$  and  $\xi$ ), such that  $|\mu' - \mu| < \delta$  implies  $|y_{\mu'}^+(x) - y_\mu^+(x)| < (x_+ - x_-)^{-1}\epsilon/6$  for all  $x \in [x_- + \xi, x_+ - \xi]$ . Therefore II  $< \epsilon/6$ . Similarly, V  $< \epsilon/6$  as well, and we are done.

Having established the above properties for the signed area  $\mathbf{a}(\mu)$ , the standard continuity argument can now be applied: By assumption we have  $\mathbf{a}(\mu_0) < 0$  and  $\mathbf{a}(\mu_1) > 0$ . Thus by the Intermediate Value Theorem there exists  $\mu_s \in (\mu_0, \mu_1)$  such that  $\mathbf{a}(\mu_s) = 0$ , which is equivalent to the existence of a connector. Since the corridors  $\mathcal{K}_1(\mu)$  by definition always have  $\tilde{S}_\pm$  on their boundary, the saddles connector also has to go from  $\tilde{S}_-$  to  $\tilde{S}_+$ .  $\square$

We will use this proposition to establish the existence of saddles connectors for both the  $\Theta$  (124) and the  $\Omega$  (128) equations. In each case we need to show that the flow satisfies the assumptions we have made in this subsection about the flow on a cylinder (139).

### 7.3.6 Existence of saddles connectors for the $\Theta$ equation

For simplicity we are only going to consider the case  $\kappa = \frac{1}{2}$ . To see that (124) is a flow on a cylinder of the type we have considered in the above, we make the following identifications:  $x = \theta$ ,  $y = \Theta$ ,  $x_- = 0$ ,  $x_+ = \pi$ . We have

$$f(\theta) = \sin \theta, \quad g(\theta, \Theta) = -2a \sin \theta \cos \theta \cos \Theta + (2aE \sin^2 \theta - 1) \sin \Theta + 2\lambda \sin \theta.$$

Therefore, the equilibria are at

$$S_- = (0, 0), \quad N_- = (0, \pi) \quad S_+ = (\pi, -\pi), \quad N_+ = (\pi, 0);$$

all of these are hyperbolic. Condition (137) is clearly satisfied:  $s_- - n_- = -\pi$  and  $n_+ - s_+ = \pi$ .

Here we make a note of the fact, easily verified, that the  $\Theta$  flow (124) possesses a discrete symmetry: Let  $\mathcal{O} = (\theta(\tau), \Theta(\tau))$  be any orbit of the flow (124). Then  $\mathcal{O}' = (\pi - \theta(-t), \pi - \Theta(-t))$  is also an orbit of (124). This is simply due to the fact that

$$f(\theta) = f(\pi - \theta), \quad g(\theta, \Theta) = g(\pi - \theta, \pi - \Theta).$$

This symmetry will prove useful in constructing corridors of given winding number for the flow. Next we check that the assumptions we made in the previous subsection about the topology of nullclines hold in the case of (124).

### 7.3.7 Topology of the nullclines

Let  $T := \tan(\Theta/2)$ . Then,

$$g_{E,\lambda}(\theta, \Theta) = \frac{1}{1+T^2} (2 \sin \theta (\lambda - a \cos \theta) T^2 + 2(2aE \sin^2 \theta - 1)T + 2 \sin \theta (\lambda + a \cos \theta)) =: \frac{q(T)}{1+T^2}$$

where  $q$  is a quadratic polynomial in  $T$ , whose discriminant we calculate to be

$$\Delta_q := (2aE \sin^2 \theta - 1)^2 - 4\lambda^2 \sin^2 \theta + 4a^2 \sin^2 \theta \cos^2 \theta.$$

Setting  $\tau := -\cot \theta$  we thus obtain

$$\Delta_q = \frac{\delta_q(\tau)}{(1+\tau^2)^2}, \quad \delta_q(\tau) = \tau^4 + 2(1 - 2aE + 2a^2 - 2\lambda^2)\tau^2 + (1 - 2aE)^2 - 4\lambda^2$$

Therefore  $\delta_q$  is a quartic polynomial which is quadratic in  $s = \tau^2$ . The discriminant of  $\delta_q(s)$  is

$$\Delta_{\delta_q} = \lambda^4 + 2a(E - a)\lambda^2 + a^2(a^2 - 2aE + 1) = (\lambda^2 + a(E - a))^2 + a^2(1 - E^2)$$

which is always positive. Thus  $\delta_q(s) = 0$  will always have two roots, which will be of opposite signs if  $|\lambda| \geq \frac{1}{2} - aE$ . In that case there exists  $s_1 = \tau_1^2 > 0$  such that  $\Delta_q(\pm\tau_1) = 0$ , and  $\Delta_q(\tau) < 0$  for  $\tau \in (-\tau_1, \tau_1)$ . It follows that the quadratic equation  $q(T) = 0$  will have two roots so long as  $|t| > \tau_1$ , it will have repeated roots when  $t = \pm\tau_1$ , and no roots when  $|t| < \tau_1$ .

If on the other hand  $|\lambda| < \frac{1}{2} - aE \leq \frac{1}{2}$  then  $s_1$  and  $s_2$ , the roots of  $\delta_q(s) = 0$ , will be of the same sign. Since the sum of these roots will be  $4(\lambda^2 - a^2 - (\frac{1}{2} - aE))$  it is easy to see that  $s_1$  and  $s_2$  will both be negative, thus they will not correspond to real values of  $\tau = \pm\sqrt{s}$ . Therefore  $\Delta_q > 0$  and hence  $q(T) = 0$  will always have two roots. Thus the critical value of the parameter  $\lambda$  (thinking of the other parameters  $a$  and  $E$  as given and fixed) is

$$\lambda_c := -\frac{1}{2} + aE.$$

Now any zero of  $q$  will be a zero of  $g$ , and thus will give us a point on the nullcline  $\Gamma$ . In addition,  $g$  may also have a zero at  $\Theta = 0$  or  $\pm\pi$ , where  $T = \pm\infty$ . For  $g$  to be zero there we need the coefficient of  $T^2$  in  $q$  to vanish, i.e. either  $\sin \theta = 0$ , which will give us the equilibrium points, or  $\cos \theta = \lambda/a$  which is impossible so long as

$$\lambda < -a.$$

Under this condition therefore, the nullclines have the topology we assumed in the previous subsection, with  $\lambda$  playing the role of the parameter  $\mu$ , i.e., given  $a \in (0, \frac{1}{2})$ ,  $E \in [0, 1]$ ,  $\lambda \in (-\infty, -a)$  and  $\lambda_c$  as in the above, the topology of  $\Theta$ -nullclines for the flow (124) changes across  $\lambda = \lambda_c$  in the manner described in assumption **(A)**. Figure 5 shows Maple plots of the  $\Theta$ -nullclines for values of  $\lambda$  below and above the critical value.

### 7.3.8 Explicit solutions of the $\Theta$ equation

One easily verifies that given  $a \in [0, \frac{1}{2})$ , there is an explicit solution of (118) (with  $\kappa = \frac{1}{2}$ ), for  $E = 1$  and  $\lambda = -1 + a$ , namely

$$\Theta_0(\theta) = -\theta.$$

This furthermore generates a saddles connector for (124):  $\mathcal{S}_0 := (\theta_0(\tau), \Theta_0(\theta_0(\tau)))$  where  $\theta_0(\tau)$  is the unique solution to the ODE  $\dot{\theta} = \sin \theta$  with  $\theta(-\infty) = 0$  and  $\theta(\infty) = \pi$ .

This solution will help us get the iteration started.

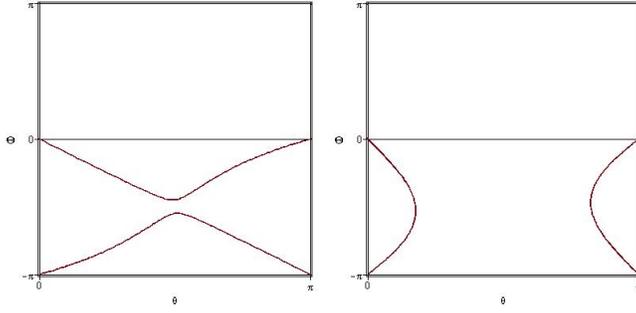


Figure 5:  $\Theta$  nullclines for  $\lambda = -0.4$  (left) and  $\lambda = -0.9$  (right), with  $a = 0.1$ ,  $E = 0.95$ .

### 7.3.9 Existence of corridors with unequal winding number

Throughout this section,  $a$  will be a fixed number in  $(0, \frac{1}{2})$ . The following two propositions will start things off:

**PROPOSITION 7.6.** *Given  $E \in [0, 1]$  and  $\lambda \leq \lambda_l := -1 - a$  the flow (124) has a corridor  $\mathcal{K}_1(E, \lambda)$  with  $w(\mathcal{K}_1) \geq 1$ .*

*Proof.* The linearization of the flow at  $S_-$  gives us eigenvalues  $\lambda_1 = 1$ ,  $\lambda_2 = -1$  and a corresponding set of eigenvectors is  $v_1 = \begin{pmatrix} 1 \\ \lambda - a \end{pmatrix}$  and  $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . The orbit  $\mathcal{W}^-$ , being the unstable manifold of  $S_-$  is tangent to the unstable direction  $v_1$ , therefore the slope of  $\mathcal{W}^-$  at  $S_-$  is  $\frac{d\Theta}{d\theta}|_{\mathcal{W}^-(S_-)} = \lambda - a < 0$ . The slope of  $\mathcal{W}^+$  at  $S_+$  is similarly calculated, and it turns out to be the same  $\frac{d\Theta}{d\theta}|_{\mathcal{W}^+(S_+)} = \lambda - a$ . On the other hand, since  $\lambda_l < \lambda_c$ , we have  $\Gamma = \Gamma_l \cup \Gamma_r$ .  $S_-$  is a terminal

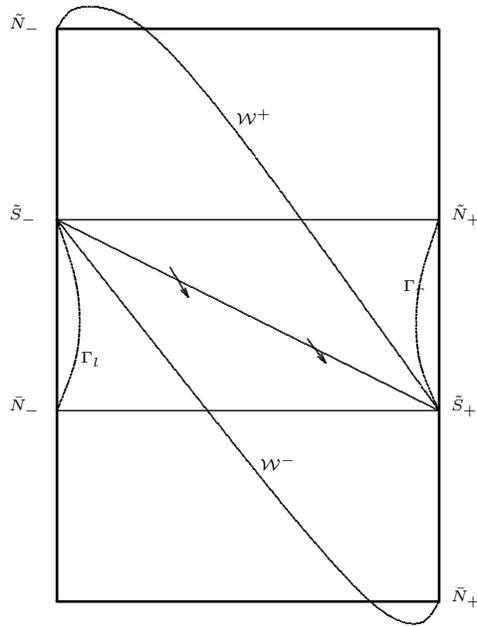


Figure 6: Existence of corridor with  $w(\mathcal{K}_1) \geq 1$

point of the curve  $\Gamma_l$ , on which  $g_{E,\lambda}(\theta, \Theta) = 0$ , hence the slope of  $\Gamma_l$  at  $S_-$  can be calculated from

implicit function theorem to be

$$\left. \frac{d\Theta}{d\theta} \right|_{\Gamma_l} (S_-) = -\frac{\partial_\theta g_{E,\lambda}}{\partial_\Theta g_{E,\lambda}}(S_-) = 2(\lambda - a) < \lambda - a.$$

Same is true for the slope of  $\Gamma_r$  at  $S_+$ , as can be easily verified. Thus  $\mathcal{W}^-$  starts off inside  $\mathcal{N}$ , which is connected, and its initial slope is  $\lambda - a < -1$  for  $\lambda < -1 + a$ . Consider the diagonal line segment  $S_-S_+$ , on which  $\Theta = -\theta$ . Let us compute the slope of orbits that cross this line, and compare it to the slope of the line:

$$\frac{g_{E,\lambda}(\theta, -\theta)}{f(\theta)} - (-1) = -2a \sin \theta (\cos \theta + E \sin \theta) + 2\lambda + 2 \leq 0 \quad \text{for } \lambda \leq -1 - a.$$

Thus  $S_-S_+$  acts as a ‘‘barrier’’, not allowing  $\mathcal{W}^-$  to cross it from below to above. Hence the  $\omega$ -limit of  $\mathcal{W}^-$  cannot be  $(\pi, 0)$ . The first possible terminal point is then  $(\pi, -2\pi)$ , hence  $w(\mathcal{W}^-) \geq 1$ .  $\square$

**PROPOSITION 7.7.** *Suppose that for some  $E \in (0, 1]$  and some  $\lambda \leq \lambda_0 := -1 + a$ , the flow (124) has a saddles connector  $\mathcal{S}^\Theta(E, \lambda)$  whose lift to the universal cover of the cylinder goes from  $\tilde{S}_- = (0, 0)$  to  $\tilde{S}_+ = (\pi, -\pi)$ . Then, for all  $E' \in [0, E)$ , there exists a corridor  $\mathcal{K}_1(E', \lambda)$  of winding number  $w(\mathcal{K}_1) = 0$  for (124).*

*Proof.* Let  $\mathcal{S}(E, \lambda) = (\theta(\tau), \Theta_{E,\lambda}(\tau))$ . Since  $\mathcal{S}^\Theta = \mathcal{W}^- = \mathcal{W}^+$ , by the calculation done in the proof of the previous Proposition, the graph of  $\Theta_{E,\lambda}$  is entirely contained in  $\mathcal{N}$ , thus it has to be monotone decreasing, and thus  $\sin \Theta_{E,\lambda}(\tau) \leq 0$  for all  $\tau$ .

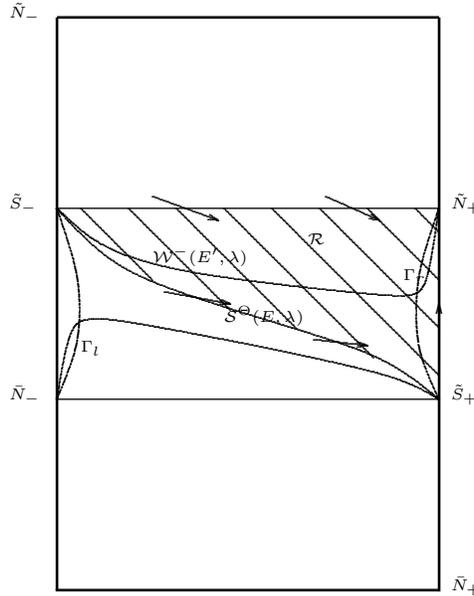


Figure 7: Trapped region  $\mathcal{R}$  and the corridor with zero winding number.

Consider the region  $\mathcal{R}$  in  $\bar{\mathcal{C}}$  whose boundary consists of the following three curves: (i) From  $S_-$  to  $S_+$  along  $\Theta_{E,\lambda}$ . (ii) From  $S_+$  to  $N_+$  along the right boundary, and (iii) from  $N_+$  to  $S_-$  along a horizontal line segment. We claim that  $\mathcal{R}$  is a trapped region for (124) at parameter values  $(E', \lambda)$  provided  $E' < E$ . Since (ii) is always an orbit, and (iii) is entirely contained in  $\mathcal{N}$ , this only needs to be checked for (i). We have

$$g_{E',\lambda}(\theta, \Theta_{E,\lambda}) = g_{E,\lambda}(\theta, \Theta_{E,\lambda}) + 2a(E' - E) \sin^2 \theta \sin \Theta_{E,\lambda} \geq \dot{\Theta}_{E,\lambda}$$

which shows that the new flow crosses the old solution from left to right. Thus  $\mathcal{R}$  is trapped and since  $\mathcal{W}^-$  starts in  $\mathcal{R}$ , it must terminate in  $\tilde{N}_+ = (\pi, 0)$ , so that  $w(\mathcal{W}^-) = 0$ .  $\square$

Setting  $E_0 = 1$  and  $\lambda_0 = -1 + a$ , let  $\mathcal{S}_0 := \mathcal{S}(E_0, \lambda_0)$  denote the explicit solution found in 7.3.8. For all  $E_1 \in (0, 1)$ , the above two propositions, together with the following immediate corollary of Proposition 7.5, establish the existence of a saddles connector  $\mathcal{S}_1 := \mathcal{S}(E_1, \lambda_1)$  for (124), for some  $\lambda_1 \in (\lambda_l, \lambda_0)$ :

**COROLLARY 7.8.** *Let  $E \in [0, 1]$  be fixed. Suppose that there exists  $\lambda_1 < \lambda_2 < 0$  such that the flow (124) has corridors  $\mathcal{K}_1(E, \lambda_1)$  and  $\mathcal{K}_1(E, \lambda_2)$  with  $w(\mathcal{K}_1(E, \lambda_2)) = 0$  and  $w(\mathcal{K}_1(E, \lambda_1)) \geq 1$ . Then there is a  $\lambda \in (\lambda_1, \lambda_2)$  such that (124) has a saddles connector  $\mathcal{S}^\Theta(E, \lambda)$  going from  $(0, 0)$  to  $(\pi, -\pi)$ .*

*Proof.* Proposition 7.5 applies, with  $-\lambda$  playing the role of the parameter  $\mu$ .  $\square$

Proceeding iteratively, suppose that given  $E_n \in [0, 1]$  a saddles connector  $\mathcal{S}_n^\Theta = \mathcal{S}^\Theta(E_n, \lambda_n)$  has been found for (124), for some  $\lambda_n < \lambda_0$ . In the next subsection we shall see how the newly-found  $\lambda_n$  can be used to prove the existence of a saddles connector for the  $\Omega$  flow (128), namely  $\mathcal{S}_n^\Omega := \mathcal{S}_n^\Omega(E_{n+1}, \lambda_n)$  for some  $E_{n+1} \in (0, 1)$ . Coming back to the  $\Theta$  flow then, a new saddles connector  $\mathcal{S}_{n+1}^\Theta$  needs to be found with the updated energy  $E_{n+1}$ , given that a saddles connector  $\mathcal{S}_n^\Theta(E_n, \lambda_n)$  already exists. Since  $E_{n+1}$  can be on either side of  $E_n$ , in addition to Prop. 7.7 we also need the following:

**THEOREM 7.9.** *Given any  $a \in (0, \frac{1}{2})$  and  $E \in [0, 1]$ , there exists a unique*

$$\lambda = \Lambda(E) \in [-1 - a, -1 + a]$$

*such that (124) has a saddles connector  $\mathcal{S}(E, \lambda)$  going from  $(0, 0)$  to  $(\pi, -\pi)$ . Moreover,  $\Lambda$  is an increasing  $C^1$  function, and  $\frac{\partial \Lambda}{\partial E} < a$ .*

*Proof.* If  $E = 1$  then  $\lambda = -1 + a$  works and  $\mathcal{S} = \mathcal{S}_0 = (\theta, -\theta)$ . For  $E < 1$  existence is guaranteed by Prop. 7.7, Prop. 7.6, and Corollary 7.8. To prove uniqueness, suppose that for a given  $E$ , there are two saddles connectors  $\mathcal{S}(E, \lambda)$  and  $\mathcal{S}'(E, \lambda')$ , with  $\lambda' > \lambda$ . Let  $\Theta_{E, \lambda}$  and  $\Theta_{E, \lambda'}$  denote the corresponding  $\Theta$ -components of  $\mathcal{S}$  and  $\mathcal{S}'$ , respectively. For  $\theta \in (0, \pi)$ ,

$$g_{E, \lambda'}(\theta, \Theta_{E, \lambda}) = g_{E, \lambda}(\theta, \Theta_{E, \lambda}) + 2(\lambda' - \lambda) \sin \theta > \dot{\Theta}_{E, \lambda}.$$

Thus orbits of the  $(E, \lambda')$  flow can only cross  $\mathcal{S}$  from below to above. On the other hand, since  $\mathcal{S}'$  is a saddles connector, near  $S_-$  it coincides with  $\mathcal{W}^-(E, \lambda')$ , and near  $S_+$  it coincides with  $\mathcal{W}^+$ . Thus from the linearization of the flow at  $S_\pm$ ,

$$\left. \frac{d\Theta_{E, \lambda'}}{d\theta} \right|_{S_\pm} = \lambda' - a > \lambda - a.$$

Therefore  $\mathcal{S}'$  must be above  $\mathcal{S}$  near  $S_-$  and below it near  $S_+$ , so  $\mathcal{S}'$  would have to cross  $\mathcal{S}$  from above to below, which is a contradiction, unless they coincide.

Given  $E$  then, let  $\Lambda(E)$  denote the unique value of  $\lambda$  for which a saddles connector  $\mathcal{S}(E, \lambda)$  exists. The fact that  $\Lambda$  is continuously differentiable (in fact analytic), and the bound on the derivative, have already been shown in [49] and [3] using analytic perturbation theory. Here we give a simple proof of monotonicity of  $\Lambda$  which also establishes the bound on the derivative:

Given  $E \in [0, 1]$  let  $\lambda = \Lambda(E)$  and let  $(\theta(\tau), \Theta_E(\tau))$  denote the unique (modulo translations in  $\tau$ ) saddles connector for (124) whose existence we have established. Let  $u := \frac{\partial \Theta_E}{\partial E}$ . By differentiating the  $\Theta$  equation in (124) with respect to  $E$  we obtain an equation for  $u$ :

$$\frac{du}{d\tau} = P(\tau)u + Q(\tau), \quad \begin{cases} P & := 2a \sin \theta \cos \theta \sin \Theta_E + (2aE \sin^2 \theta - 1) \cos \Theta_E \\ Q & := 2a \sin^2 \theta \sin \Theta_E + 2 \frac{d\Lambda}{dE} \sin \theta. \end{cases} \quad (151)$$

Let

$$U(\tau_1, \tau_2) := e^{-\int_{\tau_1}^{\tau_2} P(\tau) d\tau}.$$

Thus we have

$$U(\tau_2, \tau_1) = \frac{1}{U(\tau_1, \tau_2)}, \quad U(\tau_1, \tau_2)U(\tau_2, \tau_3) = U(\tau_1, \tau_3).$$

Solving the first-order linear ODE (151) for  $u$  we obtain

$$u(\tau) = U(\tau, \tau_1)u(\tau_1) + \int_{\tau_1}^{\tau} U(\tau, \tau')Q(\tau')d\tau'. \quad (152)$$

Note that  $P$  and  $Q$  are bounded functions of  $\tau$  and

$$\lim_{\tau \rightarrow -\infty} P(\tau) = -1, \quad \lim_{\tau \rightarrow \infty} P(\tau) = 1.$$

Therefore, for any fixed  $\tau$ ,

$$U(\tau, \tau_1) \rightarrow 0 \text{ as } \tau_1 \rightarrow -\infty, \quad U(\tau, \tau_1) \rightarrow 0 \text{ as } \tau_1 \rightarrow \infty.$$

Moreover  $u(\pm\infty) = 0$ . For any finite  $\tau$  from (152) we thus obtain two equivalent expressions for  $u(\tau)$ . For example, setting  $\tau = 0$ ,

$$u(0) = \int_{-\infty}^0 U(0, \tau')Q(\tau')d\tau' = - \int_0^{\infty} U(0, \tau')d\tau'.$$

Thus in particular

$$0 = \int_{-\infty}^{\infty} U(0, \tau')Q(\tau')d\tau' = 2a \int_{\mathbb{R}} U(0, \tau') \sin^2 \theta(\tau') \sin \Theta_E(\tau') d\tau' + 2 \frac{d\Lambda}{dE} \int_{\mathbb{R}} U(0, \tau') \sin \theta(\tau') d\tau'$$

Therefore

$$\frac{d\Lambda}{dE} = a \frac{\int_{\mathbb{R}} U(0, \tau') \sin^2 \theta(\tau') (-\sin \Theta_E(\tau')) d\tau'}{\int_{\mathbb{R}} U(0, \tau') \sin \theta(\tau') d\tau'} \geq 0.$$

We have already shown that  $\Theta_E(\tau) \in (-\pi, 0)$  and  $\theta(\tau) \in (0, \pi)$  for all  $\tau$ . Thus the numerator in the above fraction is strictly less than the denominator, hence

$$0 \leq \frac{\partial \Lambda}{\partial E} < a \quad (153)$$

□

### 7.3.10 Existence of saddles connectors for the $\Omega$ equation

We now show that the flow (128) also satisfies all the hypotheses we had made about flows on a cylinder. Once again, for simplicity we are only going to consider the case  $\kappa = \frac{1}{2}$ . The situation is somewhat more complicated than what we have done in the above for the  $\Theta$  equation, due to the presence of an extra parameter, namely  $\gamma$ , which breaks the symmetry that was present for the  $\Theta$  flow, as well as the fact that the equilibria of the  $\Omega$  flow are degenerate (non-hyperbolic).

Let us make the identifications  $x = \xi$  and  $y = \Omega$ . Thus  $x_- = -\pi/2$ ,  $x_+ = \pi/2$ , and we now have

$$f(\xi) = \cos^2 \xi, \quad g_{E,\lambda}(\xi, \Omega) = 2a \sin \xi \cos \Omega + 2\lambda \cos \xi \sin \Omega + 2\gamma \sin \xi \cos \xi + \cos^2 \xi - 2aE$$

Therefore, for  $E \in [0, 1)$ ,

$$s_- = -\pi + \cos^{-1}(E), \quad n_- = \pi - \cos^{-1}(E), \quad s_+ = -\cos^{-1}(E), \quad n_+ = \cos^{-1}(E),$$

where by  $\cos^{-1}$  we mean the principal branch of the arccosine,  $0 \leq \cos^{-1} x \leq \pi$ , and

$$S_{\pm} = (\pm \frac{\pi}{2}, s_{\pm}), \quad N_{\pm} = (\pm \frac{\pi}{2}, n_{\pm})$$

as before. We note that this time, all the equilibria are non-hyperbolic, since  $f'(\pm\pi/2) = 0$ , and that for  $\gamma = 0$ , there is a discrete symmetry:  $f(-\xi) = f(\xi)$  and  $g|_{\gamma=0}(-\xi, \pi - \Omega) = g|_{\gamma=0}(\xi, \Omega)$ , which is broken when  $\gamma$  is turned on.

Also in the case  $E = 1$  there is a further degeneracy: the two equilibria on each side coalesce into one singular point with both eigenvalues equal to zero. For these type of singular points center manifolds can be non-unique, so that the distinguished orbits  $\mathcal{W}^{\pm}$  and the index theory we have developed for the corridor they form, are not directly relevant to this case.

We now check the hypotheses about the topology of the nullclines.

### 7.3.11 Topology of the null-clines

Let  $T := \tan(\Omega/2)$ . Then

$$g_{E,\lambda}(\xi, \Omega) = \frac{2q(T)}{1 + T^2},$$

where

$$q(T) := (\gamma \sin \xi \cos \xi + \frac{1}{2} \cos^2 \xi - aE - a \sin \xi) T^2 + 2\lambda \cos \xi T + (\gamma \sin \xi \cos \xi + \frac{1}{2} \cos^2 \xi - aE + a \sin \xi).$$

Thus  $q$  is a quadratic polynomial in  $T$  with coefficients that are functions of  $\xi$ . The discriminant of  $q$  is

$$\Delta_q(\xi) := \lambda^2 \cos^2 \xi - (\gamma \sin \xi \cos \xi + \frac{1}{2} \cos^2 \xi - aE)^2 + 4a^2 \sin^2 \xi.$$

Let  $\tau := \tan \xi$ . Thus  $-\infty < \tau < \infty$ , and

$$\Delta_q(\tau) = \frac{p(\tau)}{(1 + \tau^2)^2}$$

with

$$p(\tau) := a^2(1 - E^2)\tau^4 + 2\gamma aE\tau^3 + (\lambda^2 - \gamma^2 + a^2 + 2a(\frac{1}{2} - aE))\tau^2 - 2\gamma(\frac{1}{2} - aE)\tau + \lambda^2 - (\frac{1}{2} - aE)^2.$$

Since  $|E| < 1$ ,  $p$  is an irreducible quartic in  $\tau$ . Let us write it as  $p(\tau) = \sum_{j=0}^4 c_j \tau^j$ . Suppose that  $E \in (0, 1)$  and

$$a \in (0, \frac{1}{2}), \quad \gamma \in (-\sqrt{2a(1-2a)}, 0), \quad \lambda \in [-1-a, -1+a] \quad (154)$$

It then follows that

$$c_4 > 0, \quad c_3 < 0, \quad c_2 > 0, \quad c_1 > 0, \quad c_0 > 0.$$

By Descartes' Rule of Signs, then,  $p(\tau)$  has either two or no real positive roots, and either two or no real negative roots, counting multiplicity. For a more accurate root count, one needs to use the discriminant of the quartic. Since the discriminant theory for general quartics is somewhat complicated, here we opt for a simpler analysis by estimating  $p$  from above and below with two *reducible* quartics. To this end, first we note that  $p(\tau) = Q(\tau) - (q_1(\tau))^2$  with

$$Q(\tau) = \lambda^2 + (\lambda^2 + a^2)\tau^2 + a^2\tau^4, \quad q_1(\tau) = \frac{1}{2} - aE + \gamma\tau - aE\tau^2.$$

Thus on the one hand, by completing the square,

$$Q(\tau) = \left(a\tau^2 + \frac{\lambda^2 + a^2}{2a}\right)^2 - \frac{(\lambda^2 - a^2)^2}{4a^2} \leq (q_2^+(\tau))^2, \quad q_2^+(\tau) := a\tau^2 + \frac{\lambda^2 + a^2}{2a}$$

and on the other hand,

$$Q(\tau) \geq \lambda^2 + 2|\lambda|a\tau^2 + a^2\tau^4 = (q_2^-(\tau))^2, \quad q_2^-(\tau) = a\tau^2 + |\lambda|$$

Thus we have upper and lower bounds for  $p$  in terms of factorizable quartics  $Q^\pm(\tau)$ :

$$Q^-(\tau) := (q_2^-(\tau))^2 - (q_1(\tau))^2 \leq p(\tau) \leq (q_2^+(\tau))^2 - (q_1(\tau))^2 =: Q^+(\tau)$$

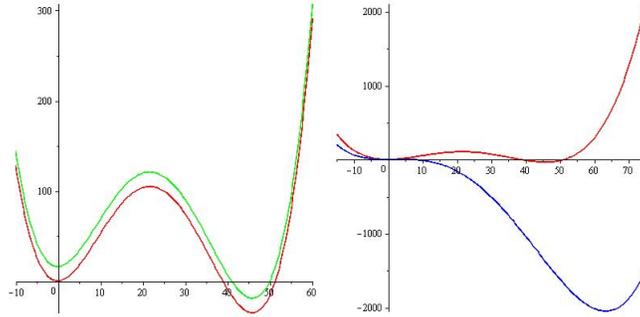


Figure 8: The bounding quartics:  $Q^+$  and  $p$  (left),  $p$  and  $Q^-$  (right) for parameter values  $a = 0.1$ ,  $\gamma = -0.2$ ,  $\lambda = -0.9$ ,  $E = 0.978$ .

Consider first the upper quartic  $Q^+$ . We have  $Q^+ = (q_2^+ + q_1)(q_2^+ - q_1) = q^+q^-$  where  $q^\pm$  are two quadratic polynomials. It is clear that if  $q^+$  has any real roots, they will be positive, and if  $q^-$  has any roots, they will be negative (recall that we are assuming  $\gamma < 0$ ). The discriminants of  $q^\pm$  are computed to be

$$\Delta_\pm := \gamma^2 - 2(1 \mp E) (\lambda^2 + a^2 \pm a(1 - 2aE)).$$

Similarly,  $Q^- = (q_2^- + q_1)(q_2^- - q_1) = \tilde{q}^+ \tilde{q}^-$  where  $\tilde{q}^\pm$  are two quadratics, and once again, any real roots of  $\tilde{q}^+$  must be positive and any real root of  $\tilde{q}^-$ , negative. The discriminants of  $\tilde{q}^\pm$  are

$$\tilde{\Delta}_\pm = \gamma^2 - 2(1 \mp E)(-2a\lambda \pm a(1 - 2aE)) \geq \Delta_\pm$$

It thus follows that there are two subsets of the  $(a, \gamma, \lambda)$  parameter space (154) that are of interest: (R1) where both  $\tilde{\Delta}_+$  and  $\tilde{\Delta}_-$  are negative; and (R2) where  $\tilde{\Delta}_+ > 0$  and  $\tilde{\Delta}_- < 0$ . For parameter values in the region (R1) the quartic  $Q^-$  will have no real zeros, and will be always positive, while for those in (R2) the quartic  $Q^+$  will have exactly two positive roots and no negative root.

Let us fix  $a, \gamma, \lambda$  as in (154). We find that the range (R1) corresponds to  $0 \leq E < E_l$ , where

$$E_l(\lambda) := \frac{1}{2a} \left[ -\lambda + a + \frac{1}{2} - \sqrt{(\lambda + a - \frac{1}{2})^2 + \gamma^2} \right],$$

while range (R2) corresponds to  $E_h < E \leq 1$ , with

$$E_h(\lambda) = \frac{1}{4a^2} \left[ \lambda^2 + 3a^2 + a - \sqrt{(\lambda^2 - a^2 + a)^2 + 4a^2\gamma^2} \right]$$

Note that  $0 < E_l < E_h < 1$  for all values of  $a, \gamma, \lambda$  as in (154).

For  $E \in (E_h, 1]$ , therefore, since  $Q^+$  has two positive roots, the quartic  $p(\tau)$  must also have at least two roots, one of which will definitely be positive. Thus by the Rule of Signs,  $p(\tau)$  has exactly two positive roots. We call them  $\tau_1$  and  $\tau_2$ , and  $p(\tau) < 0$  for  $\tau_1 < \tau < \tau_2$ . It follows that the quadratic  $q(T)$  will have two roots for  $\tau \notin [\tau_1, \tau_2]$ , double roots at  $\tau_1$  and at  $\tau_2$ , and no real roots for  $\tau \in (\tau_1, \tau_2)$ .

For  $E \in [0, E_l)$  since  $Q^-$  has no real roots,  $p$  cannot have any either. Thus  $p$  is always positive and  $q(T) = 0$  will have two roots for all  $\tau \in \mathbb{R}$ .

Combining these two, one concludes that a critical value for the energy  $E = E_c(a, \gamma, \lambda)$  exists,  $E_c \in [E_l, E_h]$ , such that assumption **(A)** is satisfied, with the role of parameter  $\mu$  played by  $E$ .

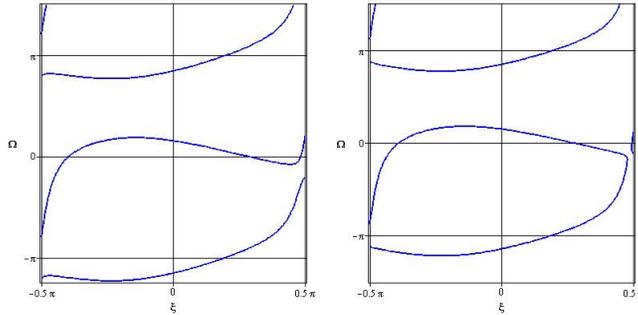


Figure 9:  $\Omega$ -nullclines for parameter values  $E = 0.8$  (left) and  $E = 0.93$  (right), with  $a = 0.1$ ,  $\gamma = -0.4$ , and  $\lambda = -0.9$ .

### 7.3.12 Existence of corridors with unequal winding number

Throughout this section,  $a$  will be a fixed number in  $(0, \frac{1}{2})$  and  $\gamma$  a fixed number in  $(-\sqrt{2a(1-2a)}, 0)$ . The following two propositions help us get started:

**PROPOSITION 7.10.** *Given  $\lambda \leq -1 + a$  there exists  $\bar{E} \in (E_h(\lambda), 1)$  such that for all  $E \in [\bar{E}, 1)$  the corridor  $\mathcal{K}_1(E, \lambda)$  of the flow (128) has winding number greater than or equal to one.*

*Proof.* We compute the slope of solution orbits that cross the following line in  $\bar{\mathcal{C}}$

$$L := \{(\xi, \Omega) \in \bar{\mathcal{C}} \mid -\frac{\pi}{2} \leq \xi \leq \frac{\pi}{2}, \quad \Omega = \frac{\pi}{2} - \cos^{-1} E - \xi\}$$

and compare it to the slope of  $L$ . Note that  $L$  passes through  $\tilde{N}_-$  and  $\tilde{S}_+$ . We have

$$\frac{g_{E,\lambda}(\xi, \Omega)}{f(\xi)} - (-1) = 2 \left( 1 - (a - \lambda)E + [\sqrt{1 - E^2}(a - \lambda) + \gamma] \tan \xi \right) \quad (155)$$

Consider first the case  $\lambda = -1 + a$ . Let

$$\xi_0(E) := \tan^{-1} \frac{1 - E}{-\gamma - \sqrt{1 - E^2}} \searrow 0 \text{ as } E \nearrow 1.$$

For  $\xi \geq \xi_0(E)$ , the slope of any orbit of the flow crossing  $L$  is less than (i.e. more negative than) the slope of  $L$ . Hence on the portion of  $L$  where  $\xi \geq \xi_0$  orbits can only cross  $L$  from above to below.

On the other hand, suppose  $\lambda < -1 + a$ . Let

$$E_m := \max \left\{ \frac{1}{a - \lambda}, \sqrt{1 - \frac{\gamma^2}{(a - \lambda)^2}}, E_h(\lambda) \right\} \in (E_h, 1)$$

For  $E \in (E_m, 1)$  we have  $\xi_0(E) \in (-\frac{\pi}{2}, 0)$ . Let

$$\eta_0 := \begin{cases} \cos^{-1} E & \lambda < -1 + a \\ \cos^{-1} E + \xi_0(E) & \lambda = -1 + a \end{cases}$$

We note that  $\eta_0 \rightarrow 0$  as  $E \rightarrow 1$ . Let us consider the horizontal line

$$L' := \{(\xi, \Omega) \in \bar{\mathcal{C}} \mid -\frac{\pi}{2} \leq \xi \leq \frac{\pi}{2}, \quad \Omega = \frac{\pi}{2} - \eta_0\}$$

We compute the slope of orbits crossing  $L'$ :

$$h(\xi) := g_{E,\lambda}(\xi, \frac{\pi}{2} - \eta_0) = 2a \sin \xi \sin \eta_0 + 2\lambda \cos \xi \cos \eta_0 + 2\gamma \sin \xi \cos \xi + \cos^2 \xi - 2aE.$$

Clearly

$$\begin{aligned} h(\xi) &\leq 2a(\sin \eta_0 - E) + |\gamma| + 2\lambda \cos \xi \cos \eta_0 + \cos^2 \xi \\ &\leq 2a(\sin \eta_0 - E) + |\gamma| + 2\lambda \cos \eta_0 + 1 \\ &\rightarrow -2a + |\gamma| + 2\lambda + 1 \quad \text{as } E \rightarrow 1 \\ &\leq |\gamma| - 1 < 0 \end{aligned}$$

Thus there exists  $E$  sufficiently close to 1 such that  $h(\xi) < 0$  for  $\xi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ .

Let  $\Upsilon$  be the curve in  $\mathcal{C}$  consisting of two line segments: From  $(-\frac{\pi}{2}, \frac{\pi}{2} - \eta_0)$  along the horizontal line  $L'$ , upto the intersection point of  $L$  and  $L'$ , and then from there along  $L$  to  $\tilde{S}_+$ . The curve  $\Upsilon$  provides a barrier for the flow: no orbit can cross it from the region below  $\Upsilon$  into the region above  $\Upsilon$ . Let  $\tilde{\mathcal{W}}^-$  be the unstable manifold of  $\tilde{S}_-$ . Thus  $\tilde{\mathcal{W}}^-$  must stay below  $\Upsilon$ , and as a result the  $\omega$ -limit of  $\mathcal{W}^-$  cannot be  $\tilde{N}_+$  so therefore its winding number is not zero or negative.  $\square$

**PROPOSITION 7.11.** *Given  $\lambda \leq -1 + a$  and  $E \in [0, E_l(\lambda))$  the corridor  $\mathcal{K}_1(E, \lambda)$  of the flow (128) has winding number equal to zero.*

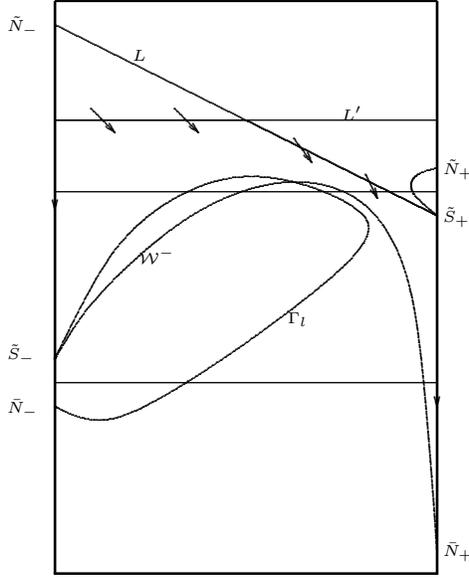


Figure 10: Construction of a barrier

*Proof.* Once again we find a barrier that prevents  $\mathcal{W}^-$  from going down: Let us compute the slope of orbits crossing the line

$$L = \{(\xi, \Omega) \mid \Omega = \xi - \frac{\pi}{2}\}$$

and compare it to the slope of this line.

$$j(\xi) := g_{E,\lambda}(\xi, \xi - \frac{\pi}{2}) - \cos^2 \xi = 2a(1 - E) - 2(\lambda + a) \cos^2 \xi + 2\gamma \sin \xi \cos \xi.$$

Thus  $j(\pm\pi/2) = 2a(1 - E) > 0$ . Any interior minimum of  $j$  must be achieved at a critical point:

$$j'(\xi_0) = 0 \implies \xi_0 = \frac{1}{2} \tan^{-1} \frac{\gamma}{\lambda + a} > 0.$$

However we have

$$j(\xi_0) = 2a(1 - E) - (\lambda + a) - (\lambda + a) \cos 2\xi_0 + \gamma \sin 2\xi_0 = 2a(1 - E) - \frac{2(\lambda + a)^3}{(\lambda + a)^2 + \gamma^2} > 2a(1 - E).$$

Thus  $j(\xi) > 0$  for all  $\xi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ . It follows that the slope of any orbit crossing the line is greater than the slope of the line. Thus orbits cannot cross this line from above to below. In particular, the orbit  $\mathcal{W}^-$  starts at  $\tilde{S}_-$ , which is above this line. Hence  $\mathcal{W}^-$  cannot end at any copy of the node  $N_+$  other than  $\tilde{N}_+$ , so its winding number cannot be positive. Since the region  $\mathcal{N}$  is connected, once  $\mathcal{W}^-$  leaves  $\mathcal{P}$  and enters  $\mathcal{N}$ , its  $\Omega$  must decrease, hence  $\mathcal{W}^-$  cannot end at any copy of the node  $N_+$  that is higher than  $\tilde{N}_+$  either, and therefore the winding number of  $\mathcal{W}^-$  cannot be negative, hence  $w(\mathcal{W}^-) = 0$ .  $\square$

Let  $\lambda_0 = -1 + a$ . The above two propositions, in conjunction with the following immediate corollary of Proposition 7.5, establish the existence a saddles connector  $\mathcal{S}_0^\Omega(E_1, \lambda_0)$  for the flow (128), for some  $E_1 \in (0, 1)$ :

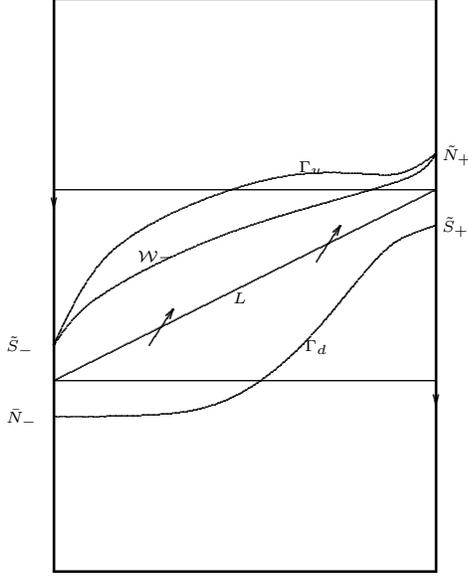


Figure 11: Barrier for zero winding number

**COROLLARY 7.12.** *Let  $\lambda \leq -1+a$  be fixed. Suppose that there exists  $0 \leq E_1 < E_2 < 1$  such that the flow (128) has corridors  $\mathcal{K}_1(E_1, \lambda)$  and  $\mathcal{K}_1(E_2, \lambda)$  with  $w(\mathcal{K}_1(E_1, \lambda)) = 0$  and  $w(\mathcal{K}_1(E_2, \lambda)) \geq 1$ . Then there is an  $E \in (E_1, E_2)$  such that (128) has a saddles connector  $\mathcal{S}(E, \lambda)$ .*

*Proof.* Proposition 7.5 applies, with  $E$  playing the role of the parameter  $\mu$ .  $\square$

Let  $n \geq 1$ . Suppose that given  $\lambda_{n-1} \leq -1+a$  a saddles connector  $\mathcal{S}_{n-1}^\Omega = \mathcal{S}^\Omega(E_n, \lambda_{n-1})$  has been found for (128), for some  $E_n \in (0, 1)$ . In the previous subsection we saw how this newly-found  $E_n$  can be used to prove the existence of a saddles connector for the  $\Theta$  flow (124), namely  $\mathcal{S}_n^\Theta := \mathcal{S}_n^\Theta(E_n, \lambda_n)$  for some  $\lambda_n \leq -1+a$ . Coming back to the  $\Omega$  flow then, given the updated value  $\lambda = \lambda_n$ , a new saddles connector  $\mathcal{S}_n^\Omega$  needs to be found with an updated energy  $E_{n+1}$ , given that a saddles connector  $\mathcal{S}_{n-1}^\Omega(E_n, \lambda_{n-1})$  already exists. Then, for all  $\lambda' \in (\lambda, -1+a)$ , there exists a corridor  $\mathcal{K}_1(E, \lambda')$  of winding number  $w(\mathcal{K}_1) = 1$  for (128). More generally, we have

**THEOREM 7.13.** *Fix  $a \in (0, \frac{1}{2})$  and  $\gamma \in (-\sqrt{2a(1-2a)}, 0)$ . Then given any  $\lambda \in [-1-a, -1+a]$ , there exists a unique*

$$E = \mathcal{E}(\lambda) \in (0, 1)$$

*such that (128) has a saddles connector  $\mathcal{S}^\Omega(E, \lambda)$ . Moreover,  $\mathcal{E}$  is a  $C^1$  function, and  $|\frac{\partial \mathcal{E}}{\partial \lambda}| < \frac{1}{a}$ .*

*Proof.* Existence of a saddles connector is guaranteed by Propositions 7.10 and 7.11, and Corollary 7.12. To see uniqueness, suppose that there exists two saddles connectors  $\mathcal{S}^\Omega(E, \lambda)$  and  $\mathcal{S}^\Omega(E', \lambda)$  for (128) for  $E$  and  $E'$  in  $(0, 1)$ , and suppose  $E < E'$ . Let  $\Omega_{E, \lambda}$  be the  $\Omega$  component of  $\mathcal{S}^\Omega(E, \lambda)$ . We have

$$g_{E', \lambda}(\xi, \Omega_{E, \lambda}) = g_{E, \lambda}(\xi, \Omega_{E, \lambda}) - 2a(E' - E) < \dot{\Omega}_{E, \lambda}.$$

It thus follows that orbits of the  $(E', \lambda)$  flow can only cross  $\mathcal{S}^\Omega(E, \lambda)$  from above to below. On the other hand, since  $E' > E$  the equilibrium point  $\tilde{S}_-(E', \lambda)$  is situated below  $\tilde{S}_-(E, \lambda)$ , while  $\tilde{S}_+(E', \lambda)$  is above  $\tilde{S}_+(E, \lambda)$ . Since  $\mathcal{S}^\Omega(E', \lambda)$  coincides with both  $\mathcal{W}^-(E', \lambda)$  and  $\mathcal{W}^+(E', \lambda)$  it begins below  $\mathcal{S}^\Omega(E, \lambda)$  and it ends above it, which is a contradiction, hence  $E' = E$ .

Given  $\lambda$ , let  $\mathcal{E}(\lambda)$  denote the unique value of  $E$  for which a saddles connector  $\mathcal{S}^\Omega(\mathcal{E}(\lambda), \lambda) = (\xi(\tau), \Omega_{\mathcal{E}(\lambda), \lambda}(\tau))$  exists. We now prove that  $\mathcal{E}$  is a  $C^1$  function: Consider the two initial value problems for  $\Omega^\pm(\tau)$ , the  $\Omega$  components of  $\mathcal{W}^\pm$ :

$$\dot{\Omega}^\pm = g_{E, \lambda}(\xi(\tau), \Omega^\pm), \quad \Omega^-( -\infty) = -\pi + \cos^{-1} E, \quad \Omega^+(\infty) = -\cos^{-1} E$$

By standard ODE theory these two problems have unique smooth solutions  $\Omega_{E, \lambda}^\pm(\tau)$  which also depend smoothly on the parameters  $\lambda$  and  $E$  (so long as  $E < 1$ ) for any finite  $\tau$ .

Next recall that  $\mathcal{W}^-$  is a saddles connector if it coincides with  $\mathcal{W}^+$ , which will be the case if these two orbits intersect at one point, e.g. if  $\Omega_{E, \lambda}^+(0) = \Omega_{E, \lambda}^-(0)$ . Let us define a smooth function

$$\Phi(E, \lambda) := \Omega_{E, \lambda}^+(0) - \Omega_{E, \lambda}^-(0) \quad (156)$$

Let  $\lambda_0 \in [-1 - a, -1 + a]$  be fixed, and set  $E_0 = \mathcal{E}(\lambda_0)$ . Then  $\Phi(E_0, \lambda_0) = 0$ . By the Implicit Function Theorem, if

$$\frac{\partial \Phi}{\partial E}(E_0, \lambda_0) \neq 0, \quad (157)$$

then there is a neighborhood  $\mathcal{I}$  of  $\lambda_0$  and a  $C^1$  function  $\tilde{\mathcal{E}}$  defined on  $\mathcal{I}$  such that  $\Phi(\tilde{\mathcal{E}}(\lambda), \lambda) = 0$  for all  $\lambda \in \mathcal{I}$ . By the uniqueness result we have already shown, we must have  $\tilde{\mathcal{E}} = \mathcal{E}$ . Thus we only need to verify the condition (157).

For  $\lambda \in [-1 - a, -1 + a]$  and  $E \in (0, 1)$ , let

$$u_\pm(\tau) := \frac{\partial}{\partial E} \Omega_{E, \lambda}^\pm(\tau).$$

Then  $u_\pm$  satisfy the linear ODEs

$$\frac{du_\pm}{d\tau} = P_\pm(\tau)u_\pm - 2a, \quad P_\pm(\tau) := -2a \sin \xi(\tau) \sin \Omega_{E, \lambda}^\pm(\tau) + 2\lambda \cos \xi(\tau) \cos \Omega_{E, \lambda}^\pm(\tau) \quad (158)$$

together with the initial conditions

$$u_-(-\infty) = \frac{-1}{\sqrt{1 - E^2}}, \quad u_+(\infty) = \frac{1}{\sqrt{1 - E^2}}.$$

Moreover,

$$\Phi(E, \lambda) = u_+(0) - u_-(0).$$

For  $\tau_1, \tau_2 \in \mathbb{R}$  let

$$U_\pm(\tau_1, \tau_2) := e^{-\int_{\tau_1}^{\tau_2} P_\pm(\tau) d\tau}.$$

Solving the ODEs (158) for  $u_\pm$  we obtain

$$U_\pm(\tau_1, \tau_2)u_\pm(\tau_2) = u_\pm(\tau_1) - 2a \int_{\tau_1}^{\tau_2} U_\pm(\tau_1, \tau) d\tau. \quad (159)$$

Note that

$$\lim_{\tau \rightarrow -\infty} P_-(\tau) = -2a\sqrt{1 - E^2} < 0, \quad \lim_{\tau \rightarrow \infty} P_+(\tau) = 2a\sqrt{1 - E^2} > 0.$$

Thus for any fixed  $\tau$ ,

$$U_-(\tau, \tau_1) \rightarrow 0 \text{ as } \tau_1 \rightarrow -\infty, \quad U_+(\tau, \tau_2) \rightarrow 0 \text{ as } \tau_2 \rightarrow \infty.$$

And so, from (159) we obtain

$$u_-(\tau) = -2a \int_{-\infty}^{\tau} U_-(\tau, \tau') d\tau', \quad u_+(\tau) = 2a \int_{\tau}^{\infty} U_+(\tau, \tau') d\tau'. \quad (160)$$

We know that  $\Omega_{E_0, \lambda_0}^+(\tau) = \Omega_{E_0, \lambda_0}^-(\tau)$  for all  $\tau$ . Hence  $U_+(\tau_1, \tau_2) = U_-(\tau_1, \tau_2) =: U(\tau_1, \tau_2)$  and thus

$$\frac{\partial \Phi}{\partial E}(E_0, \lambda_0) = 2a \int_{-\infty}^{\infty} U(0, \tau') d\tau' > 0$$

so that (157) is clearly satisfied. Since  $\lambda_0$  was arbitrary we have shown that  $\mathcal{E} \in C^1((-1-a, -1+a))$ .

We can furthermore compute the derivative of  $\mathcal{E}$  by implicit differentiation. Let

$$v_{\pm}(\tau) := \frac{\partial}{\partial \lambda} \Omega_{E, \lambda}^{\pm}(\tau).$$

Then  $v_{\pm}$  satisfy

$$\dot{v}_{\pm} = P_{\pm}(\tau)v_{\pm} + 2 \cos \xi(\tau) \sin \Omega_{E, \lambda}^{\pm}(\tau), \quad v_-(-\infty) = 0, \quad v_+(\infty) = 0.$$

Thus by a similar argument to above,

$$v_-(\tau) = \int_{-\infty}^{\tau} U_-(\tau, \tau') \cos \xi(\tau') \sin \Omega_{E, \lambda}^-(\tau') d\tau', \quad v_+(\tau) = \int_{\tau}^{\infty} U_+(\tau, \tau') \cos \xi(\tau') \sin \Omega_{E, \lambda}^+(\tau') d\tau',$$

so that

$$\frac{\partial \Phi}{\partial \lambda}(E_0, \lambda_0) = -2 \int_{-\infty}^{\infty} U(0, \tau') \cos \xi(\tau') \sin \Omega_{E_0, \lambda_0}(\tau') d\tau',$$

and thus

$$\frac{d\mathcal{E}}{d\lambda} = -\frac{\partial \Phi / \partial \lambda}{\partial \Phi / \partial E} = \frac{\int_{-\infty}^{\infty} U(0, \tau) \cos \xi(\tau) \sin \Omega_{E, \lambda}(\tau) d\tau}{a \int_{-\infty}^{\infty} U(0, \tau) d\tau}.$$

Moreover, clearly

$$\left| \int_{-\infty}^{\infty} U(0, \tau) \cos \xi(\tau) \sin \Omega_{E, \lambda}(\tau) d\tau \right| < \int_{-\infty}^{\infty} U(0, \tau) d\tau.$$

so that

$$\left| \frac{d\mathcal{E}}{d\lambda} \right| < \frac{1}{a}.$$

□

### 7.3.13 The iteration argument

**THEOREM 7.14.** *Let  $a \in (0, \frac{1}{2})$  and  $\gamma \in (-\sqrt{2a(1-2a)}, 0)$  be fixed. There exists a  $\lambda \in [-1-a, -1+a]$  and  $E \in (0, 1)$  such that (124) has a saddles connector  $\mathcal{S}^{\Theta}(E, \lambda)$  and (128) has a saddles connector  $\mathcal{S}^{\Omega}(E, \lambda)$ .*

*Proof.* Set  $\lambda_0 = -1 + a$ . For  $n \geq 1$  let

$$E_n := \mathcal{E}(\lambda_{n-1}) \in (0, 1), \quad \lambda_n := \Lambda(E_n) \in [-1-a, -1+a].$$

Thus

$$E_{n+1} = \mathcal{E}(\Lambda(E_n)).$$

By Theorems 7.13 and 7.9 we have

$$|E_{n+1} - E_n| \leq \delta |E_n - E_{n-1}|$$

where

$$\delta := \max_{0 \leq E \leq 1} \left| \frac{d\Lambda}{dE} \right| \max_{-1-a \leq \lambda \leq -1+a} \left| \frac{d\mathcal{E}}{d\lambda} \right| < a \cdot \frac{1}{a} = 1$$

Thus by the contraction mapping theorem, the sequence  $E_n$  converges, and thus so does the sequence  $\lambda_n$ . Let  $\lambda = \lim_{n \rightarrow \infty} \lambda_n \in [-1-a, -1+a]$  and  $E := \lim_{n \rightarrow \infty} E_n$ . We must have  $E = \mathcal{E}(\lambda)$  and thus  $0 < E < 1$ .  $\square$

## 8 Summary and Outlook

We have studied the Dirac equation for a point electron in static, electromagnetic, flat spacetimes with Zipoy topology which include the zero-gravity limit of the electromagnetic Kerr–Newman spacetimes as special case, but which can feature a generalization of the Appell–Sommerfeld electromagnetic fields with any charge  $Q$  and current  $I$  one wants; the zero- $G$  Kerr–Newman spacetimes correspond to  $Q = I\pi a$ . In contrast to similar-spirited studies of the Dirac equation for a point electron on the Kerr–Newman spacetime, which are plagued by the presence of a Cauchy horizon and regions of closed timelike loops, [27, 5, 17, 18, 19, 20, 3, 50, 51, 4], our zero- $G$  spacetimes do not possess any such physically troublesome features. Moreover, by working with the topologically non-trivial maximal analytical extension of zGKN and its electromagnetic fields, our treatment does not encounter physically troublesome problems like infinite charges, currents, and masses, which plague the topologically trivial Minkowski spacetime interpretations [25, 31, 21, 26].<sup>19</sup>

We proved that the spectrum of any self-adjoint extension of the pertinent Dirac Hamiltonian is symmetric about zero; this result holds for any charge  $Q$  and current  $I$  of the generalized zGKN spacetimes.<sup>20</sup> We have also shown that the formal Dirac Hamiltonian on a spacelike slice of the maximal analytically extended static zGKN spacetime is essentially self-adjoint.<sup>21</sup> We also showed that the self-adjoint Dirac operator on the zGKN spacetime has a continuous spectrum with a gap about zero which, under two smallness conditions, contains a pure point spectrum.

Our results are far from exhaustive. In the following we list a number of interesting open problems which we hope will be solved in some future work.

We begin with the Dirac point electron in zGKN spacetimes:

- **Problem 1:** Characterize the point spectrum of the Dirac Hamiltonian on zGKN in complete detail; to the extent possible, compute it analytically, or at least numerically in some representative situations.

**Remark 8.1.** *We suspect that there is a countably infinite set of energy eigenvalues which correspond with pairs of saddles connectors of arbitrary winding numbers  $(w_1, w_2) \in \mathbb{Z}^2$ ; the possibility of such saddles connectors we already established, see Theorem 7.14.*

<sup>19</sup>For  $G > 0$ , the disk singularity of the “single-leafed” truncation of the maximal analytically extended zGKN spacetime also features a negative mass density rotating at superluminal speed.

<sup>20</sup>This result was to be expected, for on the zGKN spacetime the Dirac electron “sees” a charge  $Q$  in one sheet and a charge  $-Q$  in the other, and recalling that the Dirac operator for an electron in the Coulomb potential of a point proton in Minkowski spacetime has a symmetric continuous spectrum, plus a positive point spectrum which maps into a negative mirror image of it under the switch of the coupling constant  $-e^2 \rightarrow e^2$  (by charge conjugation).

<sup>21</sup>We note that by Stone’s theorem there exists a unitary one-parameter group on the Hilbert space, generated by the unique self-adjoint extension of the Hamiltonian, which yields the time-evolution of the Dirac bi-spinors on the spacelike static slice of the zGKN spacetime. Thus the naked ring singularity does not cause any trouble.

- **Problem 2:** Discuss the generalized scattering problem for Dirac spinor fields on the  $zGKN$  spacetimes. In particular, investigate the evolution when (part of) the Dirac spinor field “dives” through the ring from one sheet to the other.

We now come to the Dirac point electron in  $zGK$  spacetimes equipped with a generalization of the Appell–Sommerfeld electromagnetic fields to arbitrary charge  $Q$  and current  $I$ :

- **Problem 3:** For the formal Dirac Hamiltonian on the  $(Q, I)$ -generalization of the  $zGKN$  spacetime (given  $a$ ), show that essential self-adjointness holds if the “coupling constant”  $(Q - I\pi a)e$  is small in magnitude; perhaps using a so-called Hardy–Dirac type estimate.
- **Problem 4:** Suppose essential self-adjointness fails if  $|(Q - I\pi a)e|$  is too large. If so, what is the sharp constant for  $|(Q - I\pi a)e|$ ? Determine the  $(Q, I)$ -parameter regimes (given  $a$ ) in which the Dirac Hamiltonian on a generalization of the  $zGKN$  spacetime with Sommerfeld fields (6) has several self-adjoint extensions, respectively has no self-adjoint extension. Amongst the self-adjoint extensions, can one identify a distinguished one?
- **Problem 5:** In the cases of self-adjointness, characterize the spectrum of the Dirac Hamiltonian in complete detail; to the extent possible, compute the spectrum analytically, or at least numerically in some representative situations. In particular:
  - **Problem 5a:** The continuous spectrum of quantum physical operator families is usually very robust. Show that Theorem 2.10 holds for the  $(Q, I)$  generalization of our Dirac operators, at least as long as  $(Q - I\pi a)e$  is small in magnitude.
  - **Problem 5b:** Same consideration as above, for Theorem 2.11; thus: Can an eigenvalue of  $\hat{H}$  on  $\mathbb{H}$  for  $Q = I\pi a$  be continuously deformed into an eigenvalue of  $\hat{H}$  on  $\mathbb{H}$  for  $Q \neq I\pi a$  as long as  $(Q - I\pi a)e$  is sufficiently small in magnitude? If so, does the size of the neighborhood of  $Q = I\pi a$  into which an eigenvalue can be continued depend on the eigenvalue, or can one have a uniform control on the spectrum w.r.t. the coupling constant  $(Q - I\pi a)e$ ? Is it possible that the point spectrum disappears completely if  $(Q - I\pi a)e$  becomes too large in magnitude?
- **Problem 6:** Same as Problem 2, now for Dirac spinor fields on  $zGK$  equipped with generalizations of the Appell–Sommerfeld fields to arbitrary  $Q$  and  $I$ .

So far our problems concern the Dirac equation on the zero-gravity limit case of the KN spacetimes, and its generalization to arbitrary  $Q$  and  $I$ . To make contact with the existing studies of Dirac’s electron in Kerr–Newman spacetimes, the following bifurcation problem suggests itself:

- **Problem 7:** Deform these zero-gravity spacetimes perturbatively by “switching on”  $G$  and discuss the Dirac equation on them perturbatively as well. In particular, is it possible to perturb into generalizations of the Kerr–Newman spacetime with gyromagnetic ratios amounting to  $g$ -factors  $g \neq g_{KN} = 2$ , or is such a perturbation feasible only if  $g = g_{KN} = 2$ , viz. if  $Q = I\pi a$ ?

In all the above problems, the energy(-density)-momentum(-density)-stress tensor is of classical electromagnetic nature. The Dirac spinor field does not influence the spacetime structure (in the zero-gravity limit the electromagnetic fields do not influence the spacetime structure, either). More to the point, Dirac’s point electron is treated as a *test particle* in this paper and in all the above

problems, which should be a good approximation if the zGKN ring singularity is much more massive and can approximately be treated as “infinitely massive.” Yet, since test particles are only physical fiction, no matter how useful practically, an obvious task is to investigate the Dirac electron *not* as a test particle. Thus:

- **Problem 8:** Treat the quantum-mechanical interaction of the Dirac electron with the ring singularity of the zero- $G$  Kerr–Newman spacetime symmetrically as a two-body problem.

**Remark 8.2.** *Of course, the same problem has not even been completely solved yet for the simpler setting of “Dirac Hydrogen” in Minkowski spacetime; i.e., how to go beyond the traditional textbook problem where the relativistic Hydrogen problem is treated by solving the Dirac equation of a point electron in flat Minkowski spacetime containing an infinitely massive positive point charge (representing the proton). Traditionally the problem of finite mass of the proton (or nucleus, more generally) is addressed by perturbation theory, starting from Pauli’s two-body equation and adding “relativistic corrections” in powers of  $1/c$ ; or by perturbative QED-type calculations. The constrained relativistic two-body approach of Bethe–Salpeter [6] is perhaps the closest one has come to solving this problem. In a similar vein one may be able to take a finite “ADM mass” of the ring singularity into account.*

- **Problem 9:** Same as Problem 8, but now with the  $(Q, I)$  generalizations of zGKN.
- **Problem 10:** Same as Problems 8 and 9, but now perturbatively for  $G > 0$ .
- **Problem 11:** Compute the feedback of the Dirac spinor field onto the spacetime structure perturbatively when  $G > 0$ . This amounts to the perturbative discussion of the so-called Einstein–Maxwell–Dirac system for small  $G$ . In this problem the energy-momentum-stress tensor, in addition to classical electromagnetic fields, also involves the Dirac spinor field which now influences the spacetime structure.

All these problems are difficult, and the amount of work needed to solve all of them can only be handled by the involvement of many mathematical physicists. In this vein we hope that our paper inspires some readers to join us in our pursuit. We ourselves have made some progress on problems 1, 3, 8, and 9, which we plan to report in forthcoming publications.

**Acknowledgement:** We thank Friedrich Hehl for his comments on an earlier version of this paper and for bringing several references to our attention. We are also thankful to the anonymous referee for constructive comments.

## References

- [1] F. E. Alzofon. *Two methods for the exact solution of diffraction problems*, volume 127. SPIE Press, 2004.
- [2] P. Appell. Quelques remarques sur la théorie des potentiels multiformes. *Math. Ann.*, 30:155–156, 1887.
- [3] D. Batic, H. Schmid, and M. Winklmeier. On the eigenvalues of the Chandrasekhar-Page angular equation. *J. Math. Phys.*, 46(1):012504, 35, 2005.
- [4] F. Belgiorno and S. L. Cacciatori. The Dirac equation in Kerr–Newman–AdS black hole background. *J. Math. Phys.*, 51(3):033517, 32, 2010.

- [5] F. Belgiorno and M. Martellini. Quantum properties of the electron field in Kerr-Newman black hole manifolds. *Physics Letters B*, 453(12):17 – 22, 1999.
- [6] H.A. Bethe and E.E. Salpeter. *Quantum Mechanics of One- and Two-Electron Atoms*. Plenum Press, 1977.
- [7] R. H. Boyer and R. W. Lindquist. Maximal analytic extension of the Kerr metric. *J. Math. Phys.*, 8(2):265–281, 1967.
- [8] D. R. Brill and J. M. Cohen. Cartan frames and the general relativistic Dirac equation. *J. Math. Phys.*, 7(2):238–243, 1966.
- [9] D. Canarutto and A. Jadczyk. Fundamental geometric structures for the dirac equation in general relativity. *Acta Appl. Math.*, 51(1):59–92, 1998.
- [10] B. Carter. Global structure of the Kerr family of gravitational fields. *Phys. Rev.*, 174:1559–1571, 1968.
- [11] B. Carter and R. G. McLenaghan. Generalized master equations for wave equation separation in a Kerr or Kerr-Newman black hole background. In R. Ruffini, editor, *Proceedings of the second Marcel Grossmann meeting on general relativity*, pages 575–585. North-Holland Publishing Company, 1982.
- [12] S. Chandrasekhar. Errata: “The solution of Dirac’s equation in Kerr geometry”. *Proc. Roy. Soc. London Ser. A*, 350(1663):565, 1976.
- [13] S. Chandrasekhar. The solution of Dirac’s equation in Kerr geometry. *Proc. Roy. Soc. London Ser. A*, 349(1659):571–575, 1976.
- [14] S. Chandrasekhar. *The Mathematical Theory of Black Holes*. Oxford University Press, New York, 1983.
- [15] P. R. Chernoff. Essential self-adjointness of powers of generators of hyperbolic equations. *J. Funct. Anal.*, 12(4):401–414, 1973.
- [16] G. C. Evans. *Lectures on Multiple-Valued Harmonic Functions in Space*. Univ. of California Press, Berkeley and Los Angeles, 1951.
- [17] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau. Erratum: “Nonexistence of time-periodic solutions of the Dirac equation in an axisymmetric black hole geometry”. *Comm. Pure Appl. Math.*, 53(9):1201, 2000.
- [18] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau. Nonexistence of time-periodic solutions of the Dirac equation in an axisymmetric black hole geometry. *Comm. Pure Appl. Math.*, 53(7):902–929, 2000.
- [19] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau. Decay rates and probability estimates for massive Dirac particles in the Kerr-Newman black hole geometry. *Comm. Math. Phys.*, 230(2):201–244, 2002.
- [20] F. Finster, N. Kamran, J. Smoller, and S.-T. Yau. The long-time dynamics of Dirac particles in the Kerr-Newman black hole geometry. *Adv. Theor. Math. Phys.*, 7(1):25–52, 2003.
- [21] J. R. Gair and D. Lynden-Bell. Electromagnetic fields of separable spacetimes. *Classical and Quantum Gravity*, 24(6):1557, 2007.

- [22] I. M. Glazman. *Direct methods of qualitative spectral analysis of singular differential operators*. Israel Program for Scientific Translations Jerusalem, 1965.
- [23] S. Hawking and G. Ellis. *The Large Scale Structure of Space-Time*. Cambridge Univ. Press, Cambridge, 1973.
- [24] C. Heinicke and F. W. Hehl. Schwarzschild and Kerr solutions of Einstein’s equations – an introduction. *World Scientific Review volume, to appear*, 2014.
- [25] W. Israel. Source of the Kerr metric. *Phys. Rev. D (3)*, 2:641–646, 1970.
- [26] G. Kaiser. Distributional sources for Newman’s holomorphic Coulomb field. *J. Phys. A*, 37(36):8735, 2004.
- [27] E. G. Kalnins and W. Miller, Jr. Series solutions for the Dirac equation in Kerr-Newman space-time. *J. Math. Phys.*, 33(1):286–296, 1992.
- [28] R. P. Kerr. Gravitational field of a spinning mass as an example of algebraically special metrics. *Phys. Rev. Lett.*, 11:237–238, 1963.
- [29] M. K.-H. Kiessling and A. S. Tahvildar-Zadeh. A novel quantum-mechanical interpretation of the Dirac equation. *arXiv:1411.2296*, submitted, 2014.
- [30] W. Kinnersley. Type D vacuum metrics. *J. Math. Phys.*, 10:1195, 1969.
- [31] D. Lynden-Bell. Electromagnetic magic: The relativistically rotating disk. *Phys. Rev. D*, 70:105017, Nov 2004.
- [32] S. F. Neustadter. *Multiple valued harmonic functions with circle as branch curve*, volume 1. University of California Press, 1951.
- [33] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence. Metric of a rotating, charged mass. *J. Math. Phys.*, 6(6):918–919, 1965.
- [34] B. O’Neill. *The Geometry of Kerr black holes*. L.K. Peters, Wellesley, 1995.
- [35] R.J. Oppenheimer and M. Snyder. On continued gravitational contraction. *Phys. Rev.*, 56:455–459, 1939.
- [36] D. Page. Dirac equation around a charged, rotating black hole. *Phys. Rev. D*, 14:1509–1510, 1976.
- [37] R. Penrose. Naked singularities. *Ann. New York Acad. Sci.*, 224:125–134, 1973.
- [38] R. Penrose. The question of cosmic censorship. *J. Astrophys. Astr.*, 20:233–248, 1999.
- [39] H. Prüfer. Neue Herleitung der Sturm-Liouvilleschen Reihenentwicklung stetiger Funktionen. *Math. Ann.*, 95(1):499–518, 1926.
- [40] A. Sommerfeld. Über verzweigte Potentiale im Raum. *Proc. London Math. Soc.*, s1-28(1):395–429, 1896.
- [41] N. Straumann. *General Relativity and Relativistic Astrophysics*. Springer-Verlag, Berlin, 1984.
- [42] K. G. Soffern, E. D. Fackerell, and C. M. Cosgrove. Eigenvalues of the Chandrasekhar-Page angular functions. *J. Math. Phys.*, 24(5):1350–1358, 1983.

- [43] A. S. Tahvildar-Zadeh. On a zero-gravity limit of the Kerr–Newman spacetimes and their electromagnetic fields. *J. Math. Phys.*, 56:042501, 2015.
- [44] S. A. Teukolsky. Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations. *Phys. Rev. Lett.*, 29(16):1114, 1972.
- [45] B. Thaller. Potential scattering of Dirac particles. *J. Phys. A: Math. Gen.*, 14:3067–3083, 1981.
- [46] B. Thaller. *The Dirac equation*. Springer-Verlag, Berlin, 1992.
- [47] N. Toop. The thermal radiation of electrons from a charged spinning black hole in a cosmological background. *preprint DAMTP, Cambridge*, 1976.
- [48] J. Weidmann. Absolut stetiges Spektrum bei Sturm-Liouville-Operatoren und Dirac-Systemen. *Math. Z.*, 180(2):423–427, 1982.
- [49] M. Winklmeier. *The Angular Part of the Dirac Equation in the Kerr-Newman Metric: Estimates for the Eigenvalues*. PhD thesis, Universität Bremen, 2005.
- [50] M. Winklmeier and O. Yamada. Spectral analysis of radial Dirac operators in the Kerr-Newman metric and its applications to time-periodic solutions. *J. Math. Phys.*, 47(10):102503, 17, 2006.
- [51] M. Winklmeier and O. Yamada. A spectral approach to the Dirac equation in the non-extreme Kerr-Newmann metric. *J. Phys. A*, 42(29):295204, 15, 2009.
- [52] D. M. Zipoy. Topology of some spheroidal metrics. *J. Math. Phys.*, 7(6):1137–1143, 1966.