Variational Methods for Quantum Hamiltonians

Sabiha Sibel Tokus

PhD Thesis

This thesis has been submitted to the PhD School of
the Faculty of Science, University of Copenhagen.

Sabiha Sibel Tokus
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 København Ø
Denmark
sabiha.tokus@math.ku.dk

Date of submission:	14th July 2021
Date of defence:	28th September 2021
Advisor:	Jan Philip Solovej University of Copenhagen, Denmark
Assessment committee:	Horia D. Cornean Aalborg University, Denmark
	Bergfinnur Durhuus (chair) University of Copenhagen, Denmark
	Maria J. Esteban CEREMADE, Université Paris-Dauphine, France

ISBN 978-87-7125-046-6

Abstract

This thesis consists of two parts, each investigating variational approaches to study spectral properties of quantum Hamiltonians in different settings.

The first part is based on a paper with Lukas Schimmer and Jan Philip Solovej in which we construct a distinguished self-adjoint extension of symmetric operators that satisfy a certain gap condition. Additionally, we prove a corresponding variational principle which gives the eigenvalues of the self-adjoint extension in the gap. A prominent example of such a gapped operator is the Coulomb–Dirac operator, describing non-interacting relativistic fermions in a Coulomb field. Another example that falls into the class of gapped operators is the Dirac operator on a bounded cylinder. For such an operator we relate the self-adjoint extension to the non-local Atiyah–Patodi–Singer boundary conditions.

In prior work, Esteban and Loss [EL] constructed a self-adjoint extension of certain gapped operators and Dolbeault, Esteban and Séré [DES00] presented a min-max principle for self-adjoint operators with a gap in the essential spectrum.

Our construction of a self-adjoint extension of gapped operators is in many ways inspired by the work of Esteban and Loss, but starts from a more general setting and shows some difference in the concrete implementation of the basic ideas. The min–max principle for the eigenvalues in the gap of such a self-adjoint extension has similarities to the one presented in [DES00]. The crucial difference and novelty in our setting, however, is that we do not need to determine the domain of the self-adjoint operator since our min–max principle can be formulated by specifying the domain of the symmetric operator only.

This characteristic feature of our min–max principle has an analogue in the case of lower semi-bounded symmetric operators, for which there is known to exist a distinguished self-adjoint extension, the Friedrichs extension. We call our self-adjoint extension a Friedrichs extension for gapped operators. In fact, the Friedrichs extension appears as a special case in our construction of a self-adjoint extension for a gapped operator.

The second part of the thesis is concerned with a variational approach to Bogoliubov's approximation theory of bosonic systems interacting via two-body potentials. This approach was introduced and analysed by Napiórkowski, Reuvers and Solovej [NRS18a] as a variational reformulation of Bogliubov's approximation. We call it the Bogoliubov variational principle.

We test this approximation method by applying it to the Lieb–Liniger model of a Bose gas in one dimension interacting via a delta-potential. The Lieb–Liniger model as introduced by [LL63] is an exactly solvable model and therefore particularly suitable for testing approximation schemes, as already pointed out by Lieb and Liniger themselves. The model has effectively one parameter, $\xi = c/\varrho$, where $2c$ is the coupling strength and ϱ the particle density. The ground state energy in the Lieb–Liniger gas is not known explicitly, but it can be expanded for small and large ξ.

We compute the ground state energy of the Lieb–Liniger gas by the Bogoliubov variational method as expansions in ξ for $\xi \to \infty$ and $\xi \to 0$. For large values of ξ we find that the ground state energy diverges as $\xi^{1/2}$. Our results thus show that the Bogoliubov variational principle is not suited to describe the model at large ξ where the ground state energy is expected to approach an asymptotic value of $\pi^2/3$.

For $\xi \to 0$ our result agrees with the first two terms of an expansion of the exact Lieb–Liniger model. A third term for the exact model has been derived by Tracy and Widom [TW16]. Our third term differs by a squared logarithmic factor from Tracy and Widom's result.

Resumé

Denne afhandling består af to dele, der begge undersøger variationelle tilgange til studiet af spektrale egenskaber for kvantemekaniske Hamiltonoperatorer i to forskellige setups.

Den første del er baseret på en artikel med Lukas Schimmer og Jan Philip Solovej, hvor vi konstruerer en særlig selvadjungeret udvidelse af en symmetrisk operator, der opfylder en bestemt 'gabbetingelse'. Vi viser yderligere et variationelt princip, der giver egenværdierne for den selvadjungerede udvidelse i gabet. Et vigtigt eksempel på en sådan operator med et gab er Coulomb–Dirac operatoren, der beskriver ikke-interagerende relativistiske fermioner i et Coulumbfelt. Et andet eksempel, der falder indenfor klassen af operatorer med gab, er Dirac operatoren på en begrænset cylinder. For sådan en operator relaterer vi den selvadjungerede udvidelse til de ikke-lokale Atiyah–Patodi–Singer grænsebetingelser.

Før vores projekt er en konstruktion af en selv-adjungeret udvidelse af særlig operator med gab blevet præsenteret af Esteban og Loss i [EL] og et min–max princip for selvadjungerede operatorer med et gab i det essentielle spectrum af Dolbeault, Esteban og Séré i [DES00].

Vores konstruktion af en selvadjungeret operator for operatorer med gab er på mange måder inspireret af Esteban and Loss' arbejde, men det tager et mere generelt udgangspunkt og adskiller sig på flere områder i den konkrete implementering af de grundlæggende idéer. Min-max princippet for egenværdierne i gabet af sådanne selvadjungerede udvidelser har ligheder med det som blev præsenteret i [DES00]. Den vigtigste forskel og nyskabelse i vores setup er, at vi ikke behøver at bestemme domænet af den selvadjungerede operator, da vores min–max princip kan formuleres ved blot at specificere domænet af den symmetriske operator.

Den karakteristiske egenskab ved min–max princippet findes tilsvarende også i tilfældet hvor vi har en nedre halvbegrænset symmetrisk operator, hvor det er kendt at der findes en særlig selvadjungeret udvidelse, Friedrichs udvidelsen. Vi kalder vores selvadjungerede udvidelse for en Friedrichs udvidelse for operatorer med gab. Faktisk kan Friedrichs udvidelsen findes som et specialtilfælde af vores konstruktion af en selvadjungeret udvidelse af en operator med gab.

Afhandlingens anden del beskæftiger sig med en variationel tilgang til Bogoliubovs approximationsteori for bosoniske systemer, der interagerer med to-legeme potentialer. Tilgangen blev introduceret og analyseret af Napiórkowski, Reuvers og Solovej [NRS18a] som en variationel omformulering af Bogliubovs approximation. Vi kalder det for det variationelle Bogoliubov princip.

Vi tester approximationsmetoden ved at anvende den på Lieb–Liniger modellen for en Bose gas i én dimension, der interagerer via et delta-potential. Lieb–Liniger modellen som blev introduceret i [LL63] kan løses eksakt og er derfor særligt god til at teste approximationsmetoder, hvilket allerede blev pointeret af Lieb og Liniger selv. Modellen har effektivt en parameter, $\xi = \frac{c}{\varrho}$, hvor $2c$ er koblingsstyrken og ϱ er partikeltætheden. Grundtilstandsenergien i Lieb–Liniger gassen er ikke kendt explicit, men kan ekspanderes i ξ.

Vi udregner grundtilstandsenergien af Lieb–Liniger gassen ved Bogoliubov variationelle metode som udvidelse i ξ for $\xi \to \infty$ og $\xi \to 0$. For store værdier af ξ finder vi, at grundtilstandsenergien divergerer som $\xi^{\frac{1}{2}}$. Vores resultater viser derfor, at Bogoliubov variationelle princip ikke er passende til at beskrive modellem for store ξ, hvor grundtilstandsenergien forventes at nærme sig værdien $\frac{\pi^2}{3}$ asymptotisk.

For $\xi \to 0$ stemmer vores resultat overens med de første to led i en ekspandering af den eksakte Lieb–Liniger model. Et tredje led for den eksakte model er blevet udledt af Tracy og Widom [TW16]. Vores tredje led er forskelligt fra Tracy og Widoms resultat med en logaritmisk faktor kvadreret.

Acknowledgements

I would like to take the opportunity to thank those people who contributed to this thesis in one or the other way.

My thanks go first and foremost to my PhD advisor Jan Philip Solovej. I am deeply grateful for his continuous advice and support at all stages of my PhD project, for his patience with all the questions I had to ask several times and for the respect and open-mindedness he is always meeting his students with.

I would also like to thank Lukas Schimmer for a great collaboration during my first PhD project about self-adjoint extensions. I have always enjoyed the inspiring discussions and I am more than grateful for an always open door and patience for my many questions.

During my PhD I spent some time at the University of Milan. I would like to thank Niels Benedikter for his hospitality and many interesting discussions which always gave me a motivational boost both in Milan and on zoom.

I was lucky to have a couple of colleagues and friends willing to proof-read parts of the thesis and translate my abstract. For their help and many valuable comments I would like to thank Anna-Sofie Schilling, Frederik Ravn Klausen, Joshua Hunt, Lukas Schimmer, Mads Friis Frand-Madsen, Marie Fialová and Paula Belzig.

I have always been happy to have chosen the University of Copenhagen and in particular the QMath centre for my PhD project. My thanks go therefore to all department members who contributed to making my time here a pleasant one. In particular, I would like to thank all the former and present members of the QMath centre for an extremely welcoming and friendly atmosphere. This really makes a difference for a PhD student. Thank you to all my colleagues I shared an office or coffee breaks with for support, many good conversations and a lot of fun.

Last but not least my thanks go to family and friends for their support and giving me opportunities to regularly break out from the world of math and thesis-writing.

A note to the reader

The two parts of this thesis are completely independent, each with its own introduction and list of references. They can be read separately and are put together in this thesis in the order in which they were carried out.

Contents

Part I

Variational methods for operators with gaps

Friedrichs extension and min–max principle for operators with a gap

1 Preliminaries

1.1 Introduction

An attempt to solve eigenvalue problems for linear differential operators inspired Friedrichs to develop and present a spectral theory for semibounded linear operators in 1934 [Fri34]. In particular in his mind were those quantum mechanical energy operators known as Schrödinger operators, that fall into this class of semibounded operators, and for which he did not have the variational approaches for the calculation of normal eigenvalues available.

Using methods of abstract Hilbert space theory, he could show how to construct from a lower bounded symmetric operator A a distinguished self-adjoint operator extending A. His construction is today known as the *Friedrichs extension* of a lower semibounded symmetric operator. For such an operator, if densely defined, one finds a distinguished self-adjoint extension, equally lower semibounded with the same bound and an operator domain that is contained in the form domain of a corresponding lower semibounded sesquilinear form. Whereas there is no reason to assume that the Friedrichs extension of a particular lower semibounded operator is the only extension that preserves its lower bound, one can show that it is the only one whose domain has the property to be a subset of the corresponding form domain.

This is also the reason why it is particularly easy to formulate and make use of a variational principle (or min–max principle) for the Friedrichs extension from which one may calculate its eigenvalues below the lowest point of the essential spectrum. Compared to other extensions one here only needs the operator domain of the original *symmetric* operator for computations which can be of advantage in practical matters.

Lower semibounded operators play an important role in the theory of quantum mechanics. In many cases that are interesting for us the Hamilton energy operator determining the dynamics of a system via the Schrödinger equation is bounded from below - the energy is bounded from below and the system is stable. Therefore the Friedrichs extension is also of importance in mathematical physics.

Description of the project: In the article [SST20], which forms the main part of the first part of this PhD thesis, we construct a self-adjoint extension and corresponding

variational principle for an operator that is not lower semibounded but satisfies a certain gap condition. By a gap condition we mean a property of a symmetric operator that may translate to a spectral gap under the process of a self-adjoint extension. The question about a generalisation from the lower semibounded to the gapped case appears naturally. At least formally we can understand lower semiboundedness as a gap with a lower limit being $-\infty$.

We understand our construction as a generalisation of the Friedrichs extension of a lower semibounded operator. This has the following reason: A key element in the construction of the Friedrichs extension is the close relation between closed sesquilinear forms and self-adjoint operators. We make use of this fact in the same manner for our construction of a gapped extension. In the way we formulate the gap condition, our type of gapped operators can be written as a block matrix. The Friedrichs construction then really appears as a special case of ours in the case of a vanishing block structure, which can be seen as a transition of the lower gap limit to $-\infty$. The fact that the self-adjoint extension can be derived from the closure of a sesquilinear form is also the crucial feature that makes it possible to formulate a min–max principle only relying on the knowledge of the domain of the symmetric operator, both in the case of the Friedrichs extension and our generalisation.

Boundary conditions are closely related to self-adjoint extensions. For example, it is well-known that the Laplace operator on $C_0^\infty(\Omega) \subseteq L^2(\Omega)$ for $\Omega \subseteq \mathbb{R}^n$ being an open set in $\mathbb{R}^n$ is lower semibounded and has a Friedrichs extension which coincides with the Dirichlet–Laplacian, with domain contained in $H_0^1(\Omega)$. That is, it is defined on functions satisfying the Dirichlet boundary condition.

In our paper we relate the generalised Friedrichs extension for the example of a first order differential operator of Dirac-type to a non-local boundary condition, known as the Atiyah–Patodi–Singer boundary condition.

Outline. In this introductory part, we aim to familiarise a possible reader in Section 1.2 with the Friedrichs extension of a lower semibounded operator, starting by briefly collecting some necessary mathematical notions and concepts, followed by a derivation of the Friedrichs extension and a proof of its min–max principle. We hope this section can also serve for highlighting the parallels between Friedrichs' and our construction.

Section 1.3 is devoted to a glimpse on the topic of self-adjoint extensions from a physical point of view. In Section 1.4, we give a short introduction to the theory of gapped operators with special focus on the example of the Coulomb–Dirac operator. The paper "Friedrichs extension and min–max principle for operators with a gap" [SST20] with a detailed derivation of the self-adjoint extension for gapped operators and its variational principle follows these introductory chapters.

1.2 The Friedrichs extension

1.2.1 Hilbert space operators

We start by shortly collecting the basic definitions, notions and conventions that form the basis for the following sections. This is only a very brief summary of a selection of the most important definitions and relations with regard to our purposes. For a deeper-going study of these concepts, there are a variety of books that discuss these in detail (e.g. [Sch12] to name one of them). The subject of this first part are linear operators on Hilbert spaces.

A **Hilbert space** $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a vector space equipped with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ which at the same time is a complete metric space in the metric induced by its inner product. We usually assume the Hilbert spaces occurring in this thesis to be complex vector spaces and the inner product to be conjugate linear in the first argument and linear in the second argument, such that for all $u, v, w \in \mathcal{H}, \alpha, \beta \in \mathbb{C}$

$$\begin{aligned}\langle \alpha u + \beta v, w\rangle_{\mathcal{H}} &= \overline{\alpha}\langle u, w\rangle_{\mathcal{H}} + \overline{\beta}\langle v, w\rangle_{\mathcal{H}}\,, \quad \text{and}\\ \langle u, \alpha v + \beta w\rangle_{\mathcal{H}} &= \alpha\langle u, v\rangle_{\mathcal{H}} + \beta\langle u, w\rangle_{\mathcal{H}}\,.\end{aligned}$$

Furthermore, all the Hilbert spaces occurring are understood to be separable Hilbert spaces, thus admitting a countable orthonormal basis. A **linear operator** T from a Hilbert space $\mathcal{H}_1$ to another Hilbert space $\mathcal{H}_2$ is a linear mapping $T : D(T) \subseteq \mathcal{H}_1 \to \mathcal{H}_2$ from a linear subspace $D(T)$ of $\mathcal{H}_1$ into $\mathcal{H}_2$. The subspace $D(T)$ is the domain of T. If $\mathcal{H}_1 = \mathcal{H}_2$, we call T a linear operator on $\mathcal{H}$. Since we are exclusively considering linear operators, we might occasionally refer to them only by the name of *operators*.

We write $T \subseteq S$ for linear operators T and S from $\mathcal{H}_1$ to $\mathcal{H}_2$ if $D(T) \subseteq D(S)$ and $T = S$ on $D(T)$. The operator S is called an **extension** of T.

Corresponding to a linear operator, starting from the inner products on $\mathcal{H}_1$ and $\mathcal{H}_2$, one constructs an inner product on $D(T)$

$$\langle\cdot,\cdot\rangle_T = \langle\cdot,\cdot\rangle_{\mathcal{H}_1} + \langle T\cdot, T\cdot\rangle_{\mathcal{H}_2}\,,$$

which itself induces a norm

$$\|\cdot\|'_T = \left(\|\cdot\|^2_{\mathcal{H}_1} + \|T\cdot\|^2_{\mathcal{H}_2}\right)^{1/2}.$$

A norm equivalent to $\|\cdot\|_T$ is

$$\|\cdot\|_T = \|\cdot\|_{\mathcal{H}_1} + \|T\cdot\|_{\mathcal{H}_2}\,.$$

Note that $\|\cdot\|_{\mathcal{H}_1} \leq \|\cdot\|_T$ (the same holds for $\|\cdot\|'_T$). Both of the norms are frequently referred to as the **graph norm** of an operator T. We will most frequently use $\|\cdot\|_T$. The **graph** of

an operator T is the set

$$\mathcal{G}(T) = \{(x, Tx) \mid x \in D(T)\} \subseteq \mathcal{H}_1 \oplus \mathcal{H}_2 \,.$$

It is clear that any linear operator T sends $0 \in \mathcal{H}_1$ to $0 \in \mathcal{H}_2$, that is, $(0, y)$ is an element of the graph only if $y = 0$. Conversely, a linear subspace $U \leq \mathcal{H}_1 \oplus \mathcal{H}_2$ is the graph of a linear operator if (and by the preceding sentence only if) $(0, y) \in U$ implies $y = 0$. These notions of the graph and the graph norm of an operator allow for characterising a special class of operators, namely **closed** and **closable** operators.

Definition 1. A linear operator $T : D(T) \subseteq \mathcal{H}_1 \to \mathcal{H}_2$ is closed if its graph is a closed subset of $\mathcal{H}_1 \oplus \mathcal{H}_2$. Equivalently, by the above definitions, T is closed if and only if $D(T)$ is complete in the graph norm, that is, if $(D(T), \langle \cdot, \cdot \rangle_T)$ forms a Hilbert space. An operator is called closable if there exists a closed linear operator S such that $T \subseteq S$.

Closability of an operator can be characterised as in the following proposition.

Proposition 1. A linear operator $T : D(T) \subseteq \mathcal{H}_1 \to \mathcal{H}_2$ is closable if and only if the closure of its graph $\mathcal{G}(T)$ in $\mathcal{H}_1 \oplus \mathcal{H}_2$ is the graph of a linear operator.

Furthermore, this is equivalent to the following situation: if $(x_n)_{n\in\mathbb{N}}$ is a sequence of vectors in $D(T)$ and we have

$$\begin{aligned} \lim_{n\to\infty} x_n &= 0 \quad \text{in } \mathcal{H}_1 \quad \text{and} \\ \lim_{n\to\infty} Tx_n &= y \quad \text{in } \mathcal{H}_2 \,, \end{aligned}$$

then $y = 0$.

If T is a closable operator, the smallest closed extension of T is called the **closure** of T and is denoted by $\overline{T}$. It is the operator corresponding to the closure of the graph of T, that is $\mathcal{G}(\overline{T}) = \overline{\mathcal{G}(T)}$. The domain of $\overline{T}$ is the closure of $D(T)$ in the graph norm $\|\cdot\|_T$. One also says in this context that $D(T)$ is a core for $\overline{T}$.

A special type of closed operators are **bounded** or **continuous** operators. The criterion for a linear operator to be continuous is stronger than the closedness condition from Proposition 1: an operator is continuous if for any sequence $(x_n)_{n\in\mathbb{N}}$ in $D(T)$ with $\lim_{n\to\infty} x_n = 0$, the sequence of Tx_n necessarily converges and we have $\lim_{n\to\infty} Tx_n = 0$.

In fact, a linear operator T is continuous if and only if it is bounded, that is, if and only if there is a constant $M > 0$ such that

$$\|Tx\|_{\mathcal{H}_2} \leq M \|x\|_{\mathcal{H}_1} \,,$$

or equivalently if

$$\sup_{x\in D(T)\setminus\{0\}} \frac{\|Tx\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}} < \infty \,.$$

It should be mentioned that bounded operators are a special class of linear operators. In many cases, they are easier to treat and one has to deal with much fewer subtleties than in the case of unbounded operators.

As an example of another class of closable operators, we can name the **densely defined**, **symmetric** operators. Densely defined here means that the domain $D(T)$ is a dense subset of the Hilbert space $\mathcal{H}_1$.

Definition 2. A linear operator $T : D(T) \subseteq \mathcal{H} \to \mathcal{H}$ is called symmetric if the following property holds:

$$\langle x, Ty\rangle = \langle Tx, y\rangle \qquad \text{for all } x, y \in D(T)\,.$$

Clearly, if $D(T)$ is dense, we find that the criterion for closability given in Proposition 1 is satisfied, hence densely defined symmetric operators are always closable.

Symmetric operators can be detected via the relation known as the *polarisation identity*[1]: using this identity it is easy to show that an operator T is symmetric if and only if the quantity $\langle x, Tx\rangle$ is a real number for all $x \in D(T)$.

We now introduce the (Hilbert-) **adjoint operator** of a densely defined linear operator.

Definition 3. Let T be a densely defined operator from a Hilbert space $\mathcal{H}_1$ to a Hilbert space $\mathcal{H}_2$. To every such T we can assign a corresponding linear operator from $\mathcal{H}_2$ to $\mathcal{H}_1$ by the following prescription. Define

$$D(T^*) \coloneqq \{y \in \mathcal{H}_2 \mid \exists v \in \mathcal{H}_1 : \langle y, Tx\rangle_{\mathcal{H}_2} = \langle v, x\rangle_{\mathcal{H}_1} \text{ for all } x \in D(T)\}\,. \tag{1.1}$$

For any $y \in D(T^*)$ we set $T^*y = v$. This defines a linear operator T^*. It is called the **adjoint operator** of T.

By Riesz' representation theorem such a $v \in \mathcal{H}_1$ exists for a given $y \in \mathcal{H}_2$ exactly when the linear functional $x \mapsto \langle Tx, y\rangle_{\mathcal{H}_2}$ is continuous, that is, if $\langle Tx, y\rangle_{\mathcal{H}_2} \leq C\|x\|_{\mathcal{H}_1}$ for some $C > 0$ and all $x \in D(T)$.

Note that it is crucial in this context to assume $D(T)$ to be dense for T^* to be a well-defined mapping. In this case, one can from the very definition directly conclude that T^* is a closed operator.

Additionally, we can see that an operator T on a Hilbert space $\mathcal{H}$ is symmetric if and only if $T \subseteq T^*$. Thus, densely defined symmetric operators are closable.

Definition 4. A densely defined linear operator on a Hilbert space $\mathcal{H}$ is called **self-adjoint** if $T = T^*$.

We say that an operator T is **essentially self-adjoint** if $\overline{T} = T^*$. As a matter of fact, the closure of T is in this case the only self-adjoint operator extending T.

[1] The polarisation identity for an operator T denotes the identity $\langle Tx, y\rangle = \frac{1}{4}\left(\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle + \mathrm{i}\langle T(x+\mathrm{i}y), x+\mathrm{i}y\rangle - \langle T(x-\mathrm{i}y), x-\mathrm{i}y\rangle\right)$.

1.2.2 Sesquilinear forms

We now introduce an additional notion that is closely connected to linear operators.

Definition 5. A **sesquilinear form** or **quadratic form**[2] (short: form) on a Hilbert space $\mathcal{H}$ is a mapping $s : D(s) \times D(s) \subseteq \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that is linear in the first argument and conjugate linear in the second argument. The linear subspace $D(s)$ of $\mathcal{H}$ is called the domain of s.

A sesquilinear form is said to be **symmetric** if $s(x, y) = \overline{s(y, x)}$ for all $x, y \in D(s)$. In the same way as for operators one finds that s is symmetric if and only if $s(x, x) \in \mathbb{R}$ for all $x \in D(s)$.

One can easily see that, given a linear operator A on a Hilbert space, we can directly construct a sesquilinear s_A form based on the inner product on $\mathcal{H}$ such that for vectors $x, y \in D(A)$

$$s_A(x, y) = \langle Ax, y \rangle .$$

A priori, we set the form domain $D(s)$ equal to the operator domain $D(A)$, although the quantity (Ax, y) might be defined for a larger set. Turning things around, we can also assign an operator to a given form s defined on a form domain $D(s)$, if $D(s)$ is densely defined. Setting

$$D(A_s) = \{x \in D(s) \mid \exists u \in \mathcal{H} : s(x, y) = \langle u, y \rangle_{\mathcal{H}} \text{ for all } y \in D(s)\} , \tag{1.2}$$

we see again by Riesz' representation theorem, that such an u exists exactly when $y \mapsto s(x, y)$ is a bounded, in this case conjugate linear functional. We then define $A_s x = u$ and we have $\langle A_s x, y \rangle = s(x, y)$. In general, $D(A_s)$ is strictly smaller than $D(s)$ and does not even have to be dense (an example is found in [Sol14], Example 5.4). Hence, starting out from a sesquilinear form, finding the corresponding operator and from there the associated form might not necessarily bring us back to the original domain we started from. This also shows us that for a general operator and form there is no simple 'correspondence rule' found without additional information. Such a correspondence is only found for what we call closed forms and self-adjoint operators. We will introduce closed forms in the remainder of this section.

Definition 6. We call a sesquilinear form on a Hilbert space $\mathcal{H}$ **lower semibounded** if there is a number $m \in \mathbb{R}$ such that $s(x, x) \geq -m\|x\|^2$ for all $x \in D(s)$. Similarly, we call a linear operator A on a Hilbert space $\mathcal{H}$ lower semibounded if there is a number $m \in \mathbb{R}$ such that for all $x \in D(A)$ it holds $\langle Ax, x \rangle \geq -m\|x\|^2$. In this case we write $s \geq -m$ and $A \geq -m$, respectively.

[2] The term quadratic form is sometimes reserved for the mapping defined by $q(v) = s(v, v)$. In fact, by the polarisation identity for forms, one finds that s is already completely determined by q.

It is clear that a lower semibounded sesquilinear form or lower semibounded operator is necessarily symmetric. If a form is lower semibounded with bound $-m$, we can form an inner product on $D(s)$ by setting

$$\langle x, y\rangle_s = s(x,y) + (m+1)\langle x, y\rangle \quad \text{for } x, y \in D(s)\,,$$

and a corresponding norm by

$$\|x\|_s = \left(s(x,x) + (m+1)\|x\|^2\right)^{1/2}\,. \tag{1.3}$$

One notices that $\|\cdot\| \leq \|\cdot\|_s$. Furthermore, the form s is bounded on $D(s)$ with norm $\|\cdot\|_s$, that is

$$|s(x,y)| \leq \|x\|_s \|y\|_s\,. \tag{1.4}$$

The existence of the norm $\|\cdot\|_s$ allows us to define a notion of 'closedness' for forms. One can notice that this is really only possible for forms that are lower semibounded. In contrast, such a condition is not needed for operators.

Definition 7. A lower semibounded form s on a Hilbert space $\mathcal{H}$ is called **closed** if $D(s) \subseteq \mathcal{H}$ is complete in the $\|\cdot\|_s$-norm, that is if $(D(s), \langle\cdot,\cdot\rangle_s)$ forms a Hilbert space. It is said to be **closable** if there exists a closed form p on $\mathcal{H}$ that extends s, that is, for which $D(s) \subseteq D(p) \subseteq \mathcal{H}$ and $s(x,y) = p(x,y)$ for $x, y \in D(s)$.

In general, completing $(D(s), \langle\cdot,\cdot\rangle_s)$ always yields a Hilbert space that could serve as the domain for a closed extension of s, we could denote it by $\mathcal{H}_s$. Furthermore, since $\|\cdot\|_{\mathcal{H}} \leq \|\cdot\|_s$, the Hilbert space $\mathcal{H}_s$ can be embedded into $\mathcal{H}$ by continuation of the continuous embedding $i : D(s) \to \mathcal{H}$. A priori it is, however, not clear if this continuation $\hat{i} : \mathcal{H}_s \to \mathcal{H}$ is also injective, hence it is not immediately obvious that the completion of $D(s)$ really is a subset of $\mathcal{H}$. Similar considerations hold for the closability of an operator when we think of closing the operator in terms of completing the operator domain in graph norm. Likewise as in the case for operators, there is a handy criterion to check whether a form is closable.

Proposition 2. A lower semibounded form s defined on $D(s) \subseteq \mathcal{H}$ is closable if and only if the following holds: if $(x_n)_{n\in\mathbb{N}}$ is a sequence of vectors in $D(s)$ such that

$$\lim_{n\to\infty} x_n = 0 \quad \text{and}$$
$$\lim_{n,m\to\infty} s(x_n - x_m, x_n - x_m) = 0$$

where the latter relation means that (x_n) is a Cauchy sequence also in the $\|\cdot\|_s$-norm, then we have $\lim_{n\to\infty} s(x_n, x_n) = 0$. This is equivalent to showing that the map $\hat{i}$ is injective.

If s is closable, then we can continue s from $D(s)$ to a form on all of $\mathcal{H}_s$ by continuity since s is bounded on $(D(s), \langle\cdot,\cdot\rangle_s)$ by (1.4). The form we obtain is called the closure of s and we denote it by $\overline{s}$.

Closed sesquilinear forms have a strong connection to self-adjoint operators (indeed, there is a one-to-one relation between closed forms and lower semibounded self-adjoint operators, cf. Corollary 10.8 in [Sch12] for an explanation). This fact is also exploited in the construction of the Friedrichs extension, which is reviewed in the following section.

1.2.3 Construction of the Friedrichs extension

In this section we show a proof of

Theorem 1. Let $A : D(A) \subseteq \mathcal{H} \to \mathcal{H}$ be a densely defined symmetric operator on a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ that is lower semibounded such that $A \geq -m$ for a constant $m \in \mathbb{R}$. Without loss of generality one may assume $m > 0$. There is a distinguished self-adjoint extension A_F of A, that preserves the lower bound of A, such that also $A_F \geq -m$. Its domain is a subset of the form domain of the closure of the sesquilinear form corresponding to A and it is the only self-adjoint extension of A with that property. The self-adjoint operator A_F is known as the **Friedrichs extension** of A.

The proof is due to Friedrichs [Fri34]. This presentation is a freely retold version of his construction, with some inspiration from Theorem VIII.15 and X.23 in [RS80a] and [RS80b], respectively. We first show

Lemma 1. Let s_A be the sesquilinear form associated to a lower semibounded operator A with domain $D(A)$ and lower bound $A \geq -m$, $m \geq 0$. It holds for $x, y \in D(A)$

$$s_A(x, y) = \langle Ax, y\rangle \,.$$

Then the form s_A is closable.

Proof. By assumption, s_A is bounded from below by $s \geq -m$. We construct an inner product on $D(A)$ as above given by

$$\langle x, y\rangle_s = s(x, y) + (m + 1)\langle x, y\rangle \qquad \text{for } x, y \in D(A)\,,$$

and a corresponding norm, defined by $\|x\|_s^2 = \langle x, x\rangle_s$. The linear space $D(A)$ can be completed in the norm $\|\cdot\|_s$ derived from $\langle\cdot,\cdot\rangle_s$ to a Hilbert space $\mathcal{H}_{s_A}$. By the remarks prior to Proposition 2, we need to show that the embedding map $\hat{i} : \mathcal{H}_{s_A} \to \mathcal{H}$ is injective. We use the criterion formulated in Proposition 2. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of vectors in $D(A)$, and $\lim_{n\to\infty} x_n = 0$ in $\mathcal{H}$. Assume furthermore that the sequence $(x_n)_n$ is also convergent in $\mathcal{H}_{s_A}$, that is, we additionally assume that $\lim_{n,m\to\infty} s(x_n - x_m, x_n - x_m) = 0$. Thus we have for any $k \in \mathbb{N}$

$$\begin{aligned} |s(x_n, x_n)| &\leq |s(x_n - x_k, x_n)| + |s(x_k, x_n)| \\ &\leq \|x_n - x_k\|_s \|x_n\|_s + |\langle Ax_k, x_n\rangle| \\ &\leq \|x_n - x_k\|_s \|x_n\|_s + \|Ax_k\| \|x_n\| \,. \end{aligned} \tag{1.5}$$

By assumption the sequence of numbers $(\|x_n\|_s)_n$ is clearly bounded. Let $\varepsilon > 0$. Choosing now k and n large enough, $k, n > N_1(\varepsilon)$, we can achieve that the first term in (1.5) is arbitrarily small, in particular smaller than $\varepsilon/2$. For the second term we have $x_k \in D(A)$ for any fixed k, hence, if additionally n is large enough, $n > N_2(\varepsilon)$, also this term will be smaller than $\varepsilon/2$. Taking all together, when selecting $k > N_1(\varepsilon)$, and at the same time $n > \max\{N_1, N_2\}$, we are able to bound (1.5) by ε. Since ε was arbitrary, $\lim_{n\to\infty} s(x_n, x_n) = 0$ and s is closable. □

Proof of Theorem 1: Associated to the symmetric operator A, there is the form s_A defined on $D(A)$ with $s_A(x, y) = \langle Ax, y\rangle$. By Lemma 1, this form is lower semibounded and closable to a closed form $\overline{s}_A$, and $\mathcal{H}_{s_A}$ is the Hilbert space formed by the completion of $D(A)$ with the inner product $\langle\cdot,\cdot\rangle_s$.

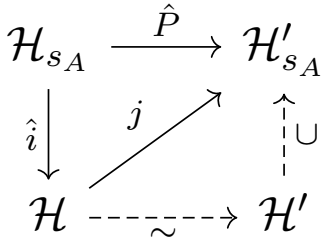
We denote by $\mathcal{H}'_{s_A}$ the dual space of $\mathcal{H}_{s_A}$, the space of bounded linear functionals on $\mathcal{H}_{s_A}$. By Riesz' representation theorem $\mathcal{H}_{s_A}$ and $\mathcal{H}'_{s_A}$ are isometrically isomorphic via the conjugate linear map $\hat{P} : \mathcal{H}_{s_A} \to \mathcal{H}'_{s_A}$, defined by

$$[\hat{P}x](y) = \langle x, y\rangle_s = \overline{s}_A(x, y) + (m+1)\langle x, y\rangle\,.$$

Likewise, there is an equivalent isomorphism between $\mathcal{H}$ and its dual space $\mathcal{H}'$. At the same time, every bounded linear functional on $\mathcal{H}$ is also a linear functional on $\mathcal{H}_{s_A} \subseteq \mathcal{H}$. Hence, we can map any vector in $\mathcal{H}$ to a linear functional on $\mathcal{H}_{s_A}$ by a map j given by

$$[j(x)](y) = \langle x, y\rangle\,.$$

Indeed, since $\|\cdot\|_{\mathcal{H}} \leq \|\cdot\|_{\mathcal{H}_{s_A}}$, the element $j(x) : y \mapsto \langle x, y\rangle \leq \|x\|\|y\|_{\mathcal{H}_{s_A}}$ is even a bounded linear functional on $\mathcal{H}_{s_A}$. Hence, j is a conjugate linear embedding of $\mathcal{H}$ into $\mathcal{H}'_{s_A}$. In total, we have the relations between the spaces as illustrated in the diagram. One should note that the upper left corner is not commuting.



We now define an operator $A_F = j^{-1}\hat{P} - (m+1)\mathbb{1}$ on the domain

$$D(A_F) = \left\{x \in \mathcal{H}_{s_A} \;\middle|\; \hat{P}x \in \operatorname{Ran} j\right\}\,.^{3}$$

We will show that A_F is a densely defined, self-adjoint operator extending A. We will

[3] A comparison shows that this is exactly a reformulation of (1.2).

prove these characteristics in the first place for the operator $P = j^{-1}\hat{P}$, with domain $D(P) = D(A_F)$, but the same results immediately follow for $A_F = P - (m+1)\mathbb{1}$ by its relation to P.

First, we establish that P is a densely defined operator on $\mathcal{H}$. Making use of the facts that $\|\cdot\|_{\mathcal{H}} \leq \|\cdot\|_{\mathcal{H}_{s_A}}$ and that $\mathcal{H}_{s_A}$ is a dense subspace of $\mathcal{H}$ as well as that $\hat{P}$ is an isometric isomorphism, it is enough for us to prove that Ran$j$ is dense in $\mathcal{H}'_{s_A}$. Therefore, assume $g \in \mathcal{H}''_{s_A}$ and $g[j(x)] = 0$ for all $x \in \mathcal{H}$. By definition of the double dual $0 = g[j(x)] = [j(x)](y_g) = \langle x, y_g\rangle$ for a unique $y_g \in \mathcal{H}_{s_A}$ and for all $x \in \mathcal{H}$. We conclude $y_g = 0$ and hence $g = 0$. Thus, Ran$j$ is dense.

Next, since $j(Px) = \hat{P}x$ on $D(A_F)$, we have for $x, y \in D(A_F)$

$$\langle Px, y\rangle = \langle x, y\rangle_s = \overline{\langle y, x\rangle_s} = \overline{\langle Py, x\rangle} = \langle x, Py\rangle\,.$$

Hence, P is symmetric.

Furthermore, we show the self-adjointness of P. First of all, we observe that P is injective and that Ran $P = \mathcal{H}$. This, by definition of the adjoint operator (1.1), tells us that $\ker P^* = \{0\}$, because if $y \in \ker P^*$, then we have $\langle Px, y\rangle = 0$ for all $x \in D(P) = D(A_F)$. Consequently, y must be orthogonal on all of $\mathcal{H}$, and therefore $y = 0$. But if $\ker P^* = \{0\}$, then we further conclude: for $x \in D(P^*)$, there is an $x' \in D(P)$ such that $P^*x = Px'$, since Ran $P = \mathcal{H}$. Because $P \subseteq P^*$, we also have $P^*x' = Px'$. Hence, $x = x'$ and $x \in D(P)$. Thus P is self-adjoint and so is A_F.

In a second-to-last step, we observe that for $x \in D(A_F)$ and $y \in D(A)$

$$\langle x, Ay\rangle = \langle x, y\rangle_s - (m+1)\langle x, y\rangle = \langle A_F x, y\rangle\,.$$

But since this holds for all $x \in D(A_F)$, we conclude that $y \in D(A_F^*) = D(A_F)$ and $Ax = A_F^*x = A_F x$, thus A_F extends A.

Finally, suppose A_0 was another self-adjoint operator extending A and with $D(A_0) \subseteq \mathcal{H}_{s_A}$. Then, similarly, for all $x \in D(A)$ and $y \in D(A_0)$ we find

$$\langle x, A_0 y\rangle = \langle A_0 x, y\rangle = \langle Ax, y\rangle = \langle A_F x, y\rangle\,,$$

from which we again conclude that $y \in D(A_F^*) = D(A_F)$, thus A_F extends A_0, that is $A_0 \subseteq A_F$. Using self-adjointness, we also have $A_F \subseteq A_0$, therefore $A_0 = A_F$. Lastly, density arguments show that the closed form $\overline{s}_A$ has the same lower bound as s_A. Hence, we also have $\langle A_F x, x\rangle \geq -m\|x\|^2$.

□

1.2.4 A min–max principle for the Friedrichs extension

The eigenvalues of an operator can be difficult to compute. A min–max principle or variational principle for an operator provides a variational formula for the computation of

eigenvalues and can be useful for an estimation of eigenvalues even in situations when for example the corresponding eigenfunctions are not known.

Min–max principles are known from linear algebra for symmetric or hermitian matrices on finite dimensional Hilbert spaces, and likewise for the closely related compact operators. As a generalisation there are also min–max principles available for lower semibounded operators as introduced in the previous section.

The formulation of the min–max principle is possible due to the close connection between lower semibounded self-adjoint operators and closed forms.

For the eigenvalues of a self-adjoint extension of a lower semibounded symmetric operator the min–max principle takes a particularly simple form for the Friedrichs extension compared to all other lower semibounded extensions: it only requires knowledge about the domain of the original symmetric operator, not that of the self-adjoint extending operator. This is a useful feature in potential applications as for instance the numerical study of eigenvalues. The reason is the close connection of the Friedrichs extension to the closure of the sesquilinear form induced by the underlying symmetric operator as can be seen in the proof of Proposition 3.

Here we present a derivation of the min–max principle for lower semibounded self-adoint operators and elaborate in particular also on the special case of the Friedrichs extension.

The min–max principle provides information about the **spectrum** of a self-adjoint operator. For a closed linear operator T on a Hilbert space $\mathcal{H}$ the spectrum is defined as the set

$$\sigma(T) \coloneqq \mathbb{C} \setminus \rho(T)$$

where $\rho(T)$ is the **resolvent set** of T:

$$\rho(T) \coloneqq \left\{ \lambda \in \mathbb{C} \;\middle|\; R_\lambda(T) = (T - \lambda\mathbb{1})^{-1} \text{exists and is a bounded operator on all of } \mathcal{H} \right\} .$$

The operator $R_\lambda(T) = (T - \lambda\mathbb{1})^{-1}$ is called the **resolvent** of T. If $\lambda \in \sigma(T)$ is such that $T - \lambda\mathbb{1}$ fails to be injective, then λ is called an **eigenvalue** of T, and $\dim\ker(T - \lambda\mathbb{1})$ is the **multiplicity** of λ. One usually distinguishes two parts of the spectrum: eigenvalues of finite multiplicities that are isolated, also known as **normal eigenvalues** and the rest of the spectrum.

Accordingly, for the for us relevant case of a self-adjoint operator[4], we define the **discrete spectrum** to be

$$\sigma_d(T) \coloneqq \{\lambda \in \mathbb{C} \mid \ker(T - \lambda\mathbb{1}) \neq \{0\}\,, \dim\ker(T - \lambda\mathbb{1}) < \infty, \\ \lambda \text{ is no accumulation point of } \sigma(T)\} \ .$$

The **essential spectrum** of a self-adjoint operator is the complement of the discrete

[4] It is also possible to introduce versions of the notions of discrete and essential spectrum for closed operators, but this is not necessary for our purposes.

spectrum,

$$\sigma_{\text{ess}}(T) = \sigma(T) \setminus \sigma_d(T)\,.$$

Additionally, we mention that for a general self-adjoint operator T it can be shown that the spectrum is real, that is $\sigma(T) \subseteq \mathbb{R}$.

There is another characterisation of the discrete and essential spectrum in the language of *spectral projections* which will prove to be useful in the derivation of the min–max principle below. Let $\Omega \subseteq \mathbb{R}$ be a measurable set. Let us denote the characteristic function of the set Ω by $\mathbb{1}_\Omega$. For a self-adjoint operator, we call the operator

$$P_\Omega(T) = \mathbb{1}_\Omega(T)$$

a **spectral projection** of T, where the right-hand side has a well-defined definition through the functional calculus available through the spectral theorem. The operator $P_\Omega(T)$ is an orthogonal projection for all Ω and the family $\{P_\Omega\}$ forms a projection valued measure (cf. e.g. [RS80a], Section VIII.3) which is the basis for the projection valued measure form of the spectral theorem.

We can reformulate (here without giving a proof) some of the above mentioned notions with the terminology of spectral projections which also will be used in the proof of the min–max principle for a lower semibounded self-adjoint operator below. As a matter of fact, it can be shown that $\lambda \in \sigma(T)$ is in the discrete spectrum $\sigma_d(T)$ of a self-adjoint operator T if and only if

$$\dim \text{Ran} P_{(\lambda-\varepsilon,\lambda+\varepsilon)}(T) < \infty \quad \text{ for some } \varepsilon > 0\,.$$

Accordingly, the essential spectrum is then given by all $\lambda \in \sigma(T)$ for which

$$\dim \text{Ran} P_{(\lambda-\varepsilon,\lambda+\varepsilon)}(T) = \infty \quad \text{ for all } \varepsilon > 0\,.$$

If $\lambda \in \sigma(T)$ is an eigenvalue of T, one finds that $\text{Ran} P_{\{\lambda\}}(T) = \ker(T - \lambda\mathbb{1})$. We call $\text{Ran} P_\lambda(T)$ the **eigenspace** of T corresponding to λ. Its dimension $\dim \text{Ran} P_\lambda(T)$ is the multiplicity of λ as defined above. In that sense, if we assume $\Omega \subseteq \mathbb{R}$ and $\Omega \cap \sigma_{\text{ess}} = \emptyset$ then the number $\dim \text{Ran} P_\Omega(T)$ counts the number of eigenvalues of T in Ω with multiplicity.

Let T be a lower semibounded self-adjoint operator on a Hilbert space $\mathcal{H}$ with $\langle Tx, x\rangle \geq -m\|x\|^2$. For simplicity, we assume $m \geq 0$. The min–max principle for a general lower semibounded self-adjoint operator T stated below in Theorem 2 gives a formula for the eigenvalues of T below the infimum of the essential spectrum. They are given by the **min–max values**. The min–max values for such an operator T are defined by

$$\mu_n(T) = \inf_{\substack{V \subset D(T) \\ \dim V = n}} \sup_{\substack{x \in V \\ x \neq 0}} \frac{\langle Tx, x\rangle}{\|x\|^2}\,. \tag{1.6}$$

In fact, since the supremum is taken over a finite dimensional vector space it is really a maximum. It can be seen quite immediately that the min–max values μ_n satisfy

$$-m \leq \mu_1 \leq \mu_2 \leq \ldots \leq \inf \sigma_{\text{ess}}(T)\,. \tag{1.7}$$

The claim that the min–max values are bounded from below by $-m$ follows directly from the fact that the operator is lower semibounded.

To prove the chain of inequalities in (1.7), let now $\varepsilon > 0$ and $n \in \mathbb{N}$. Assume that $\dim \mathcal{H} \geq n+1$, otherwise there are not more than n min–max values. Then, by definition (1.6), there is an $(n+1)$-dimensional subspace $V \subseteq \mathcal{H}$ such that

$$\mu_{n+1} \geq \max_{\substack{x \in V \\ x \neq 0}} \frac{\langle Tx, x\rangle}{\|x\|^2} - \varepsilon\,.$$

If now $V' \subseteq V$ with $\dim V' = n$ then

$$\mu_{n+1} \geq \max_{\substack{x \in V \\ x \neq 0}} \frac{\langle Tx, x\rangle}{\|x\|^2} - \varepsilon \geq \max_{\substack{x \in V' \\ x \neq 0}} \frac{\langle Tx, x\rangle}{\|x\|^2} - \varepsilon \geq \mu_n - \varepsilon\,.$$

Since ε was arbitrary, we have $\mu_{n+1} \geq \mu_n$. Finally, assuming that $\inf \sigma_{\text{ess}} < \infty$, then by the above characterisation of the essential spectrum we know that $\operatorname{Ran} P_{(\inf \sigma_{\text{ess}} - \varepsilon, \inf \sigma_{\text{ess}} + \varepsilon)}(T)$ is infinite dimensional for all $\varepsilon > 0$. Hence, for $n \in \mathbb{N}$ arbitrarily large and any $\varepsilon > 0$ we can find a subspace $V_n \subseteq P_{(\inf \sigma_{\text{ess}} - \varepsilon, \inf \sigma_{\text{ess}} + \varepsilon)}(\mathcal{H})$ with $\dim V_n = n$ such that for all $x \in V_n$ we have

$$\langle Tx, x\rangle \leq \inf \sigma_{\text{ess}} + \varepsilon\,.$$

It follows that $\mu_n \leq \inf \sigma_{\text{ess}} + \varepsilon$ and since ε was arbitrarily chosen also $\mu_n \leq \inf \sigma_{\text{ess}}$.

With this information in mind, we can now state the well-known min–max principle.

Theorem 2. Let T be a self-adjoint operator on a Hilbert space $\mathcal{H}$, bounded from below, that is $T \geq -m$ and $\sigma(T) \subseteq [-m, \infty)$ for some $m \in \mathbb{R}$. The eigenvalues of T below the essential spectrum, that is, those $\lambda \in \sigma_d(T)$ for which $\lambda < \inf \sigma_{\text{ess}}$, are precisely given by the min–max values $\mu_i(T)$ in the interval $[-m, \inf \sigma_{\text{ess}})$.

Proof. To avoid confusion we assume $m \geq 0$ without loss of generality as before. We show the following: for a given $n \in \mathbb{N}$ we either have $\mu_n < \inf \sigma_{\text{ess}}$ or $\mu_n = \inf \sigma_{\text{ess}}$.

In the first case, T has at least n eigenvalues below the essential spectrum which are given by $\mu_1 \leq \mu_2 \ldots \leq \mu_n$. In the latter case, T has at most $n-1$ eigenvalues below the essential spectrum.

Let $n \in \mathbb{N}$. Consider the case $\mu_n < \inf \sigma_{\text{ess}}$. First, we can notice that by the definition of

μ_n, for any $\varepsilon > 0$, there is a $V \subseteq D(T)$ with $\dim V = n$ such that

$$\mu_n \geq \max_{\substack{x \in V \\ x \neq 0}} \frac{\langle Tx, x \rangle}{\|x\|^2} - \varepsilon \,. \tag{1.8}$$

Suppose now that T has only $k \leq n-1$ eigenvalues below the infimum of the essential spectrum. This means, $\dim \mathrm{Ran} P_{(-\infty, \inf \sigma_{\mathrm{ess}})} = k$. Let $\varepsilon > 0$ and $V \subseteq D(T)$ an n-dimensional space such that (1.8) holds. Denoting the eigenvectors corresponding to these k eigenvalues by $x_1, \ldots, x_k$, we therefore can find a $x \in V$ such that $x \perp \mathrm{span}\{x_1, \ldots, x_k\}$, hence $x \perp \mathrm{Ran} P_{(-\infty, \inf \sigma_{\mathrm{ess}})}$. Then,

$$\frac{\langle Tx, x \rangle}{\|x\|^2} \geq \inf \sigma_{\mathrm{ess}} \,.$$

Combining this with (1.8), we find

$$\mu_n \geq \inf \sigma_{\mathrm{ess}} - \varepsilon \,.$$

Since this holds for all $\varepsilon > 0$ it contradicts the assumption. Hence, there are at least n eigenvalues below the essential spectrum. Let us denote these eigenvalues by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n < \inf \sigma_{\mathrm{ess}}$, counted with multiplicities. We show that $\lambda_i = \mu_i$ for $i = 1, \ldots, n$. By an analogous argumentation as above, when replacing $\inf \sigma_{\mathrm{ess}}$ by λ_i for $i = 1, \ldots, n$ and again making use of (1.8), we can show that from the fact that there are $i-1$ eigenvalues below λ_i it follows that $\mu_i \geq \lambda_i$.

It remains therefore to prove that $\lambda_i \geq \mu_i$ for all $i = 1, \ldots, n$. For any $1 \leq i \leq n$ we know there are i eigenvalues below the essential spectrum. Once again, we denote the eigenvectors corresponding to λ_j by $x_j \in D(T)$ and we set $V = \mathrm{span}\{x_1, \ldots, x_i\}$. The vector space $V \subseteq D(T)$ has dimension i. We then have for any $x \in V$

$$\frac{\langle Tx, x \rangle}{\|x\|^2} = \sum_{j=1}^{i} \lambda_j \frac{|\langle x_j, x \rangle|^2}{\|x\|^2} \,,$$

which is maximised by λ_i for $x = x_i$. Hence, it follows

$$\mu_i \leq \max_{x \in V, x \neq 0} \frac{\langle Tx, x \rangle}{\|x\|^2} \leq \lambda_i \,,$$

for all $1 \leq i \leq n$. This concludes the proof for the case $\mu_n < \inf \sigma_{\mathrm{ess}}$.

Assume now $\mu_n = \inf \sigma_{\mathrm{ess}}$. We show that there are at most $n-1$ eigenvalues below $\inf \sigma_{\mathrm{ess}}$. Suppose there were n such eigenvalues $\lambda_i < \inf \sigma_{\mathrm{ess}}$. Then, as in the previous paragraph, we are able to show that $\mu_n \leq \lambda_n$ which contradicts $\mu_n = \inf \sigma_{\mathrm{ess}}$.

□

For the Friedrichs extension A_F of a lower semibounded symmetric operator A, the min–max principle simplifies in the sense that we can replace the domain of A_F by the domain of $D(A)$ in the definition of the min–max values. The following proposition uses ideas from [FS11].

Proposition 3. Let $A : D(A) \to \mathcal{H}$ be a lower semibounded symmetric operator and A_F its Friedrichs extension as constructed in Theorem 1. Then Theorem 2 holds with min–max values for A_F as defined in (1.6) when replacing $D(A_F)$ by $D(A)$.

Proof. Let $\mu_n(A_F)$ be as defined in (1.6) and

$$\mu_n' = \inf_{\substack{V' \subseteq D(A) \\ \dim V' = n}} \max_{\substack{x \in V' \\ x \neq 0}} \frac{\langle Ax, x \rangle}{\|x\|^2}$$

be the min–max values of A_F when replacing $D(A_F)$ by $D(A)$. The strategy for proving this proposition is to show that for any $0 < \varepsilon < 1$

$$\mu_n' \leq \mu_n + C\varepsilon$$

for a positive constant C. Since by definition $\mu_n \leq \mu_n'$, it then follows $\mu_n' = \mu_n$.

As a first step we notice as before that for a given $0 < \varepsilon < 1$, there is an n-dimensional space $V \subseteq D(A_F)$ such that

$$\mu_n \geq \max_{\substack{x \in V \\ x \neq 0}} \frac{\langle A_F x, x \rangle}{\|x\|^2} - \varepsilon \tag{1.9}$$

holds (as in equation (1.8)). Let $x_1, \ldots, x_n$ be an orthonormal basis for V. Using the notation of Theorem 1 we have $D(A) \subseteq D(A_F) \subseteq \mathcal{H}_{s_A}$, where $\mathcal{H}_{s_A}$ is the domain of the closure of the sesquilinear form induced by A. The domain $D(A)$ is by construction a dense subset of $\mathcal{H}_{s_A}$ with respect to the norm $\|\cdot\|_s$ on $\mathcal{H}_{s_A}$ defined as in (1.3) for the closed form s_A. Hence, there are $x_1', \ldots, x_n' \in D(A)$ such that $\left\|x_j - x_j'\right\|_s \leq \varepsilon$ for $1 \leq j \leq n$. Moreover,

$$\begin{aligned} |\langle x_i', x_j' \rangle| &\leq \delta_{ij} + |\langle x_i', x_j' \rangle - \langle x_i, x_j \rangle| \\ &\leq \delta_{ij} + |\langle x_i' - x_i, x_j' \rangle| + |\langle x_i, x_j' - x_j \rangle| \\ &\leq \delta_{ij} + \|x_i' - x_i\| \left\|x_j'\right\| + \|x_i\| \left\|x_j' - x_j\right\| \\ &\leq \delta_{ij} + \varepsilon(1 + \varepsilon) + \varepsilon \leq \delta_{ij} + 3\varepsilon\,, \end{aligned} \tag{1.10}$$

where in the second to last step we used $\|\cdot\| \leq \|\cdot\|_s$. This means that the vectors x_i' form an 'almost orthonormal' set and for ε small enough they are linearly independent. They thus span an n-dimensional subspace of $D(A)$, we denote it by V', and there are $\alpha_j \in \mathbb{C}$

such that $x_0' = \sum_{j=1}^n \alpha_j x_j'$ and

$$\max_{x\in V', x\neq 0} \frac{\langle Ax, x\rangle}{\|x\|^2} = \frac{\langle Ax_0', x_0'\rangle}{\|x_0'\|^2}\,.$$

If we are able to show that for $x_0 \coloneqq \sum_{j=1}^n \alpha_j x_j$

$$\left|\frac{\langle Ax_0', x_0'\rangle}{\|x_0'\|^2} - \frac{\langle A_F x_0, x_0\rangle}{\|x_0\|^2}\right| \leq C\varepsilon \tag{1.11}$$

for a number $C > 0$ (that might depend on μ_n), then it follows

$$\mu_n' \leq \max_{x\in V', x\neq 0} \frac{\langle Ax, x\rangle}{\|x\|^2} \leq \max_{x\in V, x\neq 0} \frac{\langle A_F x, x\rangle}{\|x\|^2} + C\varepsilon \leq \mu_n + (C+1)\varepsilon$$

by (1.9). Since this holds for any $0 < \varepsilon < 1$, we then have proven $\mu_n' \leq \mu_n$. In the remainder of this proof we show (1.11). Without loss of generality we can assume that $\|x_0\|^2 = \sum_{i=1}^n |\alpha_j|^2 = 1$. Then we find

$$\begin{aligned}
&\left|\frac{\langle Ax_0', x_0'\rangle}{\|x_0'\|^2} - \frac{\langle A_F x_0, x_0\rangle}{\|x_0\|^2}\right| \\
&= \left|\sum_{j,k=1}^n \overline{\alpha}_j \alpha_k \left(\frac{1}{\|x_0'\|^2}\langle Ax_j', x_k'\rangle - \langle A_F x_j, x_k\rangle\right)\right| \\
&\leq \left|\sum_{j,k=1}^n \overline{\alpha}_j \alpha_k \left(\langle Ax_j', x_k'\rangle - \langle A_F x_j, x_k\rangle\right)\right| + \left|\sum_{j,k=1}^n \overline{\alpha}_j \alpha_k \left(\frac{1}{\|x_0'\|^2} - 1\right)\langle Ax_j', x_k'\rangle\right|\,.
\end{aligned}$$

First of all one can note that

$$\|x_i\|_s = \sqrt{\langle A_F x_i, x_i\rangle + (m+1)\langle x_i, x_i\rangle} \leq \sqrt{\mu_n + \varepsilon + (m+1)}$$

by (1.9). Hence,

$$\|x_i'\|_s \leq \sqrt{\mu_n + \varepsilon + (m+1)} + \varepsilon$$

and thus using (1.4)

$$\begin{aligned}
&\left|\langle Ax_j', x_k'\rangle - \langle A_F x_j, x_k\rangle\right| \\
\leq &\left|\langle A_F(x_j' - x_j), x_k'\rangle\right| + \left|\langle Ax_j, x_k' - x_k\rangle\right| \\
\leq &\left\|x_j' - x_j\right\|_s \|x_k'\|_s + \|x_j\|_s \|x_k' - x_k\|_s \\
\leq &2\varepsilon(\sqrt{\mu_n + \varepsilon + (m+1)} + \varepsilon) \leq 2(\sqrt{\mu_n + m + 2} + 1)\varepsilon\,.
\end{aligned}$$

Moreover, we also find

$$\begin{aligned}\left|\|x_0'\|^2 - 1\right| &\leq \sum_{j,k=1}^{n} |\alpha_j \alpha_k| |\langle x_j', x_k'\rangle - \langle x_j, x_k\rangle| \\ &\leq 3\varepsilon \sum_{j,k=1}^{n} |\alpha_j \alpha_k| \leq 3\varepsilon n \sum_{j=1}^{n} |\alpha_j|^2 = 3\varepsilon n\,,\end{aligned}$$

using what we derived in (1.10) and the Cauchy-Schwarz inequality. Altogether it follows

$$\begin{aligned}&\left|\frac{\langle Ax_0', x_0'\rangle}{\|x_0'\|^2} - \frac{\langle A_F x_0, x_0\rangle}{\|x_0\|^2}\right| \\ &\leq \sum_{j,k=1}^{n} |\alpha_j \alpha_k| \left|\langle Ax_j', x_k'\rangle - \langle A_F x_j, x_k\rangle\right| + \left|\frac{1}{\|x_0'\|^2} - 1\right| \sum_{j,k=1}^{n} |\alpha_j \alpha_k| \left|\langle Ax_j', x_k'\rangle\right| \\ &\leq 2n(\sqrt{\mu_n + m + 2} + 1)\varepsilon + \frac{3n\varepsilon}{1 - 3n\varepsilon}\left(\sqrt{\mu_n + \varepsilon + (m+1)} + \varepsilon\right)^2 n \\ &\leq C\varepsilon\end{aligned}$$

with $C = 2n(\sqrt{\mu_n + m + 2} + 1) + \frac{3n^2}{1-3n}\left(\mu_n + m + 3 + \sqrt{\mu_n + m + 2}\right)$.

□

1.3 Supplementary remarks: Self-adjoint operators in physics

Quantum Hamiltonian operators played a special motivational role for Friedrichs in his work on a self-adjoint extension of lower semibounded operators. This is connected to stability: A system whose dynamics is determined by an energy operator that is lower semibounded has a lowest possible energy[5], and is thus stable. The Hamiltonians we come across in the non-relativistic case, also known as Schrödinger operators, are usually lower semibounded.

A more basic question one might ask from the physical point of view is: why is it actually important to deal with the task of finding and classifying self-adjoint extensions of an operator?

Some thoughts that do at least partially answer this question are presented below, where we first want to focus on why one cannot avoid to specify the domain of an operator and later on looking specifically into the need for self-adjoint operators. Inspiration for this section was found amongst others in [CM21].

[5] which we call the ground state energy even if there might not exist a ground state.

1.3.1 The domain of an operator

The basic building blocks of a mathematical formulation of quantum mechanics are a separable complex Hilbert space $(\mathcal{H}, \langle ., . \rangle_{\mathcal{H}})$ whose elements ψ, modulo their length, are interpreted as physical states, as well as linear operators A on this Hilbert space corresponding to observables, physical, measurable quantities. The expectation, or average value $\langle \psi \rangle_A$ for the result of a measurement of a system in state ψ of an observable A is given by

$$\langle \psi \rangle_A = \langle \psi, A\psi \rangle \, .$$

In order to obtain a meaningful interpretation of this quantity, one demands the expectation to be a real number, which is equivalent to assuming A to be a symmetric operator.

On a formal level it is not really necessary to think further about the properties of the operator A and we can specify it only by its action on elements of the Hilbert space. At a closer look things are slightly more involved.

One of the most prominent examples that indicates that the question of an operator domain is non-trivial also in physics, is the momentum operator $\hat{P}$. In the position space representation in one dimension, where we represent physical states by elements of the Hilbert space $L^2(\mathbb{R})$, the momentum operator takes the form $\hat{P} = -\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}$ (setting Planck's constant $\hbar = 1$). This operator is easily seen to be 'formally symmetric'. But neither is every function in $L^2(\mathbb{R})$ differentiable[6], nor is it generally true that $\hat{P}f$ for $f \in L^2(\mathbb{R})$ is again square-integrable. Thus, $\hat{P}$ has no proper meaning on all of the Hilbert space $L^2(\mathbb{R})$, but as is easily seen it is still a well-defined operator on the dense subset $\mathcal{C}_0^\infty(\mathbb{R}) \subseteq L^2(\mathbb{R})$, on the smooth functions with compact support.

The fact that $\hat{P}$ cannot be defined on the whole Hilbert space is a complication we find for many operators in quantum mechanics, more precisely for all those operators that are unbounded. As long as we only deal with bounded operators, i.e. operators for which

$$\sup_{\substack{\psi \in \mathcal{H} \\ \psi \neq 0}} \frac{\|A\psi\|}{\|\psi\|} < \infty \, ,$$

we do not need to care much about the domain of an operator. Any bounded linear operator, defined on a dense subset of a Hilbert space $\mathcal{H}$ can be uniquely extended to a bounded operator on all of $\mathcal{H}$ (cf. e.g. the B.L.T theorem: Theorem I.7 in [RS80a]) and we can thus assume that the bounded linear operator is defined on the whole space.

For the unbounded, symmetric operators which, as the example of the momentum operator shows, are present in the established theory of quantum mechanics[7], however, it is not only uncertain if they are defined on all of the Hilbert space, but it is in fact

[6] Consider e.g. a *simple function*.

[7] Some thoughts on why unbounded operators might be necessary to formulate such a theory are found in [CM21].

impossible. By the Hellinger–Toeplitz theorem (see e.g. Theorem 2.10 in [Tes14]) an unbounded, symmetric operator cannot be defined on all of the Hilbert space, but has a domain which is a proper subset of the Hilbert space only. In fact, there can be several sets which possibly can serve as a domain for an operator, and also from a physical point of view there is in general no known canonical way to select a distinguished domain as the physically relevant one.

It might in this context sound like a reasonable attempt to define an operator on the set of all Hilbert space elements on which it is well-defined, i.e. on its maximal domain. However, this quickly leads to contradictions since the resulting operator is not necessarily symmetric anymore (cf. Section 2.4. in [CM21]).

Instead, it is possible to demand the domain of a physical observable to be 'maximally symmetric', i.e. to define the operator on a subspace as large as possible such that it is still symmetric on that space. This approach is for example taken by Teschl (see [Tes14], p. 65, Axiom 2 and Corollary 2.12). Together with an additional assumption on how to represent polynomials of observables this demand leads to the requirement that observables are self-adjoint operators in quantum mechanics. It is important to notice though, that maximally symmetric does not mean we can find one specific largest domain for the operator. In general, the possible domains which allow the operator to be symmetric cannot be ordered by inclusion and we can in many cases still find many different (possibly infinitely many) self-adjoint operators represented by the same formal action.

The concerns about the domain of an operator are not only important from a technical point of view. Different domains can yield completely different operators, with distinct spectral properties. Since we interpret eigenvalues as possible measurement values of pure quantum states, these considerations also have a practical physical relevance.

1.3.2 The demand for self-adjoint operators

The idea that observables in the theory of quantum mechanics should be represented by self-adjoint operators can be supported by further arguments. At least for the Hamiltonian operator there is a clear physical motivation for demanding it to be self-adjoint.

The reason is its special role as the generator of time-translations via the Schrödinger equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\psi(t) = H\psi(t)\,. \tag{1.12}$$

By H we denote the Hamiltonian of the system. Given any state of a quantum system ψ_0, its evolution in time $\psi(t)$ is, as one usually assumes, bound to satisfy certain properties:

Assuming the time evolution $\psi(t)$ is uniquely determined by the Schrödinger initial value problem with $\psi(0) = \psi_0$, we write $\psi(t) = U(t)\psi_0$ for a time-dependent operator $U(t)$. The superposition principle then already tells us that $U(t)$ should be a linear operator. Furthermore, we usually demand the state vector's normalisation to be preserved

throughout time, thus requiring

$$\|\psi_0\|^2 = \|\psi(t)\|^2 = \|U(t)\psi_0\|^2 = \langle \psi_0, U^*(t)U(t)\psi_0 \rangle \,, \tag{1.13}$$

that is, $U(t)$ is an isometry. Additionally assuming that $U(t)$ is surjective, means $U(t)$ is a unitary operator.

For any two times $t, s \in \mathbb{R}$ we then also have that

$$U(t+s) = U(t)U(s) \tag{1.14}$$

as well as $U(0) = \mathbb{1}$. And finally, one usually requires strong continuity, that is

$$\lim_{t\to 0} \|U(t)\psi_0 - \psi_0\| \to 0 \,. \tag{1.15}$$

Summarising (1.13), (1.14), (1.15), we find that the time evolution of a state ψ_0 should be determined by a family of unitary operators $\{U(t) \mid t \in \mathbb{R}\}$ that forms a *one-parameter strongly continuous group*.

The crucial observation is that such a one-parameter strongly continuous group always has a generating operator H defined by

$$H\psi = -\mathrm{i} \lim_{\varepsilon\to 0} \frac{U(\varepsilon)\psi - \psi}{\varepsilon} \tag{1.16}$$

on a set of all $\psi \in \mathcal{H}$ for which this limit exists. Furthermore, this generator is a self-adjoint operator. This statement is Stone's theorem (Theorem VIII.8 in [RS80a]). It then holds $U(t) = \mathrm{e}^{-\mathrm{i}tH}$ by the defining relation of H in (1.16) and H is the Hamiltonian operator of a corresponding Schrödinger equation (1.12).

Conversely, any self-adjoint operator H generates a one-parameter strongly continuous group $\left\{U(t) \;\middle|\; U(t) = \mathrm{e}^{-\mathrm{i}tH}, t \in \mathbb{R}\right\}$(this is proved e.g. in [CM21], Theorem 3.1) such that a state's time evolution $\psi(t) = U(t)\psi_0$ is a solution to the initial value problem posed by the Schrödinger equation with $\psi(0) = \psi_0$. A key point is that for a self-adjoint operator H, in contrast to a merely symmetric H, we have the functional calculus at hand to make sense of the operator $\mathrm{e}^{-\mathrm{i}tH}$ which is a priori only a formal solution to the Schrödinger equation.

In that sense, there is an equivalence between a self-adjoint Hamiltonian and a unique, well-defined solution to the Schrödinger inital value problem with a time evolution of states that we consider as meaningful seen from the physical point of view. For the Hamiltonian operator at least there is thus good reasoning for demanding self-adjointness.

For general observables, there are further arguments which justify the demand for self-adjointness. As an example, there are some listed in [CM21].

1.3.3 The Friedrichs extension in physics

Even when one is clear about the demand for self-adjoint operators, it is in general not obvious which of the self-adjoint realisations of an operator one wants to choose for the physical theory.

For a symmetric and lower semibounded operator there is always a distinguished self-adjoint extension available by an application of the Friedrichs' construction.

It stands out by the universal applicability of this construction to a large class of operators. Apart from that there are further properties that distinguish it from other extensions: by the min–max principle its eigenvalues below the essential spectrum are always larger than or equal to those of other self-adjoint extensions. If we assume that the Friedrichs extension is the right operator to describe a physical system we are thus assuming that the energy eigenvalues should be the highest possible. In some sense one could argue that this operator is 'furthest away from instability'.[8]

Another argument that speaks in favour of the distinction of the Friedrichs extension in physics is e.g. given by Kalf, Schmincke, Walter, Wüst in [KSWW]. For a Schrödinger operator H with a potential satisfying certain properties defined on the set of smooth functions with compact support they make the following observation: the domain of the Friedrichs extension $D(H_F)$ consists of exactly those elements in $D(H^*)$ for which the energy expectation value is finite. This can be directly seen by a comparison of $D(H_F)$ with the domain of the adjoint $D(H^*)$ (cf. (1.1)). It distinguishes the Friedrichs extension from other extensions since it is by Theorem 1 the only extension whose domain is contained in the closed form domain.

1.4 Operators with a gap

1.4.1 Self-adjoint extensions of gapped operators

As large as the class of symmetric operators is to which one can apply Friedrichs' construction of a self-adjoint extension, there are naturally very many other operators that are not of this type. Friedrichs' extension does not apply to operators that are not lower semibounded and besides, the min–max principle only characterises the eigenvalues for a self-adjoint lower semibounded operator below the essential spectrum. Any parts of the discrete spectrum that might appear above some essential spectrum are not covered.

It is therefore natural to ask for generalisations of the Friedrichs extension, or at least for the existence of a distinguished self-adjoint extension for a more general class of operators.

[8] By Krein's theory of self-adjoint extension for lower semibounded operators the Friedrichs extension is indeed the one extreme side of a class of self-adjoint extensions: All other self-adjoint extension lie in between the Friedrichs extension and the extension known as Krein's extension which is the one closest to the symmetric operator, cf. Section 1.4.1.

This can for example consist of operators satisfying a gap condition. There are different ways to formulate such a gap condition.

Krein [Kre47] developed a theory of self-adjoint extensions for lower semibounded operators and formulated a generalisation to the gapped case at the same time. His results are partly reviewed in [BN94] by Brasche and Neidhardt.

That a lower semibounded operator has at least one lower semibounded self-adjoint extension was already known by Friedrichs' work. Krein developed this theory further and was able to show that there exists a whole family of lower semibounded extensions that all fall in between two distinguished extreme extensions: the soft and hard extension, nowadays also known as the *Krein–von Neumann extension* and the Friedrichs extension, respectively. The Friedrichs extensions is by this theory the largest of all lowerbounded self-adjoint extensions and the Krein–von Neumann the smallest.

Krein's approach to a generalisation for operators with a gap is based on the following definition of a gapped operator: He defined a symmetric operator A on a Hilbert space to have a gap if there exist $-\infty < a < b < \infty$ such that for all $x \in D(A)$ it holds

$$\left\| \left(A - \frac{a+b}{2} \right) x \right\| \geq \frac{b-a}{2} \|x\| \,. \tag{1.17}$$

An elementary reformulation of (1.17),

$$\|Ax\|^2 - (a+b)\langle Ax, x\rangle + ab\|x\|^2 \geq 0 \tag{1.18}$$

emphasises in which sense this gap condition can be thought of as a generalisation of lower semiboundedness. Dividing (1.18) by a and then letting $a \to -\infty$ we rediscover the condition for lower semiboundedness with lower bound b.

Krein's analysis showed that for a symmetric operator with a gap there is always at least one self-adjoint condition that preserves the gap. Brasche and Neidhardt's aim in [BN94] was to parametrise the gap-preserving self-adjoint extensions based on Krein's extension theory for semibounded operators. In particular they explicitly mark two extensions in the gap-theory that should correspond to Friedrichs and Krein-von Neumann extension in the semibounded theory. They identify the distinguished self-adjoint extensions for the gapped case as limits of the Friedrichs and Krein–von Neumann extension for the gap parameter $a \to -\infty$, respectively.

In the paper [SST20], we construct a self-adjoint extension for a densely defined, symmetric operator A on a Hilbert space $\mathcal{H}$ satisfying a certain gap condition. We see this construction as an analogue to the Friedrichs extension in the lower semibounded case. The gap condition we consider in [SST20] differs from the gap introduced by Krein. In particular, it is based on an orthogonal decomposition of the underlying Hilbert space and the resulting self-adjoint extension we construct is dependent on the additional structure of the Hilbert space. In detail, the assumptions we make are the existence of orthogonal

projections on $\mathcal{H}$ with $\Lambda_+ + \Lambda_- = \mathbb{1}_{\mathcal{H}}$ such that

$$\mathcal{H} = \Lambda_+\mathcal{H} \oplus \Lambda_-\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$$

as well as

$$F_\pm := \Lambda_\pm D(A) \subset D(A)\,,$$

for $D(A)$ being the domain of A. In particular, this means that the operator can be written in a block structure

$$A = \begin{pmatrix} A_{++} & A_{+-} \\ A_{-+} & A_{--} \end{pmatrix},$$

where $A_{++} = \Lambda_+ A|_{\mathcal{H}_+}$, $A_{+-} = \Lambda_+ A|_{\mathcal{H}_-}$ et cetera. The gap condition we impose then reads

$$\sup_{y_-\in F_-\setminus\{0\}} \frac{\langle y_-, Ay_-\rangle_{\mathcal{H}}}{\|y_-\|^2_{\mathcal{H}}} =: \lambda_0 < \lambda_1 := \inf_{x_+\in F_+\setminus\{0\}} \sup_{y_-\in F_-} \frac{\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}}}{\|x_+ + y_-\|^2_{\mathcal{H}}}\,. \quad (1.19)$$

This gap condition is not purely inherent in the symmetric operator as opposed to the more general Krein condition since it presupposes the splitting of the Hilbert space. Krein's gap condition is a more general notion: it can be seen that our gap condition does imply Krein's condition as is shown in Remark 3 of the paper.

The lower semibounded case can in our model at least be formally rediscovered when taking $\lambda_0 \to -\infty$. In this case the block structure vanishes, leaving us with F_+ being the entire domain of A. The gap condition (1.19) then assures the lower semiboundedness of A. Likewise, our construction of a self-adjoint extension is essentially Friedrichs' construction in the case of a vanishing block structure of A.

1.4.2 The Coulomb–Dirac operator

With the Dirac equation for relativistic quantum mechanics, the physical theory provides us with an example of an operator that is not bounded from below but has a gap. Specifically the Dirac operator for massive particles with Coulomb potential in three dimensions has been in focus in the work of Esteban, Loss, Dolbeault and Séré ([EL07],[EL],[DES00]) that has been inspiration for our construction for operators with a gap.

In three dimensions the Hamiltonian for a free relativistic fermion is the free Dirac operator. Acting on the Hilbert space $\mathcal{H}_0 = L^2(\mathbb{R}^3;\mathbb{C}^4)$ (setting $c = 1$ and $\hbar = 1$) it takes the form

$$H_0 = -\mathrm{i}\boldsymbol{\alpha}\cdot\nabla + m\beta$$

where m is the particle mass. We set $m = 1$ in our paper. Furthermore, β is a complex 4×4 matrix and the symbol $\boldsymbol{\alpha}$ denotes the 3-dimensional vector of complex 4×4-matrices

$\alpha_1, \alpha_2, \alpha_3$ that satisfy

$$\begin{aligned} \left\{\alpha^i, \alpha^j\right\} &= 2\delta_{ij}\mathbb{1} \qquad & i,j &= 1,2,3 \\ \left\{\alpha^i, \beta\right\} &= 0 & i &= 1,2,3 \\ \beta^2 &= \mathbb{1}\,. \end{aligned}$$

Multiplying the Schrödinger equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\psi(t) = H_0\psi(t)$$

from the left by β we reproduce the Dirac equation in its common form

$$\mathrm{i}\gamma^\mu\partial_\mu\psi - m\psi = 0$$

with gamma matrices γ^μ, $\mu = 0,\ldots,3$ with $\gamma^0 = \beta$, $\gamma^j = \beta\alpha^j$, $j = 1,2,3$. By their anticommutation relations $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ the gamma matrices generate a matrix representation of the Dirac algebra, that is, of the Clifford algebra $Cl_{1,3}(\mathbb{R})$.

There are different representations for the matrices α^i and β. In the standard representation known as the Dirac basis the matrices take the form

$$\alpha^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix}, \quad i = 1,2,3\,, \qquad \beta = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix},$$

with σ^i being the Pauli matrices. In this representation, the free Dirac operator then has the block form

$$H_0 = \begin{pmatrix} m & -\mathrm{i}\boldsymbol{\sigma}\cdot\nabla \\ -\mathrm{i}\boldsymbol{\sigma}\cdot\nabla & -m \end{pmatrix},$$

with $\boldsymbol{\sigma}$ being the vector of Pauli matrices $\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$. It is easily seen that H_0 is symmetric on the dense domain $\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4) \subseteq \mathcal{H}_0$. It is furthermore a well-known fact (cf. e.g. Theorem 1.1 in [BT92]) that H_0 is essentially self-adjoint on this domain[9] and its closure is an operator H_0 defined on the Sobolev space $H^1(\mathbb{R}^3;\mathbb{C}^4)$, where it is self-adjoint.

The spectrum of the self-adjoint operator H_0 has no discrete part and is given by

$$\sigma(H_0) = (-\infty, -m] \cup [m, \infty)\,.$$

We can see that in the case of the free Dirac operator, there is a spectral gap taking up the full interval $(-m, m)$, thereby separating positive from negative energy states. The fact that the Dirac operator is not bounded from below is reflected in the existence of negative energy states with arbitrary negative energy. To make physical sense of this operator,

[9] This holds equally if we consider the operator on $\mathcal{C}_0^\infty(\mathbb{R}^3 \setminus \{0\};\mathbb{C}^4)$.

Dirac used the idea to interpret the negative energy states as those that can be taken by antiparticles.

Adding a Coulomb potential to H_0 defines the Coulomb–Dirac operator

$$H_\nu = H_0 - \frac{\nu}{|x|}\,,$$

where the parameter ν in the atomic case represents the constant $\nu = Z\alpha$, with Z being the atomic number and α the fine structure constant whose value is approximately given by $\alpha \approx 1/137$. The operator is symmetric on $\mathcal{C}_0^\infty(\mathbb{R}^3 \setminus \{0\}; \mathbb{C}^4)$.

During the last decades extensive research has been done on the existence of self-adjoint extensions of the Coulomb–Dirac operator, also raising the question about the existence of a (physically) distinguished one.

We collect some known results, citing from [BT92], [Gal], [GM18],[BE11], [Nen76], [EL07], [EL], [DES00], a list which is naturally not exhaustive. Some remarks on this topic are also found on page 3 of our paper [SST20].

The existence and nature of self-adjoint extensions of H_ν depends on the value of ν. But for any self-adjoint realisation of H_ν and value of ν it holds that the essential spectrum remains stable under the perturbation of H_0 by the Coulomb term. That is,

$$\sigma_{\text{ess}}(H_\nu) = (-\infty, -m] \cup [m, \infty)$$

and the discrete spectrum is entirely contained in the interval $(-m, m)$,

$$\sigma_d(H_\nu) \subseteq (-1, 1)\,.$$

As is shown in [BT92] in Section 4.3.4., this follows from a theorem by Weyl about the stability of the essential spectrum (cf. for example Theorem 8.12 in [Sch12]).

The interesting question about the spectrum of the Dirac–Coulomb operator is therefore about the eigenvalues in the spectral gap between $-m$ and m. We distinguish a number of regimes, depending on the value for ν:

i) The sub-critical regime $|\nu| \leq \sqrt{3}/2$ (corresponding to $Z \leq 118$):

   The operator H_ν is essentially self-adjoint on the domain $\mathcal{C}_0^\infty(\mathbb{R}^3 \setminus \{0\}; \mathbb{C}^4)$ and has unique self-adjoint extension, cf. e.g. Example 4.17 in [BT92]. For $\nu < 1/2$ this can be directly proved by viewing the Coulomb potential as a perturbation of the free Dirac operator via the Kato-Rellich theorem. For $1/2 \leq \nu \leq \sqrt{3}/2$ generalisations or other approaches are needed.

   The Dirac equation can be solved explicitly by a partial wave decomposition from which one finds that the Dirac–Coulomb operator is unitarily equivalent to the direct sum of radial symmetric operators. This solution includes an explicit formula for the

discrete energy eigenvalues which takes the form

$$E_{n,\kappa} = m\left(1 + \frac{\nu^2}{(n + \sqrt{\kappa^2 - \nu^2})^2}\right)^{-1/2}, \qquad n \in \mathbb{N}_0 \tag{1.20}$$

where κ takes values in $\mathbb{N}$. These are exactly the eigenvalues of the (unique) self-adjoint realisation of H_ν in the gap of the essential spectrum (cf. Theorem 3.1.5 in [BE11]).

ii) The critical regime $\sqrt{3}/2 < |\nu| < 1$ ($118 \leq Z \leq 137$):

The operator H_ν has several self-adjoint extensions. There is a distinguished one pointed out in the literature. It has been constructed independently in different ways ([Wüs75], [Wüs77], [Nen76], [Sch72]) and has the property that its domain is contained in $D(|H_0|^{1/2})$ as well as in $D(H_\nu^*) \cap D(|x|^{-1/2})$, that is, the elements in the domain have individually finite kinetic and potential energy.

Esteban and Loss [EL07] reconstructed this distinguished self-adjoint extension by a different approach, using a Hardy-like inequality.

The formula (1.20), which in the critical range of ν can be derived by implementing certain boundary conditions for the radial wave functions at $|x| = 0$, gives also in this regime the eigenvalues of the distinguished self-adjoint extension of H_ν (Remark 3.1.6 in [BE11], a derivation is also found in [GM18]).

iii) The critical case $|\nu| = 1$:

Esteban and Loss [EL07] showed that their construction can be extended up to and including the case $\nu = 1$, which by the earlier known approaches had not been possible. Their self-adjoint extension for $\nu = 1$ evolves as a norm-resolvent limit from the distinguished extension for $\nu < 1$, cf. [ELS19]. This gives us good reason to interpret their extension for H_1 as the distinguished one. The domain of this self-adjoint extension consists of elements that have finite total energy, even if kinetic and potential energy are in general not finite individually.

Since the discrete eigenvalues converge under a norm-resolvent limit, we find the eigenvalues for the distinguished extension again by formula (1.20).

iv) $|\nu| > 1$ (corresponding to $Z > 137$):

The eigenvalue formula (1.20) in this case yields complex values which indicates that it is related to a non-self-adjoint extension. This appears to be corresponding to the usual assumption that an atomic number $Z = 137$ is the highest number possible for a single nucleus atom. One can find self-adjoint extensions by regularising the Coulomb singularity (cf. [BT92], Section 7.4.5).

The construction we present in [SST20] is applicable to the Dirac–Coulomb operator, as we elaborate in Section 3 of [SST20].

1.4.3 Perspectives for future research

There are naturally many questions related to our work that remain unanswered our paper. We state some of the potential topics for further investigation:

- In Section 1.4.1 we mentioned the study of Brasche and Neidhardt [BN94] who classified the self-adjoint extensions in Krein's gap theory and pointed out a distinguished one corresponding to a generalisation of the Friedrichs extension. It could be an interesting task to further investigate the exact relation of the extension that Brasche and Neidhardt call the Friedrichs extension for gapped operators and our extension.

- As has been pointed out in Section 1.4.1, our gap condition relies on an orthogonal decomposition of the underlying Hilbert space which is a rather strong premise. The choice of the splitting might influence the nature of the extension significantly. Krein's gap condition on the other hand is far more general. It might be worth to make an attempt to weaken the assumption we make on the operator and find a gap condition that is more general, even if possibly still more restrictive than Krein's.

- In Section 4 of our paper [SST20], we relate our self-adjoint extension for a Dirac-type operator to a special type of boundary conditions, known as the APS boundary condition. By the assumptions we make our considerations are applicable to the situation with an underlying geometry of cylindrical type. It could be interesting to investigate the possibilities to extend these results to more general shapes.

  Besides that we do not exclude that other boundary conditions may also play a role in relation to the special self-adjoint extension we present. This is a question that would need more and deeper-going research.

Bibliography I

[BE11] Alexander A. Balinsky and William Desmond Evans, *Spectral Analysis of Relativistic Operators*, Imperial College Press, London, 2011.

[BN94] Johannes-Friedemann Brasche and Hagen Neidhardt, *Some Remarks on Krein's extension theory*, Mathematische Nachrichten **165** (1994), 159-181.

[CM21] Andrea Cintio and Alessandro Michelangeli, *Self-adjointness in Quantum Mechanics: a pedagogical path* (2021), available at `https://arxiv.org/abs/2012.14490`.

[DES00] Jean Dolbeault, Maria J. Esteban, and Eric Séré, *On the Eigenvalues of Operators with Gaps. Application to Dirac Operators*, Journal of Functional Analysis **174** (2000), 208–22.

[ELS19] Maria J. Esteban, Mathieu Lewin, and Eric Séré, *Domains for Dirac-Coulomb min–max levels*, Revista Matemática Iberoamericana **35** (2019), no. 3, 877–924.

[EL07] Maria J. Esteban and Michael Loss, *Self-adjointness for Dirac operators via Hardy-Dirac inequalities*, Journal of Mathematical Physics **48** (2007), no. 11, 112107.

[EL] Maria J. Esteban and Michael Loss, *Self-adjointness via partial Hardy-Dirac inequalities*, Mathematical Results in Quantum Mechanics, WORLD SCIENTIFIC, 2008, pp. 41–47.

[FS11] Søren Fournais and Jan Philip Solovej, *Lecture notes MatPhys 2, The min-max theorem*, 2011. an earlier version is available at `http://web.math.ku.dk/~solovej/MANYBODY/minmax.pdf`.

[Fri34] Kurt Friedrichs, *Spektraltheorie halbbeschränkter Operatoren und Anwendung auf die Spektralzerlegung von Differentialoperatoren*, Mathematische Annalen **109** (1934), 465–487.

[Gal] Matteo Gallone, *Self-Adjoint Extensions of Dirac Operator with Coulomb Potential*, Advances in Quantum Mechanics: Contemporary Trends and Open Problems (Alessandro Michelangeli and Dell'Antonio Gianfausto, eds.), Springer INdAM Series, Springer Nature, 2017.

[GM18] Matteo Gallone and Alessandro Michelangeli, *Discrete spectra for critical Dirac–Coulomb Hamiltonians*, Journal of Mathematical Physics **59** (2018), 062108.

[KSWW] Hubert Kalf, Upke-Walther Schmincke, Johann Walter, and Rainer Wüst, *On the spectral theory of Schrödinger and Dirac operators with strongly singular potentials*, Symposium (Dundee, Scotland, 1974), Spectral Theory and Differential Equations, Lecture Notes in Mathematics, vol. 448, Springer-Verlag, Berlin, Heidelberg, New York, 1975, pp. 182–226.

[KW79] M. Klaus and Rainer Wüst, *Characterization and uniqueness of distinguished selfadjoint extensions of Dirac operators*, Communications in Mathematical Physics **64** (1979), no. 2, 171–176.

[Kre47] Mark G. Krein, *The theory of self-adjoint extensions of semi-bounded Hermitian transformations and its applications. I*, Matematicheskii Sbornik **20** (1947), no. 3, 431-495. in Russian, English summary from page 491.

[Nen76] Gheorghe Nenciu, *Self-adjointness and invariance of the essential spectrum for Dirac operators defined as quadratic forms*, Communications in Mathematical Physics **48** (1976), no. 3, 235–247.

[RS80a] Michael Reed and Barry Simon, *Methods of Modern Mathematical Physics*, Vol. 1, Academic Press, Inc., Harcourt Brace Jovanovich, Publishers, 1980.

[RS80b] ———, *Methods of Modern Mathematical Physics*, Vol. 2, Academic Press, Inc., Harcourt Brace Jovanovich, Publishers, 1980.

[SST20] Lukas Schimmer, Jan Philip Solovej, and Sabiha Tokus, *Friedrichs Extension and Min-Max Principle for Operators with a Gap*, Annales Henri Poincaré **21** (2020), 327-357.

[Sch72] Upke-Walther Schmincke, *Essential selfadjointness of Dirac operators with a strongly singular potential*, Mathematische Zeitschrift **126** (1972), 71–81.

[Sch12] Konrad Schmüdgen, *Unbounded Self-adjoint Operators on Hilbert Space*, Graduate Texts in Mathematics, vol. 265, Springer Dordrecht Heidelberg New York London, 2012.

[Sol14] Jan Philip Solovej, *Many Body Quantum Mechanics*, 2014. lecture notes, draft, with corrections and additions of P. T. Nam. An earlier version was published at http://web.math.ku.dk/ solovej/mbnotes.

[BT92] Bernd Thaller, *The Dirac equation*, Texts and Monographs in Physics, Springer-Verlag, Berlin Heidelberg, 1992.

[Tes14] Gerald Teschl, *Mathematical Methods in Quantum Mechanics: With Applications to Schrödinger Operators, Second Edition*, Graduate Studies in Mathematics, vol. 157, American Mathematical Society, Providence, Rhode Island, 2014.

[Wüs75] Rainer Wüst, *Distinguished self-adjoint extensions of Dirac operators constructed by means of cut-off potentials*, Mathematische Zeitschrift **141** (1975), 93–98.

[Wüs77] ———, *Dirac operations with strongly singular potentials: Distinguished self-adjoint extensions constructed with a spectral gap theorem and cut-off potentials*, Mathematische Zeitschrift **3** (1977), 259–271.

2 | Article: Friedrichs extension and min–max principle for operators with a gap

FRIEDRICHS EXTENSION AND MIN–MAX PRINCIPLE FOR OPERATORS WITH A GAP

LUKAS SCHIMMER[1], JAN PHILIP SOLOVEJ, AND SABIHA TOKUS

Abstract. Semibounded symmetric operators have a distinguished self-adjoint extension, the Friedrichs extension. The eigenvalues of the Friedrichs extension are given by a variational principle that involves only the domain of the symmetric operator. Although Dirac operators describing relativistic particles are not semibounded, the Dirac operator with Coulomb potential is known to have a distinguished extension. Similarly, for Dirac-type operators on manifolds with a boundary a distinguished self-adjoint extension is characterised by the Atiyah–Patodi–Singer boundary condition. In this paper we relate these extensions to a generalisation of the Friedrichs extension to the setting of operators satisfying a gap condition. In addition we prove, in the general setting, that the eigenvalues of this extension are also given by a variational principle that involves only the domain of the symmetric operator. We also clarify what we believe to be inaccuracies in the existing literature.

1. Introduction and Main Result

For a symmetric, semibounded operator A with dense domain $D(A)$ on a Hilbert space $\mathcal{H}$ there exists a distinguished self-adjoint extension, the Friedrichs extension A_F. This extension was introduced by Friedrichs [17] in 1934. Its eigenvalues can be computed by a variational principle.

More precisely, if A is bounded from below by λ_1, where

$$\lambda_1 = \inf_{z\in D(A)} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|^2_{\mathcal{H}}} > -\infty\,, \tag{1}$$

a variational principle (see e.g. [7, Theorem 4.5.2]) states that the values

$$\lambda_k = \inf_{\substack{V\subset D(A)\\ \dim V = k}} \sup_{z\in V} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|^2_{\mathcal{H}}} \tag{2}$$

for $k \geq 1$ are the discrete spectrum of A_F in the interval $(-\infty, \sup_{k\geq 1}\lambda_k)$, counted with multiplicities

$$d_k \coloneqq \#\{j \geq 1 : \lambda_j = \lambda_k\}$$

as long as $d_k < \infty$. If $d_k = \infty$ then λ_k is in the essential spectrum of A_F. While similar variational principles hold for all semibounded self-adjoint

1991 *Mathematics Subject Classification.* 49R05, 49S05, 47B25, 81Q10.

Key words and phrases. Self-adjoint extension; Spectral gap; Variational principle; Schur complement; Dirac operator; Atiyah–Patodi–Singer boundary condition.

[1]Corresponding author

The authors were supported by ERC Advanced Grant no. 321029 and VILLUM FONDEN through the QMATH Centre of Excellence (grant no. 10059).

extensions of A, we stress that for the calculation of the λ_k's according to (2) only the domain $D(A)$ is needed, making the spectrum of the Friedrichs extension especially accessible to numerical methods. This is a consequence of $D(A)$ being a form core for A_F.

For symmetric operators A that are not semibounded, Friedrichs' construction is not applicable. Of particular interest is the case where the self-adjoint extension of A is expected to have a gap in its spectrum. In a similar way to the semibounded case, one would like to solve the following problems.

(P1) Define a distinguished self-adjoint extension A_F of A.
(P2) Provide a simple variational principle that allows to compute the eigenvalues of A_F, ideally only from the symmetric operator A.

In this paper, we will generalise the construction of the Friedrichs extension A_F to symmetric operators A where the lower semiboundedness (1) is replaced by a gap condition. We will furthermore relate the extension to a variational principle that only involves the domain of the symmetric operator A hence providing solutions to both problems, (P1) and (P2). An important example of an operator that our results apply to is the Dirac operator H_ν on $L^2(\mathbb{R}^3;\mathbb{C}^4)$ with Coulomb potential $-\nu/|x|$. The operator H_ν is not semibounded and for $\nu > \sqrt{3}/2$ it is not essentially self-adjoint on the space of smooth, compactly supported functions $\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$.

Our results also apply to Dirac-type operators on manifolds with a boundary. For these operators, there exists a distinguished self-adjoint extension which can be characterised by a non-local boundary condition, as first introduced by Atiyah, Patodi and Singer in the proof of their index theorem [1]. We will show that this boundary condition naturally arises from the construction given in this paper.

The problem (P1) has been studied already by Krein [23] for a symmetric operator A that is not semibounded but satisfies a gap condition of the form ($\lambda_0 < \lambda_1$)

$$\left\| \left(A - \frac{\lambda_0 + \lambda_1}{2} \right) z \right\|_{\mathcal{H}} \geq \frac{\lambda_1 - \lambda_0}{2} \|z\|_{\mathcal{H}} \,. \tag{3}$$

In Krein's work it is proved that such an operator has a self-adjoint extension that preserves the gap, i.e. the interval (λ_0, λ_1) belongs to the resolvent set of the extension. Subsequently Brasche and Neidhart [4] parametrised all gap-preserving self-adjoint extensions of A by using a suitable representation for their inverses. The authors' parametrisation allowed them to identify one of the extensions as the Friedrichs extension in the limit $\lambda_0 \to -\infty$.

The type of operators we wish to consider here satisfy a gap condition which is seen to imply Krein's gap condition (3), as will be proved in Remark 3. In analogy to the Friedrichs extension preserving the lower-semiboundedness, our extension A_F preserves (3).

More recently, different forms of gap conditions have been considered. Esteban and Loss [14] considered a block-matrix operator

$$\begin{pmatrix} P & Q \\ T & -S \end{pmatrix} \tag{4}$$

densely defined on a domain $\mathcal{D}_0 \times \mathcal{D}_0 \subset \mathcal{H}_0 \times \mathcal{H}_0$ where $P = P^*, S = S^*, Q = T^*$ and $S \geq -\lambda_0 > 0$. Furthermore they assumed that $P, Q, S, T, S^{-1}T$ and $QS^{-1}T$ map $\mathcal{D}_0$ into $\mathcal{H}_0$. Their gap condition was phrased in terms of the assumption that for some $\lambda_1 > 0$ and all $z \in \mathcal{D}_0$

$$q_{\lambda_1}(z,z) \coloneqq \langle (S+\lambda_1)^{-1}Tz, Tz\rangle_{\mathcal{H}} + \langle (P-\lambda_1)z, z\rangle_{\mathcal{H}} \geq 0\,.$$

In the case of Dirac operators H_ν with Coulomb potentials this assumption constitutes a Hardy inequality that was previously proved analytically by Dolbeault, Esteban, Loss and Vega [9]. In this way Loss and Esteban [13] were able to define a distinguished self-adjoint extension for H_ν up to and including the critical value $\nu = 1$. For $\nu < 1$ their extension coincides with the previously known distinguished extension established separately by Schmincke [30], Nenciu [27] and Wüst [34] (which were all proved to be equal by Klaus and Wüst [21]). Recently, Gallone and Michelangeli classified all self-adjoint extensions of the Dirac–Coulomb operator for $\nu < 1$ [18] and characterised their eigenvalues as roots of transcendental equations[19]. The case of a two-body Dirac operator with Coulomb interaction was recently addressed by Deckert and Oelker [8] who were able to establish a distinguished self-adjoint extension. This case differs from our problem setting in that the two-body operator does not exhibit a gap but the methods used are similar to those applied in [14].

Regarding the second problem (P2), variational principles have been studied by several authors for self-adjoint operators with gaps. For Dirac operators with negative potentials Talman [31] as well as Datta and Deviah [6] suggested a way to compute the first eigenvalue. The idea was to split the optimisation in the variational principle. Decomposing the Hilbert space into a direct sum

$$L^2(\mathbb{R}^3;\mathbb{C}^4) = (L^2(\mathbb{R}^3;\mathbb{C}^2) \times \{0\}) \oplus (\{0\} \times L^2(\mathbb{R}^3;\mathbb{C}^2))$$

corresponding to the upper and lower spinors, the first eigenvalue would be given by first maximising the quadratic form over one component and then minimising over the other. More precisely, for suitably chosen spaces

$$F_+ \subset L^2(\mathbb{R}^3;\mathbb{C}^2) \times \{0\}\,, \qquad F_- \subset \{0\} \times L^2(\mathbb{R}^3;\mathbb{C}^2)$$

the authors suggested that

$$\lambda_1 = \inf_{x_+ \in F_+ \setminus \{0\}} \sup_{y_- \in F_-} \frac{\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}}}{\|x_+ + y_-\|^2_{\mathcal{H}}}\,.$$

For Dirac operators H_ν with Coulomb potential $-\nu/|x|$ such a variational principle describing the discrete spectrum was proved by Dolbeault, Esteban and Séré [10] in the case of essentially self-adjointness $\nu \in [0, \sqrt{3}/2]$ where

they could choose $F_+ = \mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)\times\{0\}$ and $F_- = \{0\}\times\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)$. Their argument for $\nu \in (\sqrt{3}/2, 1)$ was not complete. For $\nu < 1$ Morozov and Müller [25, 26] showed that $F_+ = H^{1/2}(\mathbb{R}^3;\mathbb{C}^2)\times\{0\}$ and $F_- = \{0\}\times H^{1/2}(\mathbb{R}^3;\mathbb{C}^2)$ are valid choices to obtain a variational principle for the distinguished extension.

In the general setting of a self-adjoint operator with spectral gap, variational principles that use an orthogonal decomposition of the Hilbert space were investigated by Griesemer and Siedentop [20]. Abstract variational principles were also proved in [10, 25] and with different assumptions by Kraus, Langer and Tretter [22] (see also [33]). In all these results however, the operator is a-priori assumed to be self-adjoint or essentially self-adjoint.

Only recently Esteban, Lewin and Séré [12] extended the variational principle for Dirac operators with Coulomb potentials to all $\nu \in [0, 1]$ and discussed its connections to the distinguished self-adjoint extension. Building upon the results of [10] they showed that for any $\nu \in [0, 1]$ it is sufficient to choose $F_+ = \mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)\times\{0\}$ and $F_- = \{0\}\times\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)$ to obtain the eigenvalues of the distinguished extension, evoking similarities to the Friedrichs extension in this special case.

Our main result, Theorem 1 clarifies the connection between a distinguished self-adjoint extension and a variational principle in the case of operators satisfying a general gap condition. It applies, in particular, to the Dirac–Coulomb operator, thus generalising the result of [12].

Theorem 1. *Let A be a densely defined symmetric operator on a Hilbert space $\mathcal{H}$ and let $\langle x, Ay\rangle_{\mathcal{H}}$ be the corresponding real quadratic form with form domain equal to the operator domain $D(A)$. Furthermore the following assumptions are made.*

(i) **Orthogonal decomposition:** *There are orthogonal projections $\Lambda_\pm$ on $\mathcal{H}$ with $\Lambda_+ + \Lambda_- = \mathbb{I}_{\mathcal{H}}$ such that*

$$\mathcal{H} = \Lambda_+\mathcal{H} \oplus \Lambda_-\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$$

and

$$F_\pm \coloneqq \Lambda_\pm D(A) \subset D(A)\,.$$

(ii) **Gap condition:**

$$\sup_{y_-\in F_-\setminus\{0\}} \frac{\langle y_-, Ay_-\rangle_{\mathcal{H}}}{\|y_-\|_{\mathcal{H}}^2} =: \lambda_0 < \lambda_1 \coloneqq \inf_{x_+\in F_+\setminus\{0\}} \sup_{y_-\in F_-} \frac{\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}}}{\|x_+ + y_-\|_{\mathcal{H}}^2}\,.$$

(iii) The operator $\Lambda_- A|_{F_-} : F_- \to \mathcal{H}_-$ is essentially self-adjoint.

Then there exists a self-adjoint extension A_F of A such that for $k \geq 1$ the numbers

$$\lambda_k \coloneqq \inf_{\substack{V\subset F_+\\ \dim V = k}} \sup_{z\in(V\oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2} \tag{5}$$

are the eigenvalues of A_F in the set $(\lambda_0, \sup_{\ell\geq 1}\lambda_\ell)$ counted with multiplicities

$$d_k := \#\{j \geq 1 : \lambda_j = \lambda_k\}$$

as long as $d_k < \infty$. If $d_k = \infty$ then λ_k is in the essential spectrum of A_F. The operator A_F is the unique self-adjoint extension with the property that $D(A_F) \subset \mathcal{F}_+ \oplus \mathcal{H}_-$, for a subspace $\mathcal{F}_+ \subset \mathcal{H}_+$ defined in the proof.

Remark 2. Theorem 1 is very similar to [10, Theorem 1.1]. In [10], however, the variational principle is given for an operator which is already self-adjoint whereas the important aspect of our theorem is that we give the variational principle in terms of a symmetric operator. We therefore do not need to determine the domain of the self-adjoint extension in order to apply the variational principle. This is analogous to the case of the Friedrichs extension where the variational principle may be given in terms of the semibounded symmetric operator. Thus, our Theorem 1 would be a generalisation of the theorem in [10] except that our condition (*iii*) is not included in there. We do, however, believe this assumption is needed, also in the case in [10].

Remark 3. Assumptions (*i*) and (*ii*) of Theorem 1 imply that A satisfies Krein's gap condition (3). To see this, we let $x = x_+ + x_- \in D(A)$ and for given $\varepsilon > 0$ choose $y_-^\varepsilon \in D(A)$ such that

$$\langle x_+ + y_-^\varepsilon, A(x_+ + y_-^\varepsilon)\rangle_{\mathcal{H}} \geq (\lambda_1 - \varepsilon)\|x_+ + y_-^\varepsilon\|_{\mathcal{H}}^2 . \tag{6}$$

Then with $\lambda := (\lambda_0 + \lambda_1)/2$

$$\begin{aligned}\|(A-\lambda)\,x\|_{\mathcal{H}} &\geq \sup_{z\in D(A)} \frac{|\Re\langle (A-\lambda)x, z\rangle_{\mathcal{H}}|}{\|z\|_{\mathcal{H}}} \\ &= \sup_{z\in D(A)} \frac{|\langle x+z, (A-\lambda)(x+z)\rangle_{\mathcal{H}} - \langle x-z, (A-\lambda)(x-z)\rangle_{\mathcal{H}}|}{4\|z\|_{\mathcal{H}}}.\end{aligned}$$

Choosing $z := x_+ - x_- + 2y_-^\varepsilon \in D(A)$ and using (6) together with the definition of λ_0 we obtain the lower bound

$$\begin{aligned}&\|(A-\lambda)\,x\|_{\mathcal{H}} \\ &\geq \frac{\left|\langle x_+ + y_-^\varepsilon, (A-\lambda)(x_+ + y_-^\varepsilon)\rangle_{\mathcal{H}} - \langle x_- - y_-^\varepsilon, (A-\lambda)(x_- - y_-^\varepsilon)\rangle_{\mathcal{H}}\right|}{\|z\|_{\mathcal{H}}} \\ &\geq \left(\frac{\lambda_1-\lambda_0}{2} - \varepsilon\right)\frac{\|x+z\|_{\mathcal{H}}^2 + \|x-z\|_{\mathcal{H}}^2}{4\|z\|_{\mathcal{H}}} .\end{aligned}$$

Using the parallelogram law and the fact that $a + 1/a \geq 2$ for any $a > 0$ we obtain

$$\begin{aligned}\|(A-\lambda)\,x\|_{\mathcal{H}} &\geq \left(\frac{\lambda_1-\lambda_0}{2} - \varepsilon\right)\frac{\|x\|_{\mathcal{H}}}{2}\left(\frac{\|x\|_{\mathcal{H}}}{\|z\|_{\mathcal{H}}} + \frac{\|z\|_{\mathcal{H}}}{\|x\|_{\mathcal{H}}}\right) \\ &\geq \left(\frac{\lambda_1-\lambda_0}{2} - \varepsilon\right)\|x\|_{\mathcal{H}} .\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, Krein's gap condition (3) holds.

By an application of the spectral theorem the same holds true for the extension A_F.

Conversely, if a symmetric operator A satisfies Krein's gap condition we do not know whether there is a universally applicable strategy to find orthogonal projections $\Lambda_\pm$ such that the conditions of Theorem 1 are satisfied. According to Krein there exist self-adjoint extensions that preserve the gap, as mentioned above. In an attempt to apply our theorem to this problem a natural choice (and in the very general setting also the only possible candidates we could propose) for $\Lambda_\pm$ are the spectral projections onto $(-\infty, \lambda_0]$ and $[\lambda_1, \infty)$, respectively, for one of these self-adjoint extensions. However, for such a choice of projections the assumption (i) of Theorem 1 may not be satisfied as $\Lambda_\pm D(A)$ is not necessarily contained in $D(A)$. As an example one can consider the Schrödinger–Coulomb operator, i.e. $A = -\Delta - \frac{\nu}{|x|}$, on $\mathcal{C}_0^\infty(\mathbb{R}^3) \subset L^2(\mathbb{R}^3)$ with $\nu > 0$. By Hardy's inequality together with the Kato–Rellich theorem A is essentially self-adjoint for all $\nu > 0$. The spectrum of the unique self-adjoint extension $\overline{A}$ consists of discrete eigenvalues $(\lambda_k)_{k\geq 0}$ converging to the bottom of the essential spectrum $[0, \infty)$. The lowest eigenvalue $\lambda_0 = -\nu^2/4$ is non-degenerate with eigenfunction $\mathrm{e}^{-\nu|x|/2}$. Since the symmetric operator A is contained in $\overline{A}$ it satisfies Krein's gap condition with parameters λ_0 and λ_1. However, the spectral projection Λ_- of $\overline{A}$ onto $(-\infty, \lambda_0]$ is simply the projection onto $\mathrm{e}^{-\nu|x|/2}$, which does not leave $\mathcal{C}_0^\infty(\mathbb{R}^3)$ invariant. Nevertheless, it is worth mentioning that in this specific case of a lower semibounded operator we can still find suitable projections that would allow us to apply our result, since the choice $\Lambda_- = 0$ satisfies all the assumptions of Theorem 1.

We construct A_F as an analogue to the Friedrichs extension of a semibounded operator (see e.g. [28, Theorem VIII.15] and [29, Theorem X.23] as well as [3, pp. 224]). We closely follow [10], the main idea being the following. If A is a bounded self-adjoint operator such that $F_\pm = \mathcal{H}_\pm$, then for any $E \notin \sigma(\Lambda_- A|_{\mathcal{H}_-})$ the decomposition

$$\begin{pmatrix} \Lambda_+ A|_{\mathcal{H}_+} & \Lambda_+ A|_{\mathcal{H}_-} \\ \Lambda_- A|_{\mathcal{H}_+} & \Lambda_- A|_{\mathcal{H}_-} \end{pmatrix} - E\mathbb{I} = \begin{pmatrix} \mathbb{I} & -L_E^* \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} Q_E & 0 \\ 0 & -(B+E) \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ -L_E & \mathbb{I} \end{pmatrix} \tag{7}$$

holds (see e.g. [33, Proposition 1.6.2]), where

$$\begin{aligned} B &= -\Lambda_- A|_{\mathcal{H}_-}\,, \\ L_E &= (B+E)^{-1}\Lambda_- A|_{\mathcal{H}_+}\,, \\ Q_E &= (\Lambda_+ A - E)|_{\mathcal{H}_+} + \Lambda_+ A|_{\mathcal{H}_-}(B+E)^{-1}\Lambda_- A|_{\mathcal{H}_+}\,. \end{aligned}$$

The operator Q_E is one of two Schur complements of A.

In Section 2, we will construct A_F by defining these three operators. The definition of B in Subsection 2.1 is straightforward and yields an operator with form domain denoted by $\mathcal{F}_- \subset \mathcal{H}_-$. Complications arise from the fact that the Schur complement is only defined in terms of a quadratic

form q_E which is not necessarily closable on $\mathcal{H}_+$. Thus a new Hilbert space $\mathcal{G}_+$, which is obtained when considering the closure L_E of the operator $(B+E)^{-1}\Lambda_- A|_{F_+}$, has to be introduced in Subsection 2.2. That $(B+E)^{-1}\Lambda_- A|_{F_+}$ is closable is non-trivial and does not seem to hold true without assumption (*iii*). For this reason we believe (*iii*) is necessary to guarantee that $\mathcal{G}_+$ can be identified with a subspace of $\mathcal{H}_+$. On $\mathcal{G}_+$, we can close q_E and define the corresponding operator Q_E, as done in Subsection 2.3. Particular consideration has to be given to the fact that the construction does not depend on the explicit choice of $E > \lambda_0$. In Subsection 2.4 the definition of the self-adjoint extension A_F is given in a form that resembles the above decomposition (7). In Subsection 2.6 the variational principle stated in Theorem 1 will be proved.

Table 1 summarises the Hilbert spaces that need to be defined while Table 2 lists all the additional spaces.

In Section 3 we will apply Theorem 1 to the Dirac–Coulomb operator.

In Section 4 we will introduce the APS-boundary condition for generalised Dirac-operators and prove that the self-adjoint extension constructed according to Theorem 1 is exactly characterised by these boundary conditions.

Remark 4. Our construction of the distinguished self-adjoint extension differs from [14]. Phrasing our assumptions in terms of the block-matrix notation (4), we do not require P and S to be self-adjoint nor that $Q = T^*$. In addition, we do not make any assumption about the domain of $QS^{-1}T$. With the setup as in [14] the quadratic form q_E is closable on $\mathcal{H}_+$. In a general setting this is not likely to be the case. Moreover in [14] it is important that the closure of q_E is independent of E. This independence is claimed but not proved. In our construction the introduction of $\mathcal{G}_+$ is necessary to guarantee both that q_E is closable on $\mathcal{G}_+$ and that the domain of the closure and hence the self-adjoint extension do not depend on the choice of E. Nevertheless our construction is inspired by the approach in [13] and [14].

Space	(Equivalent) norms	Inclusion	Description	Page
$\mathcal{H}$	$\|\cdot\|_{\mathcal{H}}$			p.4
$\mathcal{H}_+$	$\|\cdot\|_{\mathcal{H}}$	$\mathcal{H}_+ \subset \mathcal{H}$	$\Lambda_+\mathcal{H}$	p.4
$\mathcal{H}_-$	$\|\cdot\|_{\mathcal{H}}$	$\mathcal{H}_- \subset \mathcal{H}$	$\Lambda_-\mathcal{H}$	p.4
$\mathcal{F}_-$	$\|\cdot\|_{\mathcal{F}_-}$	$\mathcal{F}_- \subset \mathcal{H}_-$	Form domain of B	p.8
$\mathcal{G}_+$	$\|\cdot\|_E, E > \lambda_0$	$\mathcal{G}_+ \subset \mathcal{H}_+$	Domain of L_E	p.9
$\mathcal{F}_+$	$\|\cdot\|_{\mathcal{F}_+,E}, E > \lambda_0$	$\mathcal{F}_+ \subset \mathcal{G}_+$	Form domain of Q_E	p.12

TABLE 1. The required Hilbert spaces

Space	Contained in	Description	Page
$D(A_F)$	$D(A_F) \subset \mathcal{F}_+ \oplus \mathcal{H}_-$	Domain of A_F	p.16
$D(A)$	$D(A) \subset D(A_F)$	Domain of A	p.4
F_+	$F_+ \subset \mathcal{F}_+$	$\Lambda_+ D(A)$	p.4
F_-	$F_- \subset \mathcal{F}_-$	$\Lambda_- D(A)$	p.4

TABLE 2. The additionally required vector spaces

2. The Proof of Theorem 1

2.1. **The Definition of B.** We start by setting

$$\langle y_-, z_- \rangle_{F_-} := (\lambda_0 + 1)\langle y_-, z_- \rangle_{\mathcal{H}} - \langle y_-, Az_- \rangle_{\mathcal{H}}$$

which by definition of λ_0 is an inner product on F_- with corresponding norm

$$\|y_-\|^2_{F_-} = (\lambda_0 + 1)\|y_-\|^2_{\mathcal{H}} - \langle y_-, Ay_- \rangle_{\mathcal{H}} .$$

Since the quadratic form $\langle \cdot, \cdot \rangle_{F_-}$ comes from a symmetric operator it is closable, i.e. it extends to a closed quadratic form on the form domain $\mathcal{F}_- \subset \mathcal{H}_-$, which is the closure of F_- with respect to the norm $\|\cdot\|_{F_-}$. If we denote the continuous extension of the quadratic form $\langle \cdot, \cdot \rangle_{F_-}$ to $\mathcal{F}_-$ by $\langle \cdot, \cdot \rangle_{\mathcal{F}_-}$, then $(\mathcal{F}_-, \langle \cdot, \cdot \rangle_{\mathcal{F}_-})$ forms a Hilbert space.

By assumption $\Lambda_- A|_{F_-}$ is essentially self-adjoint, hence there exists a unique self-adjoint extension given by its closure, which we will denote by $-B$. It is then clear that $B + \lambda_0 + 1$ coincides with the self-adjoint operator associated with the closed quadratic form $\langle \cdot, \cdot \rangle_{\mathcal{F}_-}$ such that

$$\langle y_-, z_- \rangle_{\mathcal{F}_-} = \langle y_-, Bz_- \rangle_{\mathcal{H}} + (\lambda_0 + 1)\langle y_-, z_- \rangle_{\mathcal{H}}.$$

The form domain of B is $\mathcal{F}_-$ and its operator domain $D(B)$ is a subset of $\mathcal{F}_-$. For $E > \lambda_0$ the self-adjoint operator $B + E$ is strictly positive and its inverse $(B + E)^{-1}$ is well-defined and bounded on all of $\mathcal{H}_-$.

Remark 5. Since $\Lambda_- A|_{F_-}$ is essentially self-adjoint, the operator B coincides with the Friedrichs extension of the semibounded operator $-\Lambda_- A|_{F_-}$. For the convenience of the reader and to evoke connections to our construction, we recall the definition of this extension. Using Riesz' theorem we first define the operator $\widehat{P}$ as the isometric isomorphism between the Hilbert space $\mathcal{F}_-$ and its dual $\mathcal{F}'_-$, i.e. for any $z_- \in \mathcal{F}_-$ we define $\widehat{P} z_- \in \mathcal{F}'_-$ to be the unique continuous functional such that

$$[\widehat{P} z_-](y_-) = \langle z_-, y_- \rangle_{\mathcal{F}_-} .$$

With the embedding $j_- : \mathcal{H}_- \to \mathcal{F}'_-$ given by $[j_-(y_-)](z_-) = \langle y_-, z_- \rangle_{\mathcal{H}}$ (identifying $\mathcal{H}_-$ with its dual space $\mathcal{H}'_-$) we can show that on the domain

$$D(P) = \left\{ z_- \in \mathcal{F}_- \subset \mathcal{H}_- : \widehat{P} z_- \in j_-(\mathcal{H}_-) \right\} \subset \mathcal{H}_-$$

the operator $P = j_-^{-1} \circ \widehat{P}$ is a self-adjoint extension of $-\Lambda_- A|_{F_-} + \lambda_0 + 1$. The Friedrichs extension B is then defined as $B = P - \lambda_0 - 1$ with domain $D(B) = D(P)$. In particular, the quadratic form $\langle y_-, Bz_- \rangle_{\mathcal{H}}$ has a

continuous extension to all $x_-, y_- \in \mathcal{F}_-$ given by $[\widehat{B}(y_-)](z_-)$ where $\widehat{B} = \widehat{P} - (\lambda_0 + 1)j_-$. The form domain of B is consequently $\mathcal{F}_-$.

2.2. **The Definition of L_E.** Let $E > \lambda_0$. Then for $x_+ \in F_+$ the mapping $x_+ \mapsto (B+E)^{-1}\Lambda_- A x_+$ defines a linear operator from F_+ into $\mathcal{H}_-$. The proof of the second part of the following lemma is adapted from [10, Lemma 2.1].

Lemma 6. *The operator $(B+E)^{-1}\Lambda_- A$ defined on F_+ is closable. We denote its closure by L_E with graph norm*

$$\|x_+\|_E^2 = \|x_+ + L_E x_+\|_{\mathcal{H}}^2 = \|x_+\|_{\mathcal{H}}^2 + \|L_E x_+\|_{\mathcal{H}}^2 .$$

For $\lambda_0 < E \le E'$ the norms $\|\cdot\|_E$ and $\|\cdot\|_{E'}$ are equivalent on F_+ with

$$\|x_+\|_{\mathcal{H}} \le \|x_+\|_{E'} \le \|x_+\|_E \le C_{E,E'}\|x_+\|_{E'} , \tag{8}$$

where $C_{E,E'} = (E'-\lambda_0)/(E-\lambda_0) \ge 1$.

Proof. We first show that the operator is closable. Consider a sequence of $x_n \in F_+$ with $\|x_n\|_{\mathcal{H}} \to 0$ and $y \in \mathcal{H}_-$ with $\|(B+E)^{-1}\Lambda_- A x_n - y\|_{\mathcal{H}} \to 0$. We have to show that $y = 0$. Let $z \in (B+E)F_- \subset \mathcal{H}_-$. Then

$$\begin{aligned}|\langle z, (B+E)^{-1}\Lambda_- A x_n\rangle_{\mathcal{H}}| &= |\langle (B+E)^{-1} z, \Lambda_- A x_n\rangle_{\mathcal{H}}| \\ &= |\langle A(B+E)^{-1} z, x_n\rangle_{\mathcal{H}}| \le \|A(B+E)^{-1} z\|_{\mathcal{H}}\|x_n\|_{\mathcal{H}}\end{aligned}$$

and the right-hand side converges to zero. Since $\Lambda_- A|_{F_-}$ is essentially self-adjoint, we can conclude that $(B+E)F_-$ is dense in $\mathcal{H}_-$ and thus $y = 0$.

Next, assume $\lambda_0 < E \le E'$. The first inequality in (8) follows directly from the definition of the norm $\|\cdot\|_E$. For a bound on $\|\cdot\|_E$ in terms of $\|\cdot\|_{E'}$ we note that by the spectral theorem for $x \in D(B)$

$$\left\|(B+E)^{-1}(B+E')x\right\|_{\mathcal{H}}^2 \le \sup_{\lambda \ge -\lambda_0} \frac{|\lambda + E'|^2}{|\lambda + E|^2}\|x\|_{\mathcal{H}}^2 \le \frac{(E'-\lambda_0)^2}{(E-\lambda_0)^2}\|x\|_{\mathcal{H}}^2 .$$

As a consequence we obtain with $C_{E,E'} \coloneqq (E'-\lambda_0)/(E-\lambda_0)$ for any $x_+ \in F_+$

$$\begin{aligned}\|L_E x_+\|_{\mathcal{H}} = \left\|(B+E)^{-1}\Lambda_- A x_+\right\|_{\mathcal{H}} &\le C_{E,E'}\left\|(B+E')^{-1}\Lambda_- A x_+\right\|_{\mathcal{H}} \\ &= C_{E,E'}\|L_{E'} x_+\|_{\mathcal{H}} ,\end{aligned}$$

which proves the last inequality in (8). A bound on $\|\cdot\|_{E'}$ in terms of $\|\cdot\|_E$ follows in the same way. □

We conclude that the domain of L_E, meaning the closure of F_+ with respect to the norm $\|\cdot\|_E$, can be identified for all values of $E > \lambda_0$ and we will denote this vector space by $\mathcal{G}_+$. Together with the inner product

$$\langle x_+, z_+\rangle_E \coloneqq \langle x_+, z_+\rangle_{\mathcal{H}} + \langle L_E x_+, L_E z_+\rangle_{\mathcal{H}}$$

it forms a Hilbert space $(\mathcal{G}_+, \langle\cdot,\cdot\rangle_E)$ and we have the vector space inclusions

$$F_+ \subset \mathcal{G}_+ \subset \mathcal{H}_+ ,$$

where the last equation also holds in the sense of Hilbert spaces.

As an operator from $(\mathcal{G}_+, \|\cdot\|_E)$ to $(\mathcal{H}_-, \|\cdot\|_{\mathcal{H}})$, L_E is then bounded. We will later consider L_E as an operator on an even smaller Hilbert space, where it is consequently also bounded.

2.3. **The Definition of Q_E.** For $E > \lambda_0$ we now define the quadratic form q_E on $F_+ \times F_+$

$$q_E(x_+, z_+) \coloneqq \langle x_+, (A-E)z_+\rangle_{\mathcal{H}} + \langle \Lambda_- A x_+, (B+E)^{-1}\Lambda_- A z_+\rangle_{\mathcal{H}}\,.$$

It is the quadratic form related to one of the Schur complements of the matrix representation of A. We will see that q_E can be closed as a lower-semibounded form on $(\mathcal{G}_+, \|\cdot\|_E)$ such that the closure is independent of E. To this end we first derive the following result which can also be found in [10, pp. 210].

Lemma 7. *For $E > \lambda_0$ and $x_+ \in F_+$ let $\varphi_{E,x_+} : F_- \to \mathbb{R}$ be the function defined as*

$$\varphi_{E,x_+}(y_-) \coloneqq \langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}} - E\|x_+ + y_-\|^2_{\mathcal{H}}\,.$$

The quadratic form q_E is then related to φ_{E,x_+} by

$$q_E(x_+, x_+) = \sup_{y_- \in F_-} \varphi_{E,x_+}(y_-)\,.$$

In particular, $\varphi_{E,x_+}(\cdot)$ can be extended to $\mathcal{F}_-$ and the extension attains its maximum at the unique point $y_{\max} = L_E x_+ = (B+E)^{-1}\Lambda_- A x_+$.

Proof. For $y_- \in F_-$ we write

$$\varphi_{E,x_+}(y_-) = \langle x_+, (A-E)x_+\rangle_{\mathcal{H}} + 2\Re\langle y_-, Ax_+\rangle_{\mathcal{H}} - \langle y_-, (B+E)y_-\rangle_{\mathcal{H}}. \quad (9)$$

It is then clear that the functional $\varphi_{E,x_+}(\cdot)$ naturally extends to $\mathcal{F}_-$, see also Remark 5. We denote the continuous extension by $\overline{\varphi}_{E,x_+}$.

The quadratic polynomial $f : \mathbb{R} \to \mathbb{R}$ which we can define for any $y_-, z_- \in \mathcal{F}_-$ as

$$f(h) = \overline{\varphi}_{E,x_+}(y_- + h(z_- - y_-))\,, h \in \mathbb{R}$$

is strictly concave and thus we have

$$f(1) < f(0) + f'(0)\,. \quad (10)$$

Now assume that $y_- \in \mathcal{F}_-$ satisfies the Euler equation, that is

$$\overline{\varphi}'_{E,x_+}(y_-; (z_- - y_-)) = 0$$

for all $z_- \in \mathcal{F}_-$. Then we must have $\langle w_-, \Lambda_- A x_+ - (B+E)y_-\rangle_{\mathcal{H}} = 0$ for all $w_- \in \mathcal{F}_-$ or equivalently

$$y_- = (B+E)^{-1}\Lambda_- A x_+ \quad (11)$$

and by (10) for all $z_- \in \mathcal{F}_-$, $z_- \neq y_-$

$$\overline{\varphi}_{E,x_+}(z_-) < \overline{\varphi}_{E,x_+}(y_-)\,.$$

Consequently $\overline{\varphi}_{E,x_+}(\cdot)$ has a unique global maximum at the point $y_{\max} = (B+E)^{-1}\Lambda_- A x_+ \in D(B)$. Inserting (11) into (9) we obtain

$$\overline{\varphi}_{E,x_+}(y_{\max}) = \langle x_+, (A-E)x_+\rangle_{\mathcal{H}} + \langle \Lambda_- A x_+, L_E x_+\rangle_{\mathcal{H}}\,.$$

□

The following lemma establishes important properties of the quadratic form q_E. The proof can also be found in [10, Lemma 2.1].

Lemma 8. *Let $\lambda_0 < E \leq E'$.*

(i) On F_+ the quadratic forms q_E and $q_{E'}$ satisfy

$$q_{E'}(x_+,x_+)+(E'-E)\|x_+\|^2_{E'} \leq q_E(x_+,x_+) \leq q_{E'}(x_+,x_+)+(E'-E)\|x_+\|^2_E\,.$$

(ii) There is a constant $\kappa_E \geq 1$ such that for any $x_+ \in F_+$

$$q_E(x_+,x_+) + \kappa_E \langle x_+,x_+\rangle_E \geq \langle x_+,x_+\rangle_E\,.$$

Thus, the quadratic form q_E is bounded from below on $F_+ \subset \mathcal{G}_+$.

(iii) It holds that

$$\begin{aligned} &q_E(x_+,x_+) > 0 \text{ for all } x_+ \in F_+\setminus\{0\} \qquad \text{if and only if} \qquad E < \lambda_1\,,\\ &q_E(x_+,x_+) \geq 0 \text{ for all } x_+ \in F_+\setminus\{0\} \qquad \text{if and only if} \qquad E \leq \lambda_1\,. \end{aligned}$$

Proof. For (i) we use the resolvent identity to compute for any $\lambda, \lambda' > \lambda_0$,

$$\begin{aligned} &q_{\lambda'}(x_+,x_+)\\ &= q_\lambda(x_+,x_+)+(\lambda-\lambda')\Big[\|x_+\|^2_{\mathcal{H}}+\langle \Lambda_- A x_+, (B+\lambda)^{-1}(B+\lambda')^{-1}\Lambda_- A x_+\rangle_{\mathcal{H}}\Big]\,. \end{aligned}$$

The result then follows by setting $\lambda = E$, $\lambda' = E'$ and $\lambda = E'$, $\lambda' = E$, respectively, and using $(B+E')^{-1} \leq (B+E)^{-1}$ to bound the last term.

We continue to show (iii) and subsequently (ii). We start by proving that the right-hand sides imply the left-hand sides. First, let $\lambda_0 < E < \lambda_1$ and $x_+ \in F_+\setminus\{0\}$ arbitrary. By the definition of λ_1, for any $\varepsilon > 0$ there exists a $y_-^\varepsilon \in F_-$ such that

$$\frac{\langle x_+ + y_-^\varepsilon, A(x_+ + y_-^\varepsilon)\rangle_{\mathcal{H}}}{\|x_+ + y_-^\varepsilon\|^2_{\mathcal{H}}} \geq \lambda_1 - \varepsilon$$

and consequently

$$\begin{aligned} q_E(x_+,x_+) \geq \varphi_{E,x_+}(y_-^\varepsilon) &= \langle x_+ + y_-^\varepsilon, A(x_+ + y_-^\varepsilon)\rangle_{\mathcal{H}} - E\|x_+ + y_-^\varepsilon\|^2_{\mathcal{H}}\\ &\geq (\lambda_1 - \varepsilon - E)\|x_+ + y_-^\varepsilon\|^2_{\mathcal{H}}\,. \end{aligned}$$

We can conclude that for all $x_+ \in F_+\setminus\{0\}$ and all $\lambda_0 < E < \lambda_1$

$$q_E(x_+,x_+) \geq (\lambda_1 - E)\|x_+\|^2_{\mathcal{H}} > 0\,.$$

We next prove that $q_{\lambda_1}(x_+,x_+) \geq 0$. Setting $E' \coloneqq \lambda_1$ and using (i) we have that

$$q_{\lambda_1}(x_+,x_+) \geq q_E(x_+,x_+) - (\lambda_1 - E)\|x_+\|^2_E\,,$$

which in the case $E \to \lambda_1$ shows $q_{\lambda_1} \geq 0$ since $\|x_+\|_E \to \|x_+\|_{\lambda_1}$ by (8).

We now prove the reverse implications. If $E > \lambda_1$ then again by definition of λ_1, for any $\varepsilon > 0$ with $\lambda_1 < E - \varepsilon$ there exists $x_+^\varepsilon \in F_+ \setminus \{0\}$ with

$$\frac{\langle (x_+^\varepsilon + y_-), A(x_+^\varepsilon + y_-)\rangle_{\mathcal{H}}}{\|x_+^\varepsilon + y_-\|_{\mathcal{H}}^2} \leq (E - \varepsilon)$$

for all $y_- \in F_-$ and consequently

$$q_E(x_+^\varepsilon, x_+^\varepsilon) = \sup_{y_- \in F_-} \varphi_{E,x_+^\varepsilon}(y_-) \leq - \inf_{y_- \in F_-} \varepsilon \|x_+^\varepsilon + y_-\|_{\mathcal{H}}^2 \leq -\varepsilon \|x_+^\varepsilon\|_{\mathcal{H}}^2$$

which finishes the proof of (*iii*).

The statement in (*ii*) is now clear for $E \leq \lambda_1$. If $E > \lambda_1$ we use (8) to compute that

$$\begin{aligned} q_E(x_+, x_+) &\geq q_{\lambda_1}(x_+, x_+) - (E - \lambda_1)\|x_+\|_{\lambda_1}^2 \\ &\geq q_{\lambda_1}(x_+, x_+) - (E - \lambda_1) C_{\lambda_1, E} \|x_+\|_E^2 \,. \end{aligned}$$

Choosing $\kappa_E = 1 + \max(0, (E - \lambda_1) C_{\lambda_1, E})$ then gives the result.

□

Remark 9. In the first part of the proof we needed bounds on terms of the type $\|L_E x_+\|_{\mathcal{H}}$. This was done by using the new norm $\|\cdot\|_E$. Without specifying further assumptions on the operator A it is not possible to estimate the difference between quadratic forms q_E for different values of E by a $\|\cdot\|_{\mathcal{H}}$-norm only. Introducing the Hilbert space $\mathcal{G}_+$ thus turns out to be essential.

An immediate consequence of Lemma 8 together with Lemma 6 is that the completion $\mathcal{F}_+$ of F_+ with respect to the norm $\|\cdot\|_{F_+,E}$ induced by the inner product

$$\langle x_+, z_+ \rangle_{F_+,E} = q_E(x_+, z_+) + \kappa_E \langle x_+, z_+ \rangle_E$$

is independent of E. In the remainder we fix $E > \lambda_0$ and denote the extension of the inner product $\langle \cdot, \cdot \rangle_{F_+,E}$ to $\mathcal{F}_+$ by $\langle \cdot, \cdot \rangle_{\mathcal{F}_+,E}$. A priori, it is not clear that $\mathcal{F}_+$ is a subspace of $\mathcal{H}_+$. However, the following holds.

Lemma 10. *The semibounded quadratic form q_E is closable on the Hilbert space $\mathcal{G}_+$ for $E > \lambda_0$. The closure $\overline{q_E}$ has the form domain $\mathcal{F}_+$, independent of E, and can be identified with a subspace of $\mathcal{G}_+$ and subsequently also of $\mathcal{H}_+$.*

Proof. We show that the positive form $\langle \cdot, \cdot \rangle_{F_+,E} = q_E(\cdot, \cdot) + \kappa_E \langle \cdot, \cdot \rangle_E$ is closable. The argument used here can also be found in e.g. Lemma 6, Chapter 10.1 of [3]. Consider a sequence $x_n \in F_+$ which is a Cauchy sequence with respect to $\|\cdot\|_{\mathcal{F}_+,E}$ and which satisfies $\|x_n\|_E \to 0$. Then for any $z \in F_+$

$$\begin{aligned} |\langle z, x_n \rangle_{\mathcal{F}_+,E}| &\leq \kappa_E |\langle z, x_n \rangle_E| + |\langle z, (A - E) x_n \rangle_{\mathcal{H}}| + |\langle \Lambda_- A z, L_E x_n \rangle_{\mathcal{H}}| \\ &= \kappa_E |\langle z, x_n \rangle_E| + |\langle (A - E) z, x_n \rangle_{\mathcal{H}}| + |\langle \Lambda_- A z, L_E x_n \rangle_{\mathcal{H}}| \\ &\leq \kappa_E \left(\|z\|_E + \|(A - E) z\|_{\mathcal{H}} + \|\Lambda_- A z\|_{\mathcal{H}} \right) \|x_n\|_E \end{aligned}$$

and thus $\langle z, x_n\rangle_{\mathcal{F}_+,E} \to 0$, where we again crucially need the assumption that $x_n \to 0$ in the $\|\cdot\|_E$-norm, which is stronger than the $\|\cdot\|_{\mathcal{H}}$-norm. Since the sequence x_n is Cauchy it converges to some element in $\mathcal{F}_+$. Using the fact that F_+ is dense in $\mathcal{F}_+$ with respect to $\|\cdot\|_{\mathcal{F}_+,E}$, we conclude from the above calculation that this element must be zero. □

We thus have the Hilbert space inclusions

$$\mathcal{F}_+ \subset \mathcal{G}_+ \subset \mathcal{H}_+\,,$$

with their respective inner products implicit, and the corresponding inclusions of the associated dual spaces

$$\mathcal{H}'_+ \subset \mathcal{G}'_+ \subset \mathcal{F}'_+\,.$$

By Riesz' theorem there exists an isometric isomorphism $i_{X\to X'}(x) = \langle x, \cdot\rangle_X$ between each Hilbert space X and its dual space X'. In general we will not explicitly write the isomorphisms $i_{\mathcal{H}_\pm\to\mathcal{H}'_\pm}$ thus identifying $\mathcal{H}$ and its dual space $\mathcal{H}'$.

Furthermore for each of the Hilbert space inclusions $X \subset Y$ there is a corresponding embedding of dual spaces $j_{Y'\to X'} : Y' \to X'$, in the sense

$$[j_{Y'\to X'}\ell](x) = \ell(x) = \langle i^{-1}_{Y\to Y'}\ell, x\rangle_Y\,, \qquad \ell \in Y',\ x \in X\,.$$

All these embeddings are bounded in norm by one.

Associated to the closed quadratic form $\overline{q_E}$ there is an operator defined on all of the form domain $\mathcal{F}_+$, as well as a self-adjoint operator with domain a subset of $\mathcal{F}_+$.

Lemma 11. *Let $E > \lambda_0$. There exists an operator $\widehat{Q_E} : \mathcal{F}_+ \subset \mathcal{G}_+ \to \mathcal{F}'_+$ with the following properties.*

(i) For all $x_+, z_+ \in \mathcal{F}_+$ the closure $\overline{q_E}$ of q_E on $\mathcal{F}_+$ is given by

$$\overline{q_E}(x_+, z_+) = [\widehat{Q_E}x_+](z_+).$$

(ii) The operator $\widehat{Q_E}$ is bounded and if additionally $E < \lambda_1$ then its inverse $\widehat{Q_E}^{-1}$ is also bounded.

(iii) On the dense domain

$$D(Q_E) = \left\{z_+ \in \mathcal{F}_+ : \widehat{Q_E}z_+ \in j_{\mathcal{G}'_+\to\mathcal{F}'_+}(\mathcal{G}'_+)\right\} \subset \mathcal{G}_+$$

the operator

$$Q_E \coloneqq i^{-1}_{\mathcal{G}_+\to\mathcal{G}'_+} \circ j^{-1}_{\mathcal{G}'_+\to\mathcal{F}'_+} \circ \widehat{Q_E} : D(Q_E) \to \mathcal{G}_+$$

is self-adjoint and $Q_E + \kappa_E \geq 1$. If additionally $E < \lambda_1$ then Q_E is also positive.

Proof. We define $\widehat{S} : \mathcal{F}_+ \to \mathcal{F}'_+$ using Riesz' theorem as the unique operator such that

$$[\widehat{S}z_+](y_+) = \langle z_+, y_+\rangle_{\mathcal{F}_+,E}\,.$$

The operator $\widehat{Q_E} = \widehat{S} - \kappa_E j_{\mathcal{G}'_+ \to \mathcal{F}'_+} i_{\mathcal{G}_+ \to \mathcal{G}'_+}$ then has the claimed properties. □

2.4. **The Definition of A_F.** We consider once more the operator L_E, viewed now as a mapping from $(\mathcal{F}_+, \|\cdot\|_{\mathcal{F}_+,E})$ into $(\mathcal{H}_-, \|\cdot\|_{\mathcal{H}})$. This operator is bounded and we denote its adjoint by $L'_E : (\mathcal{H}_-, \|\cdot\|_{\mathcal{H}}) \to (\mathcal{F}'_+, \|\cdot\|_{\mathcal{F}'_+,E})$, which is related to the Hilbert adjoint L^*_E by $L^*_E = i^{-1}_{\mathcal{F}_+ \to \mathcal{F}'_+} L'_E i_{\mathcal{H}_- \to \mathcal{H}'_-}$.

This allows us to define the operator $\widehat{R_E} : D(\widehat{R_E}) \subset \mathcal{F}_+ \times \mathcal{H}_- \to \mathcal{F}'_+ \times \mathcal{H}_-$ for $E > \lambda_0$ as

$$\begin{aligned}\widehat{R_E}\begin{pmatrix} x_+ \\ y_- \end{pmatrix} &= \begin{pmatrix} \mathbb{I} & -L'_E \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \widehat{Q_E} & 0 \\ 0 & -(B+E) \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ -L_E & \mathbb{I} \end{pmatrix} \begin{pmatrix} x_+ \\ y_- \end{pmatrix} \\ &= \begin{pmatrix} \widehat{Q_E} x_+ + L'_E(B+E)(y_- - L_E x_+) \\ -(B+E)(y_- - L_E x_+) \end{pmatrix}\end{aligned}$$

on the domain

$$D(\widehat{R_E}) = \left\{ \begin{pmatrix} x_+ \\ y_- \end{pmatrix} \in \mathcal{F}_+ \times \mathcal{H}_- : \, y_- - L_E x_+ \in D(B) \right\} \subset \mathcal{H}_+ \times \mathcal{H}_-.$$

The construction of $\widehat{R_E}$ should be compared to the decomposition (7). By the resolvent identity for any $x_+ \in F_+$

$$L_E x_+ - L_{E'} x_+ = (E' - E)(B+E)^{-1} L_{E'} x_+$$

and by Lemma 6 this identity extends to $\mathcal{F}_+$. Thus $D(\widehat{R_E})$ is independent of $E > \lambda_0$ and the same holds for the corresponding subset $\mathcal{F}$ of $\mathcal{H}$

$$\mathcal{F} \coloneqq \{x_+ + y_- \in \mathcal{F}_+ \oplus \mathcal{H}_- : \, y_- - L_E x_+ \in D(B)\} \subset \mathcal{H}_+ \oplus \mathcal{H}_- \,.$$

If $E < \lambda_1$ the operator $\widehat{Q_E}$ is invertible and in this case $\widehat{R_E}$ has an inverse defined by

$$\begin{aligned}\widehat{R_E}^{-1}\begin{pmatrix} x_+ \\ y_- \end{pmatrix} &= \begin{pmatrix} \mathbb{I} & 0 \\ L_E & \mathbb{I} \end{pmatrix} \begin{pmatrix} \widehat{Q_E}^{-1} & 0 \\ 0 & -(B+E)^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{I} & L'_E \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \ell_+ \\ k_- \end{pmatrix} \\ &= \begin{pmatrix} \widehat{Q_E}^{-1}(\ell_+ + L'_E k_-) \\ L_E \widehat{Q_E}^{-1}(\ell_+ + L'_E k_-) - (B+E)^{-1} k_- \end{pmatrix}\end{aligned}$$

for all $(\ell_+, k_-) \in \mathcal{F}'_+ \times \mathcal{H}_-$. It is straightforward to see that this operator maps into the domain of $\widehat{R_E}$ and vice versa.

Using the notation j_+ for the embedding $j_{\mathcal{H}_+ \to \mathcal{F}'_+}$ we now define $D(R) \subset \mathcal{F} \subset \mathcal{F}_+ \oplus \mathcal{H}_-$ to be the set

$$D(R) = \{x_+ + y_- \in \mathcal{F} : \widehat{Q_E} x_+ + L'_E(B+E)(y_- - L_E x_+) \in j_+(\mathcal{H}_+)\}$$

which will be proved to be independent of E. On this domain we define for $E > \lambda_0$ the family of operators $R_E : D(R) \to \mathcal{H}$ acting as

$$R_E(x_+ + y_-) = j_+^{-1}(\widehat{Q_E} x_+ + L'_E(B+E)(y_- - L_E x_+)) - (B+E)(y_- - L_E x_+) \,.$$

In the following we prove that R_E is an extension of $A-E$, that its domain is indeed independent of E and that it is self-adjoint.

To see that R_E is an extension of $A-E$, we note that for $x_+ \in F_+, y_- \in F_-$ we have $y_- - L_E x_+ = y_- - (B+E)^{-1}\Lambda_- A x_+ \in D(B)$. Furthermore, for any $u_+ \in F_+$ we compute that

$$\begin{aligned}&[\widehat{Q_E}x_+ + L'_E(B+E)(y_- - L_E x_+)](u_+)\\ &= \langle x_+, (A-E)u_+\rangle_{\mathcal{H}} + \langle (B+E)y_-, L_E u_+\rangle_{\mathcal{H}} = \langle (A-E)x_+, u_+\rangle_{\mathcal{H}} + \langle Ay_-, u_+\rangle_{\mathcal{H}}.\end{aligned}$$

The linear functional $\langle (A-E)x_+, \cdot\rangle_{\mathcal{H}} + \langle Ay_-, \cdot\rangle_{\mathcal{H}}$ is bounded on F_+ and extends continuously to $\mathcal{F}_+$. Hence $\widehat{Q_E}x_+ + L'_E(B+E)(y_- - L_E x_+) \in j_+(\mathcal{H}_+)$ and $F_+ \oplus F_- \subset D(R)$. Since in addition for any $v_- \in F_-$

$$\begin{aligned}-\langle (B+E)(y_- - L_E x_+), v_-\rangle_{\mathcal{H}} &= -\langle (B+E)y_-, v_-\rangle_{\mathcal{H}} + \langle \Lambda_- A x_+, v_-\rangle_{\mathcal{H}}\\ &= \langle (A-E)y_-, v_-\rangle_{\mathcal{H}} + \langle Ax_+, v_-\rangle_{\mathcal{H}}\end{aligned}$$

we obtain that for all $x_+, u_+ \in F_+, y_-, v_- \in F_-$

$$\langle R_E(x_+ + y_-), u_+ + v_-\rangle_{\mathcal{H}} = \langle (A-E)(x_+ + y_-), u_+ + v_-\rangle_{\mathcal{H}}$$

which allows us to conclude that R_E is an extension of $A-E$.

To show that $D(R)$ is independent of E, we first note that by the above for any $x_+, u_+ \in F_+$ and $y_- \in F_-$

$$\begin{aligned}&\overline{q_E}(x_+, u_+) + \langle y_- - L_E x_+, (B+E)L_E u_+\rangle_{\mathcal{H}} + (E-E')\langle x_+, u_+\rangle_{\mathcal{H}}\\ &= \overline{q_{E'}}(x_+, u_+) + \langle y_- - L_{E'}x_+, (B+E')L_{E'}u_+\rangle_{\mathcal{H}}\,.\end{aligned} \tag{12}$$

Let now $x_+ + y_- \in D(R) \subset \mathcal{F}_+ \oplus \mathcal{H}_-$, such that for some $E > \lambda_0$

$$y_- - L_E x_+ \in D(B)\,, \quad [\widehat{Q_E}x_+ + L'_E(B+E)(y_- - L_E x_+)] \in j_+(\mathcal{H}_+)\,.$$

We have already seen that then also $y_- - L_{E'}x_+ \in D(B)$ for any $E' > \lambda_0$. We can approximate $x_+ + y_-$ by elements of $x_+^{(n)} + y_-^{(n)} \in F_+ \oplus F_-$ such that

$$\left\|x_+ - x_+^{(n)}\right\|_{\mathcal{F}_+,E} \to 0\,, \quad \left\|y_- - y_-^{(n)}\right\|_{\mathcal{H}} \to 0\,.$$

By continuity (12) extends to $x_+ + y_-$ and we obtain that for all $u_+ \in F_+$

$$\begin{aligned}&[\widehat{Q_E}x_+ + L'_E(B+E)(y_- - L_E x_+)](u_+) + (E-E')\langle x_+, u_+\rangle_{\mathcal{H}}\\ &= [\widehat{Q_{E'}}x_+ + L'_{E'}(B+E')(y_- - L_{E'}x_+)](u_+)\end{aligned}$$

and thus also

$$y_- - L_{E'}x_+ \in D(B)\,, \quad [\widehat{Q_{E'}}x_+ + L'_{E'}(B+E')(y_- - L_{E'}x_+)] \in j_+(\mathcal{H}_+)\,.$$

To prove that R_E is symmetric, we compute that for given $u_+ + v_-$, $x_+ + y_- \in D(R)$

$$\begin{aligned}
&\langle R_E(x_+ + y_-), u_+ + v_- \rangle_{\mathcal{H}} \\
&= \langle j_+^{-1}(\widehat{Q_E} x_+ + L_E'(B+E)(y_- - L_E x_+)) - (B+E)(y_- - L_E x_+), u_+ + v_- \rangle_{\mathcal{H}} \\
&= \langle j_+^{-1}(\widehat{Q_E} x_+ + L_E'(B+E)(y_- - L_E x_+), u_+ \rangle_{\mathcal{H}} - \langle (B+E)(y_- - L_E x_+), v_- \rangle_{\mathcal{H}} \\
&= [\widehat{Q_E} x_+](u_+) + \langle (B+E)(y_- - L_E x_+), L_E u_+ \rangle_{\mathcal{H}} - \langle (B+E)(y_- - L_E x_+), v_- \rangle_{\mathcal{H}} \\
&= \overline{q_E}(x_+, u_+) - \langle (B+E)(y_- - L_E x_+), (v_- - L_E u_+) \rangle_{\mathcal{H}} .
\end{aligned}$$

This last expression is symmetric in interchanging $u_+ + v_-$ and $x_+ + y_-$ and hence R_E is a symmetric operator.

For $E < \lambda_1$, the operator $R_E^{-1} : \mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_- \to \mathcal{H}$ defined as

$$\begin{aligned}
&R_E^{-1}(x_+ + y_-) \\
&\quad = \widehat{Q_E}^{-1}(j_+(x_+) + L_E' y_-) + L_E \widehat{Q_E}^{-1}(j_+(x_+) + L_E' y_-) - (B+E)^{-1} y_-
\end{aligned}$$

is the inverse of R_E. It is itself symmetric and since defined on all of $\mathcal{H}$, self-adjoint. By the Hellinger–Toeplitz theorem it is also a bounded operator, hence closed. But then R_E itself as a bijective, symmetric and closed operator is also self-adjoint. The self-adjointness then extends to R_E for any $E > \lambda_0$.

Lastly, we define the self-adjoint extension A_F of A as $A_F \coloneqq R_E + E$ on $D(A_F) \coloneqq D(R)$.

2.5. **The Uniqueness of A_F.** Let $\widetilde{A}$ be another self-adjoint extension of A with $D(\widetilde{A}) \subset \mathcal{F}_+ \oplus \mathcal{H}_-$. We first show that then necessarily $D(\widetilde{A}) \subset \mathcal{F}$. For $x_+ + y_- \in D(\widetilde{A})$ and $v_- \in F_-$ we compute

$$\begin{aligned}
&\langle (\widetilde{A} - E)(x_+ + y_-), v_- \rangle_{\mathcal{H}} = \langle x_+ + y_-, (A - E) v_- \rangle_{\mathcal{H}} = \langle x_+ + y_-, (A_F - E) v_- \rangle_{\mathcal{H}} \\
&= -\langle y_- - L_E x_+, (B+E) v_- \rangle_{\mathcal{H}}
\end{aligned}$$

and we can conclude that $v_- \mapsto \langle y_- - L_E x_+, (B+E) v_- \rangle_{\mathcal{H}}$ is a continuous functional for all $v_- \in F_-$. This implies $y_- - L_E x_+ \in D(\Lambda_- A|_{F_-}^*) = D(B)$ and thus $x_+ + y_- \in \mathcal{F}$. Taking $u_+ \in F_+$ we further compute

$$\begin{aligned}
&\langle (\widetilde{A} - E)(x_+ + y_-), u_+ \rangle_{\mathcal{H}} = \langle x_+ + y_-, (A - E) u_+ \rangle_{\mathcal{H}} = \langle x_+ + y_-, (A_F - E) u_+ \rangle_{\mathcal{H}} \\
&= \langle x_+, j_+^{-1}(\widehat{Q_E} u_+ - L_E'(B+E) L_E u_+) \rangle_{\mathcal{H}} + \langle y_-, (B+E) L_E u_+ \rangle_{\mathcal{H}} \\
&= \overline{[\widehat{Q_E} u_+](x_+) - \langle (B+E) L_E u_+, L_E x_+ \rangle_{\mathcal{H}} + \langle (B+E) L_E u_+, y_- \rangle_{\mathcal{H}}} \\
&= [\widehat{Q_E} x_+](u_+) + \langle (B+E)(y_- - L_E x_+), L_E u_+ \rangle_{\mathcal{H}} \\
&= [\widehat{Q_E} x_+ + L_E'(B+E)(y_- - L_E x_+)](u_+) .
\end{aligned}$$

From this we can conclude that $x_+ + y_- \in D(A_F) = D(R)$ and thus $D(\widetilde{A}) \subset D(A_F)$. Conversely, by self-adjointness, $D(A_F) = D(A_F^*) \subset D(\widetilde{A}^*) = D(\widetilde{A})$, which proves the desired $\widetilde{A} = A_F$.

2.6. **The Proof of the Variational Principle.** It remains to prove that the variational principle holds. The min–max levels of Q_E on $(\mathcal{G}_+, \langle\cdot,\cdot\rangle_E)$ are given by

$$\begin{aligned}\mu_k(Q_E) &= \inf_{\substack{V\subset\mathcal{F}_+\\ \dim V=k}} \sup_{x_+\in V\setminus\{0\}} \frac{\overline{q_E}(x_+,x_+)}{\|x_+\|_E^2}\\ &= \inf_{\substack{V\subset F_+\\ \dim V=k}} \sup_{x_+\in V\setminus\{0\}} \frac{q_E(x_+,x_+)}{\|x_+\|_E^2}\end{aligned}$$

where we used that F_+ is a form core of $\overline{q_E}$. The numbers $\mu_k(Q_E)$ satisfy $\mu_k(Q_E) \le \inf\sigma_{ess}(Q_E)$ and if $\mu_k(Q_E) < \inf\sigma_{ess}(Q_E)$ then μ_k is an eigenvalue of Q_E with multiplicity

$$m_k(Q_E) = \#\{j\ge 1 : \mu_j(Q_E) = \mu_k(Q_E)\}\,.$$

We need the following result, which can be found in [10, Lemma 2.2].

Lemma 12. *Under the assumptions of Theorem 1, it holds that:*

(i) For any $x_+ \in F_+\setminus\{0\}$ the real number

$$E(x_+) := \sup_{z\in(\mathrm{span}(x_+)\oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2}$$

is the unique solution in $(\lambda_0,+\infty)$ of

$$q_E(x_+,x_+) = 0\,,$$

which may also be written as

$$E\|x_+\|_{\mathcal{H}}^2 = \langle x_+, Ax_+\rangle_{\mathcal{H}} + \langle \Lambda_- Ax_+, L_E x_+\rangle_{\mathcal{H}}\,.$$

(ii) The variational principle (5) is equivalent to

$$\lambda_k = \inf_{\substack{V\subset F_+\\ \dim V=k}} \sup_{x_+\in V\setminus\{0\}} E(x_+)\,.$$

(iii) For any $k\ge 1$ the real number λ_k given by (5) is the unique solution of

$$\mu_k(Q_\lambda) = 0\,.$$

Proof. First note that for fixed $x_+\in F_+\setminus\{0\}$, $q_E(x_+,x_+)$ is by Lemma 8 a strictly decreasing, continuous function of E with $q_{\lambda_1}(x_+,x_+)\ge 0$ and $\lim_{E\to\infty} q_E(x_+,x_+) = -\infty$, as can be seen from the definition. We can conclude that $q_E(x_+,x_+)=0$ has precisely one solution in $[\lambda_1,+\infty)$. Denote this solution by $\widetilde{E}(x_+)$. We now prove that $\widetilde{E}(x_+) = E(x_+)$.

If $E < \widetilde{E}(x_+)$ then necessarily $q_E(x_+,x_+) > 0$ and thus there exists a $y_-\in F_-$ such that

$$\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}} - E\|x_+ + y_-\|_{\mathcal{H}}^2 > 0\,.$$

We obtain that

$$E(x_+) = \sup_{z\in(\mathrm{span}(x_+)\oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2} \ge \frac{\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}}}{\|x_+ + y_-\|_{\mathcal{H}}^2} > E\,.$$

If $E > \widetilde{E}(x_+)$ then necessarily $q_E(x_+, x_+) \leq -\varepsilon < 0$ for some ε and thus for all $y_- \in F_-$

$$\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}} - E\|x_+ + y_-\|_{\mathcal{H}}^2 \leq -\varepsilon\,.$$

Consequently,

$$E(x_+) = \sup_{z \in (\mathrm{span}(x_+) \oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2} \leq -\varepsilon + E < E\,.$$

This proves that $\widetilde{E}(x_+) = E(x_+)$.

The statement (ii) is an immediate consequence of the definitions of $E(x_+)$ and λ_k as well as the observation that, since $\lambda_0 < \lambda_1$, for any k-dimensional subspace $V \subset F_+$

$$\sup_{z \in (V \oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2} = \sup_{\substack{z \in V \oplus F_- \\ \Lambda_+ z \neq 0}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2}\,.$$

Note that $\mu_k(Q_\lambda)$ is a continuous function of λ with $\mu_k(Q_{\lambda_1}) \geq 0$ by Lemma 8. Furthermore $\lim_{\lambda\to+\infty} \mu_k(Q_\lambda) = -\infty$ and we can conclude that $\mu_k(Q_\lambda) = 0$ has at least one solution in $[\lambda_1, +\infty)$. Denote this solution by $\widetilde{\lambda}_k$. We now prove that $\widetilde{\lambda}_k = \lambda_k$.

Assume $\lambda < \widetilde{\lambda}_k$. For all $V \subset F_+$ with $\dim V = k$ there exists an $x_+^V \in V \setminus \{0\}$ such that

$$q_{\widetilde{\lambda}_k}(x_+^V, x_+^V) \geq -\frac{\widetilde{\lambda}_k - \lambda}{2}\left\|x_+^V\right\|_{\widetilde{\lambda}_k}^2$$

and thus by Lemma 8

$$q_\lambda(x_+^V, x_+^V) \geq q_{\widetilde{\lambda}_k}(x_+^V, x_+^V) + (\widetilde{\lambda}_k - \lambda)\left\|x_+^V\right\|_{\widetilde{\lambda}_k}^2 \geq \frac{\widetilde{\lambda}_k - \lambda}{2}\left\|x_+^V\right\|_{\widetilde{\lambda}_k}^2 > 0\,.$$

This implies the existence of $y_-^V \in F_-$ such that

$$\varphi_{\lambda, x_+^V}(y_-^V) = \langle x_+^V + y_-^V, A(x_+^V + y_-^V)\rangle_{\mathcal{H}} - \lambda\left\|x_+^V + y_-^V\right\|_{\mathcal{H}}^2 \geq 0\,.$$

We obtain that

$$\sup_{z \in (V \oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|_{\mathcal{H}}^2} \geq \frac{\langle x_+^V + y_-^V, A(x_+^V + y_-^V)\rangle_{\mathcal{H}}}{\|x_+^V + y_-^V\|_{\mathcal{H}}^2} \geq \lambda$$

and thus $\lambda \leq \lambda_k$. Assume $\lambda > \widetilde{\lambda}_k$. There exists a vector space V_0 such that

$$q_{\widetilde{\lambda}_k}(x_+, x_+) \leq \frac{\lambda - \widetilde{\lambda}_k}{2C_{\widetilde{\lambda}_k, \lambda}}\|x_+\|_{\widetilde{\lambda}_k}^2 \leq \frac{\lambda - \widetilde{\lambda}_k}{2}\|x_+\|_\lambda^2$$

for all $x_+ \in V_0$ and thus by Lemma 8

$$q_\lambda(x_+, x_+) \leq q_{\widetilde{\lambda}_k}(x_+, x_+) - (\lambda - \widetilde{\lambda}_k)\|x_+\|_\lambda^2 \leq -\frac{\lambda - \widetilde{\lambda}_k}{2}\|x_+\|_\lambda^2 < 0$$

for all $x_+ \in V_0 \setminus \{0\}$. This implies that

$$\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}} - \lambda\|x_+ + y_-\|_{\mathcal{H}}^2 \leq 0$$

for all $x_+ \in V_0 \setminus \{0\}$ and all $y_- \in F_-$. We can conclude that

$$\begin{aligned}\lambda_k &\leq \sup_{z\in(V_0\oplus F_-)\setminus\{0\}} \frac{\langle z, Az\rangle_{\mathcal{H}}}{\|z\|^2_{\mathcal{H}}} \\ &\leq \max\Big(\sup_{x_+\in V_0\setminus\{0\}} \sup_{y_-\in F_-} \frac{\langle x_+ + y_-, A(x_+ + y_-)\rangle_{\mathcal{H}}}{\|x_+ + y_-\|^2_{\mathcal{H}}}, \sup_{y_-\in F_-\setminus\{0\}} \frac{\langle y_-, Ay_-\rangle_{\mathcal{H}}}{\|y_-\|^2_{\mathcal{H}}}\Big) \\ &\leq \max(\lambda, \lambda_0) = \lambda\end{aligned}$$

and together with the above $\lambda_k = \widetilde{\lambda}_k$.

□

To prove that the real numbers λ_k are in the spectrum of A_F, we use an argument presented in [10, Section 2] and construct a sequence of subspaces X_n of dimension d_k such that

$$\lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \sup_{y\in D(A_F)\setminus\{0\}} \frac{|\langle x, (A_F-\lambda_k)y\rangle_{\mathcal{H}}|}{\sqrt{\|y\|^2_{\mathcal{H}} + \|A_F y\|^2_{\mathcal{H}}}} = 0\,.$$

First note that

$$d_k \coloneqq \#\{j\geq 1:\, \lambda_j = \lambda_k\} = \#\{j\geq 1:\, \mu_j(Q_{\lambda_k}) = \mu_k(Q_{\lambda_k}) = 0\} = m_k(Q_{\lambda_k})$$

and by the min–max principle for Q_{λ_k}, there exists a sequence of spaces $X_n^+ \subset D(Q_{\lambda_k})$ of dimension d_k such that

$$\lim_{n\to\infty} \sup_{\substack{x_+\in X_n^+\\ \|x_+\|_{\lambda_k}=1}} \|Q_{\lambda_k} x_+\|_{\lambda_k} = 0\,,$$

which also implies that

$$\lim_{n\to\infty} \sup_{\substack{x_+\in X_n^+\\ \|x_+\|_{\lambda_k}=1}} \left\|\widehat{Q_{\lambda_k}} x\right\|_{\mathcal{F}'_+,\lambda_k} = \lim_{n\to\infty} \sup_{\substack{x_+\in X_n^+\\ \|x_+\|_{\lambda_k}=1}} \sup_{y_+\in\mathcal{F}_+} \frac{|[\widehat{Q_{\lambda_k}} x_+](y_+)|}{\|y_+\|_{\mathcal{F}_+,\lambda_k}} = 0\,. \quad (13)$$

Let $X_n \coloneqq (1 + L_{\lambda_k})X_n^+ \subset \mathcal{F}$. We observe that for all $x_+ \in \mathcal{F}_+$ and $y \in D(A_F)$

$$[\widehat{Q_{\lambda_k}}(x_+)](\Lambda_+ y) = \langle x_+ + L_{\lambda_k} x_+, (A_F - \lambda_k)y\rangle_{\mathcal{H}}\,. \quad (14)$$

Furthermore for all $y \in D(A_F)$

$$\begin{aligned}[\widehat{Q_{\lambda_k}}(\Lambda_+ y)](\Lambda_+ y) &= \langle \Lambda_+ y + L_{\lambda_k}\Lambda_+ y, (A_F - \lambda_k)y\rangle_{\mathcal{H}} \\ &\leq (\|y\|_{\mathcal{H}} + \|\Lambda_- y - L_{\lambda_k}\Lambda_+ y\|_{\mathcal{H}})(1+|\lambda_k|)(\|y\|_{\mathcal{H}} + \|A_F y\|_{\mathcal{H}})\end{aligned}$$

and using

$$\begin{aligned}\|\Lambda_- y - L_{\lambda_k}\Lambda_+ y\|_{\mathcal{H}} &= \left\|(B+\lambda_k)^{-1}\Lambda_-(A_F-\lambda_k)y\right\|_{\mathcal{H}} \\ &\leq \frac{1+|\lambda_k|}{\lambda_k - \lambda_0}(\|y\|_{\mathcal{H}} + \|A_F y\|_{\mathcal{H}})\end{aligned}$$

we can see that there exists a constant $C_{\lambda_k} > 0$ such that for all $y \in D(A_F)$

$$[\widehat{Q_{\lambda_k}}(\Lambda_+ y)](\Lambda_+ y) \leq C_{\lambda_k}(\|y\|^2_{\mathcal{H}} + \|A_F y\|^2_{\mathcal{H}})\,.$$

This allows us to bound

$$\begin{aligned}\|\Lambda_+ y\|^2_{\mathcal{F}_+,\lambda_k} &= \kappa_{\lambda_k}\|\Lambda_+ y_+\|^2_{\mathcal{H}} + \kappa_{\lambda_k}\|L_{\lambda_k}\Lambda_+ y\|^2_{\mathcal{H}} + [\widehat{Q_{\lambda_k}}(\Lambda_+ y)](\Lambda_+ y)\\ &\leq C'_{\lambda_k}(\|y\|^2_{\mathcal{H}} + \|A_F y\|^2_{\mathcal{H}})\end{aligned}$$

with some constant $C'_{\lambda_k} > 0$ for all $y \in D(A_F)$. Together with (13), (14) and the fact that $\|x_+ + L_{\lambda_k} x_+\|_{\mathcal{H}} = \|x_+\|_{\lambda_k}$ we can conclude that

$$\lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \sup_{y\in D(A_F)\setminus\{0\}} \frac{|\langle x, (A_F-\lambda_k)y\rangle_{\mathcal{H}}|}{\sqrt{\|y\|^2_{\mathcal{H}} + \|A_F y\|^2_{\mathcal{H}}}} = 0\,. \tag{15}$$

As $1 + L_{\lambda_k} : X_n^+ \to X_n$ is a surjective isometry we obtain that $\dim X_n = d_k$.

Consider the case $d_k < \infty$. Let P be the spectral measure of A_F. Suppose now for some $\varepsilon > 0$ we had that $\dim \operatorname{ran} P((\lambda_k - \varepsilon, \lambda_k + \varepsilon)) \leq d_k - 1$. Then there exists a sequence of $x_n \in X_n$ with $\|x_n\|_{\mathcal{H}} = 1$ and $P((\lambda_k - \varepsilon, \lambda_k + \varepsilon))x_n = 0$. We write $x_n = w_n + z_n$ with $w_n \in \operatorname{ran} P((-\infty, \lambda_k - \varepsilon])$ and $z_n \in \operatorname{ran} P([\lambda_k + \varepsilon, \infty))$. Unless λ_k is in the essential spectrum of A_F, we also observe that for some $\nu \in (\lambda_k - \varepsilon, \lambda_k + \varepsilon)$ necessarily $\nu \in \rho(A_F)$. Choosing $y_n = (A_F - \nu)^{-1} x_n \in D(A_F)$ we compute that, if $\lambda_k - \varepsilon < \nu \leq \lambda_k$,

$$\begin{aligned}\langle x_n, (A-\lambda_k)y_n\rangle_{\mathcal{H}} &= \int_{-\infty}^{\lambda_k-\varepsilon} \frac{\lambda-\lambda_k}{\lambda-\nu}\,\mathrm{d}P_{w_n,w_n}(\lambda) + \int_{\lambda_k+\varepsilon}^{\infty} \frac{\lambda-\lambda_k}{\lambda-\nu}\,\mathrm{d}P_{z_n,z_n}(\lambda)\\ &\geq \|w_n\|^2_{\mathcal{H}} + \frac{\varepsilon}{\lambda_k+\varepsilon-\nu}\|z_n\|^2_{\mathcal{H}}\\ &\geq \min\left(1, \frac{\varepsilon}{\lambda_k+\varepsilon-\nu}\right)\|x_n\|^2_{\mathcal{H}}\end{aligned}$$

and similarly, if $\lambda_k \leq \nu < \lambda_k + \varepsilon$,

$$\langle x_n, (A-\lambda_k)y_n\rangle_{\mathcal{H}} \geq \min\left(\frac{\varepsilon}{\nu-\lambda_k+\varepsilon}, 1\right)\|x_n\|^2_{\mathcal{H}}\,.$$

Since $\|x_n\|_{\mathcal{H}} = 1$ and

$$\|y_n\|^2_{\mathcal{H}} + \|A_F y_n\|^2_{\mathcal{H}} = \left\|(A_F-\nu)^{-1}x_n\right\|^2_{\mathcal{H}} + \left\|x_n + \nu(A_F-\nu)^{-1}x_n\right\|^2_{\mathcal{H}} \leq C\|x_n\|^2_{\mathcal{H}}$$

we obtain that for some constant $C' > 0$

$$|\langle x_n, (A_F-\lambda_k)y_n\rangle_{\mathcal{H}}| \geq C'\sqrt{\|y_n\|^2_{\mathcal{H}} + \|A_F y_n\|^2_{\mathcal{H}}}\,,$$

which contradicts (15). Thus necessarily $\dim \operatorname{ran} P((\lambda_k - \varepsilon, \lambda_k + \varepsilon)) \geq d_k$ for any $\varepsilon > 0$.

In the case $d_k = \infty$, we can use the above argument to conclude that $\dim \operatorname{ran} P(\lambda_k - \varepsilon, \lambda_k + \varepsilon) = \infty$ for all $\varepsilon > 0$. As a consequence $\lambda_k \in \sigma(A_F)$ and λ_k is larger or equal to the k-th eigenvalue $\mu_k(A_F)$ of A_F in $(\lambda_0, \sup_{\ell\geq 1}\lambda_\ell)$.

Before we prove that the λ_k are all the points in $\sigma(A_F) \cap (\lambda_0, \sup_{\ell\geq 1}\lambda_\ell)$, we first note that (see Remark 5)

$$\begin{aligned}\lambda_0 = \sup_{y_-\in F_-\setminus\{0\}} \frac{-\langle y_-, By_-\rangle_{\mathcal{H}}}{\|y_-\|^2_{\mathcal{H}}} &= \sup_{y_-\in \mathcal{F}_-\setminus\{0\}} \frac{-[\widehat{B}y_-](y_-)}{\|y_-\|^2_{\mathcal{H}}} \\ &= \sup_{y_-\in \mathcal{F}_-\cap D(A_F)\setminus\{0\}} \frac{\langle y_-, A_F y_-\rangle_{\mathcal{H}}}{\|y_-\|^2_{\mathcal{H}}}\end{aligned} \tag{16}$$

as an immediate consequence of the continuity of $\widehat{B}$ with respect to $\|\cdot\|_{\mathcal{F}_-}$.

Now assume that $\lambda \in \sigma(A_F) \cap (\lambda_0, \sup_{\ell\geq 1}\lambda_\ell)$ with spectral multiplicity d. We have to show that $\lambda = \lambda_k$ for some $k \in \mathbb{N}$, or equivalently that $\mu_k(Q_\lambda) = 0$ for some $k \in \mathbb{N}$. By assumption there exist spaces $X_n \subset D(A_F)$ with $\dim X_n = d$ such that

$$\lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \|(A_F - \lambda)x\|_{\mathcal{H}} = 0\,.$$

In particular we obtain that

$$\lim_{n\to\infty} \sup_{x\in X_n} \frac{\|(B+\lambda)(\Lambda_- x - L_\lambda\Lambda_+ x)\|_{\mathcal{H}}}{\|x\|_{\mathcal{H}}} = 0$$

and since $(B+\lambda)^{-1}$ is a bounded operator

$$\lim_{n\to\infty} \sup_{x\in X_n} \frac{\|\Lambda_- x - L_\lambda\Lambda_+ x\|_{\mathcal{H}}}{\|x\|_{\mathcal{H}}} = 0\,.$$

We can conclude that there exists an $N \in \mathbb{N}$ such that

$$\|\Lambda_- x - L_\lambda\Lambda_+ x\|_{\mathcal{H}} \leq \frac{\|x\|_{\mathcal{H}}}{2}$$

for all $x \in X_n$ with $n \geq N$. In the remainder we assume without loss of generality that $N = 1$. Note that

$$\begin{aligned}0 &= \lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \|(A_F-\lambda)x\|_{\mathcal{H}} = \lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \sup_{\substack{y\in\mathcal{H}\\ \|y\|_{\mathcal{H}}=1}} |\langle (A_F-\lambda)x, y\rangle_{\mathcal{H}}| \\ &= \lim_{n\to\infty} \sup_{\substack{x\in X_n, y\in\mathcal{F}\\ \|x\|_{\mathcal{H}}=1, \|y\|_{\mathcal{H}}=1}} |[\widehat{Q_\lambda}\Lambda_+ x](\Lambda_+ y) - \langle (B+\lambda)(\Lambda_- x - L_\lambda\Lambda_+ x), (\Lambda_- y - L_\lambda\Lambda_+ y)\rangle_{\mathcal{H}}|\end{aligned}$$

and in particular

$$\begin{aligned}0 &= \lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \sup_{\substack{y\in(1+L_\lambda)F_+\\ \|y\|_{\mathcal{H}}=1}} |\langle (A_F-\lambda)x, y\rangle_{\mathcal{H}}| \\ &= \lim_{n\to\infty} \sup_{\substack{x\in X_n\\ \|x\|_{\mathcal{H}}=1}} \sup_{\substack{y_+\in F_+\\ \|y_+\|_\lambda=1}} |[\widehat{Q_\lambda}(\Lambda_+ x)](y_+)|\,.\end{aligned}$$

Now let $X_n^+ := \Lambda_+ X_n \subset \mathcal{F}_+$. If $x \in X_n$ is an element of $\mathcal{F}_-$, then by (16)

$$\langle (A_F-\lambda)x, x\rangle_{\mathcal{H}} \leq (\lambda_0 - \lambda)\|x\|^2_{\mathcal{H}}$$

and thus

$$|\langle (A_F - \lambda)x, x\rangle_{\mathcal{H}}| \geq (\lambda - \lambda_0)\|x\|_{\mathcal{H}}^2$$

which is a contradiction to the definition of X_n. Thus $\dim X_n^+ = d$. Furthermore for $x \in X_n$ by an application of the lower triangle inequality

$$\|\Lambda_+ x\|_\lambda = \|\Lambda_+ x_+ + L_\lambda \Lambda_+ x\|_{\mathcal{H}} = \|x - (\Lambda_- x - L_\lambda \Lambda_+ x)\|_{\mathcal{H}} \geq \frac{\|x\|_{\mathcal{H}}}{2}\,.$$

As a consequence

$$\lim_{n\to\infty} \sup_{\substack{x_+ \in X_n^+ \\ \|x_+\|_\lambda = 1}} \sup_{\substack{y_+ \in F_+ \\ \|y_+\|_\lambda = 1}} |[\widehat{Q_\lambda}(x_+)](y_+)| = 0$$

and thus also

$$\lim_{n\to\infty} \sup_{\substack{x_+ \in X_n^+ \\ \|x_+\|_\lambda = 1}} \left\|\widehat{Q_\lambda} x_+\right\|_{\mathcal{F}'_+,\lambda} = 0\,. \tag{17}$$

This implies that zero is in the spectrum of Q_λ by a generalised version of Weyl's criterion, which can for example be found in [24]. To prove this taking into account the multiplicity d, we let $\varepsilon > 0$ and let P be the spectral measure of Q_λ. If $\dim \operatorname{ran} P((-\varepsilon,\varepsilon)) \leq d-1$ then we can find a sequence of $x_n \in X_n^+$ with $\|x_n\|_\lambda = 1$ and $P((-\varepsilon,\varepsilon))x_n = 0$. Using the embedding $j = j_{\mathcal{G}'_+ \to \mathcal{F}'_+} \circ i_{\mathcal{G}_+ \to \mathcal{G}'_+}$ we can compute that for any $x \in \mathcal{F}_+$

$$\begin{aligned}
\left\|\widehat{Q_\lambda} x\right\|_{\mathcal{F}'_+,\lambda} &= \left\|(\widehat{Q_\lambda} + \kappa_\lambda j)^{-1} \widehat{Q_\lambda} x\right\|_{\mathcal{F}_+,\lambda} \\
&= \left\|(Q_\lambda + \kappa_\lambda)^{\frac{1}{2}} (x - \kappa_\lambda (\widehat{Q_\lambda} + \kappa_\lambda j)^{-1} j(x))\right\|_\lambda \\
&= \left\|(Q_\lambda + \kappa_\lambda)^{\frac{1}{2}} (x - \kappa_\lambda (Q_\lambda + \kappa_\lambda)^{-1} x)\right\|_\lambda = \left\|Q_\lambda (Q_\lambda + \kappa_\lambda)^{-\frac{1}{2}} x\right\|_\lambda .
\end{aligned}$$

Together with the spectral theorem we obtain that

$$\left\|\widehat{Q_\lambda} x_n\right\|_{\mathcal{F}'_+,\lambda}^2 = \int_{-\kappa_\lambda+1}^{-\varepsilon} \frac{t^2}{t+\kappa_\lambda}\,\mathrm{d}P_{x_n,x_n}(t) + \int_\varepsilon^\infty \frac{t^2}{t+\kappa_\lambda}\,\mathrm{d}P_{x_n,x_n}(t) \geq \frac{\varepsilon^2}{\varepsilon+\kappa_\lambda}\|x\|_\lambda^2$$

which is a contradiction to (17). It remains to prove that $0 = \mu_k(Q_\lambda)$, for some $k \in \mathbb{N}$.

Since $\lambda < \sup_{\ell\geq 1} \lambda_\ell$ there exists an $\ell \in \mathbb{N}$ such that $\lambda < \lambda_\ell$. By definition $\mu_\ell(Q_{\lambda_\ell}) = 0$ and thus for any subspace $V \subset F_+$ of dimension ℓ there exists an $x_+^V \in V$ such that

$$q_{\lambda_\ell}(x_+, x_+) \geq -\varepsilon \|x_+\|_\lambda^2\,.$$

By Lemma 8 we obtain that

$$q_\lambda(x_+^V, x_+^V) \geq q_{\lambda_\ell}(x_+^V, x_+^V) \geq -\varepsilon \left\|x_+^V\right\|_{\lambda_\ell}^2 \geq -\varepsilon \left\|x_+^V\right\|_\lambda^2 .$$

This implies that $\mu_\ell(Q_\lambda) \geq 0$, and consequently $0 = \mu_k(Q_\lambda)$ for some $k < \ell$. We can conclude that $\lambda = \lambda_k$, which completes the proof of Theorem 1.

3. Application to the Dirac–Coulomb Operator

Let $H_0 = -\mathrm{i}\alpha\cdot\nabla+\beta$ be the free Dirac operator where $\alpha^1,\alpha^2,\alpha^3,\beta\in\mathbb{C}^{4\times4}$ with

$$\alpha^i\alpha^j+\alpha^j\alpha^i = 2\delta_{ij}\mathbb{I}_{\mathbb{C}^4}\,,\quad \alpha^i\beta+\beta\alpha^i = 0\,,\quad \beta^2=\mathbb{I}_{\mathbb{C}^4}\,.$$

We choose the representation

$$\alpha^i = \begin{pmatrix}0 & \sigma^i\\ \sigma^i & 0\end{pmatrix},\quad \beta=\begin{pmatrix}\mathbb{I}_{\mathbb{C}^2} & 0\\ 0 & -\mathbb{I}_{\mathbb{C}^2}\end{pmatrix}.$$

The free Dirac operator is essentially self-adjoint on $\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$. The Dirac–Coulomb operator $H_\nu = H_0-\nu/|x|$ is symmetric on $D(H_\nu)=\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)\subset L^2(\mathbb{R}^3;\mathbb{C}^4)=:\mathcal{H}$. Let $\Lambda_\pm$ be the Talman projections,

$$\Lambda_+\begin{pmatrix}\varphi\\ \psi\end{pmatrix}=\begin{pmatrix}\varphi\\ 0\end{pmatrix},\quad \Lambda_-\begin{pmatrix}\varphi\\ \psi\end{pmatrix}=\begin{pmatrix}0\\ \psi\end{pmatrix}.$$

Then clearly $\Lambda_\pm D(H_\nu)\subset D(H_\nu)$ and thus the first assumption of Theorem 1 is satisfied. We further compute that

$$\lambda_0 = \sup_{\psi\in\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)\setminus\{0\}}\frac{\int_{\mathbb{R}^3}(-1-\frac{\nu}{|x|})|\psi(x)|^2\,\mathrm{d}x}{\|\psi\|_{\mathcal{H}}^2} = \sup_{x\in\mathbb{R}^3}(-1-\nu/|x|) = -1\,.$$

Dolbeault, Esteban, Loss and Vega [9] proved the Hardy inequality

$$\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\psi(x)|^2}{1+\frac{1}{|x|}}\,\mathrm{d}x+\int_{\mathbb{R}^3}\left(1-\frac{1}{|x|}\right)|\psi(x)|^2\,\mathrm{d}x\geq 0 \tag{18}$$

for all $\psi\in H^1(\mathbb{R}^3;\mathbb{C}^2)$ by analytic methods. Following similar computations in [13, 12] we can use (18) to prove that $q_0(\psi,\psi)\geq 0$ for all $\psi\in\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)$ and all $\nu\in[0,1]$. Here q_E is the Schur complement

$$q_E(\psi,\psi)=\int_{\mathbb{R}^3}\left(1-\frac{\nu}{|x|}-E\right)|\psi(x)|^2\,\mathrm{d}x+\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\psi(x)|^2}{1+\frac{\nu}{|x|}+E}\,\mathrm{d}x\,.$$

As a consequence of Lemma 8 (*iii*) we obtain $\lambda_1\geq 0>\lambda_0$. Note that this statement can also be proved by means of an abstract continuation principle [10, Section 3] and can then in turn be used to establish the Hardy inequality (18) [10, Section 4]. Since in addition $-1-\nu/|x|$ is essentially self-adjoint on $\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)$, all the conditions of Theorem 1 are satisfied and thus for any $\nu\in[0,1]$ there exists a self-adjoint extension of H_ν with eigenvalues given by

$$\lambda_k = \inf_{\substack{V\subset\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2)\\ \dim V=k}}\sup_{\psi\in(V\times\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^2))\setminus\{0\}}\frac{\langle\psi,H_\nu\psi\rangle_{\mathcal{H}}}{\|\psi\|_{\mathcal{H}}^2}\,.$$

The self-adjoint extension coincides with extension constructed in [13] and thus for $\nu<1$ also with the extensions of Schmincke [30], Wüst [34, 35] and Nenciu [27] (which were all proved to be equal by Klaus and Wüst [21]). The variational principle for this distinguished extension is the same as the one obtained in [12].

To establish a second variational principle, we can choose $\Lambda_\pm$ to be spectral projections of the free Dirac operator,

$$\Lambda_+ = P_{H_0}[0,\infty)\,, \quad \Lambda_- = P_{H_0}(-\infty,0)\,.$$

Let H_ν again denote the Dirac operator with Coulomb potential acting on the domain $D(H_\nu) = F_+ \oplus F_- \subset L^2(\mathbb{R}^3;\mathbb{C}^4) =: \mathcal{H}$ with $F_\pm = \Lambda_\pm \mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$. The operator H_ν is symmetric and the first assumption of Theorem 1 is satisfied. Again we can compute λ_0 to be

$$\lambda_0 = \sup_{\psi\in F_-\setminus\{0\}} \frac{\langle\psi, (-\sqrt{1-\Delta} - \nu/|x|)\psi\rangle_{\mathcal{H}}}{\|\psi\|^2_{\mathcal{H}}} \leq \sup_{x\in\mathbb{R}^3}(-1-\nu/|x|) = -1\,.$$

Using an abstract continuation principle, it was proved in [10] that $\lambda_1 \geq 0 > \lambda_0$ for $\nu \in [0,1)$. To extend this result to the endpoint $\nu = 1$ we note that by the above for any $\nu \in [0,1)$ the Schur complement q_E

$$\begin{aligned} q_E(\psi,\psi) = &\left\langle \psi, \Big(\sqrt{1-\Delta} - \frac{\nu}{|x|} - E\Big)\psi\right\rangle_{\mathcal{H}} \\ &+ \left\langle \Lambda_- \frac{\psi}{|x|}, \Big(\Lambda_-\Big(\sqrt{1-\Delta} + \frac{\nu}{|x|} + E\Big)\Lambda_-\Big)^{-1}\Lambda_- \frac{\psi}{|x|}\right\rangle_{\mathcal{H}} \end{aligned}$$

satisfies $q_0(\psi,\psi) \geq 0$ for all $\psi \in \Lambda_+ H^{1/2}(\mathbb{R}^3)$. Taking the limit $\nu \to 1$ one obtains (see [12, Lemma 15]) the analogue of (18)

$$\left\langle \psi, \Big(\sqrt{1-\Delta} - \frac{1}{|x|}\Big)\psi\right\rangle_{\mathcal{H}} + \left\langle \Lambda_- \frac{\psi}{|x|}, \Big(\Lambda_-\Big(\sqrt{1-\Delta} + \frac{1}{|x|}\Big)\Lambda_-\Big)^{-1}\Lambda_- \frac{\psi}{|x|}\right\rangle_{\mathcal{H}} \geq 0\,.$$

In contrast to the case of the Talman projections, we are not aware of an analytic proof of this inequality. By Lemma 8 (*iii*) this inequality proves that $\lambda_1 \geq 0 > \lambda_0$ still holds in the endpoint case $\nu = 1$. As discussed in the appendix, the operator $\Lambda_-(\sqrt{1-\Delta} + \nu/|x|)|_{F_-}$ is essentially self-adjoint on $\Lambda_-\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4) \subset \Lambda_- L^2(\mathbb{R}^3;\mathbb{C}^4)$. Thus all the conditions of Theorem 1 are satisfied and for any $\nu \in [0,1]$ we obtain a a self-adjoint extension of H_ν with eigenvalues given by

$$\lambda_k = \inf_{\substack{V\subset\Lambda_+\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)\\ \dim V = k}} \sup_{\psi\in(V\oplus\Lambda_-\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4))\setminus\{0\}} \frac{\langle\psi, H_\nu\psi\rangle_{\mathcal{H}}}{\|\psi\|_{\mathcal{H}}}\,.$$

This is the same result as in [12].

4. Self-adjoint Extensions and the APS Boundary Condition

We consider an operator of the form

$$A = \sigma(\partial_x + B) \tag{19}$$

acting on functions in $\mathcal{H} = L^2((0,1);\mathcal{K})$ with $\mathcal{K}$ being a complex Hilbert space. The operator B is densely defined on a domain $D(B) \subset \mathcal{K}$ and does not depend on x. It is self-adjoint and has discrete spectrum. The map σ is an automorphism on $\mathcal{K}$, equally independent of x. Furthermore, we make the following assumptions:

(i) $\sigma^2 = -\mathbb{I}$, $\sigma^* = -\sigma$,

(ii) $\{B, \sigma\} = B\sigma + \sigma B = 0$,
(iii) $\dim\ker B < \infty$, $\ker B = \mathcal{N}_+ \oplus \mathcal{N}_-$ for some subspaces $\mathcal{N}_\pm \subset D(B)$ and
(iv) $\sigma(\mathcal{N}_-) = \mathcal{N}_+$.

On the domain $D(A) = \mathcal{C}_0^1((0,1); D(B))$, A is a well-defined symmetric operator. Continuity and differentiability on the set $D(A)$ are defined with respect to the graph norm of B on $D(B)$.

Remark 13.

- As an example we could take $\mathcal{K}$ to be the Hilbert space of square integrable functions on $\mathbb{S}^1$, parametrised by a variable y. If then B is defined on the continuously differentiable periodic functions as $B = i\sigma_3\partial_y$ and $\sigma = i\sigma_2$, with the Pauli matrices σ_i, we may identify A with the Dirac operator on the cylinder of height one.
- More generally, we may think of B being any first-order differential operator on some closed manifold Σ such that A represents a differential operator (of first order) on a generalised cylinder. In fact, any first-order elliptic operator on a compact manifold M with boundary Σ takes the form $A = \sigma(\partial_x + B)$ on a collar neighbourhood of the boundary [1], but B and σ are not necessarily independent of x.

These special cases are included in our considerations but we do not restrict ourselves to them.

From assumptions (i), (iii) and (iv) we conclude that the kernel of B is of even dimension, hence $\dim\ker B = 2N_0$ for some $N_0 \in \mathbb{N}_0$. Furthermore, the vanishing anticommutator $\{B, \sigma\}$ implies that σ maps elements from the positive spectral subspace of B to the negative spectral subspace and vice versa.

It is well-known (cf. e.g. [11], [16]) that $A|_{\mathcal{C}_0^1((0,1);D(B))}$ has a self-adjoint extension characterised by a non-local boundary condition known as the 'Atiyah–Patodi–Singer boundary condition'. Let us denote by $P^+_{B>0}$ the projection onto the sum of $\mathcal{N}_+$ and the positive spectral subspace of B. Then the following holds.

Proposition 14 (APS)**.** *The operator $A = \sigma(\partial_x + B)$ from* (19) *is self-adjoint on the domain*

$$D(A_{APS}) = \{f \in L^2((0,1); D(B)) \cap H^1((0,1); \mathcal{K}) :\ P^+_{B>0} f|_{x=0} = 0 \\ \text{and } P^+_{B>0} f|_{x=1} = 0\}. \tag{20}$$

Also here, $D(B)$ and $\mathcal{K}$ are to be understood as Hilbert spaces with their respective norms, in the sense that $f \in D(A_{APS})$ is a function such that $\|f\|$, $\|Bf\|$ and $\|\partial_x f\|$ are all square-integrable. A proof of this proposition follows by explicit calculation.

Starting from the symmetric operator $A = \sigma(\partial_x + B)$ in (19) we can show that it falls into the class of gapped operators for which our construction of a self-adjoint extension applies. Indeed we find:

Theorem 15. *Theorem 1 applies to the operator $A = \sigma(\partial_x + B)$ defined on $\mathcal{C}^1_0((0,1); D(B))$. The self-adjoint extension constructed in this way coincides with the Atiyah–Patodi–Singer extension A_{APS} from Proposition 14.*

In this sense the extension from Theorem 1 is characterised by the global boundary conditions from (20). We prove Theorem 15 in the remainder of this section.

Let $P^-_{B<0}$ the orthogonal projection complementary to $P^+_{B>0}$, i.e. the projection onto the sum of $\mathcal{N}_-$ and the negative spectral subspace of B. If we write $\mathcal{K}_\pm = P^\pm_{B\gtrless 0}\mathcal{K}$ and $\mathcal{H}_\pm = L^2((0,1);\mathcal{K}_\pm)$ then $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$ is an orthogonal decomposition of $\mathcal{H}$ with corresponding orthogonal projections $\Lambda_\pm$ such that

$$F_\pm = \Lambda_\pm D(A) = \mathcal{C}^1_0((0,1); \mathcal{K}_\pm \cap D(B)) \subset D(A)\,.$$

It is easy to see that $\sigma(\mathcal{K}_+) = \mathcal{K}_-$ and $\sigma(\mathcal{K}_-) = \mathcal{K}_+$. We can conclude that $\Lambda_- A \Lambda_- = 0$ and hence $\Lambda_- A \Lambda_-$ is essentially self-adjoint on $\mathcal{C}^1_0((0,1); D(B))$ and $\lambda_0 = 0$.

Let ℓ_k, $k \in \mathbb{Z}\setminus\{0\}$ be the eigenvalues of B such that $\ell_k \geq \ell_{k'}$ if $k > k'$ and $\ell_k = 0$ for $-N_0 \leq k \leq N_0$. We denote the corresponding eigenvectors of B by φ_k and assume they are chosen such that $\sigma\varphi_k = -\varphi_{-k}$ and $\sigma\varphi_{-k} = \varphi_k$ for $k > 0$. Any function $u \in F_+$ has then an expansion

$$u(x) = \sum_{k>0} u_k(x)\varphi_k$$

with functions $u_k \in \mathcal{C}^1_0((0,1);\mathbb{C})$ and similarly for $v \in F_-$.

We can then write for any $u \in F_+$ and any $v \in F_-$

$$\begin{aligned}
&\langle u+v, \sigma(\partial_x + B)(u+v)\rangle \\
&= \sum_{k>0}\sum_{l<0} \langle u_k\varphi_k, \sigma(\partial_x + B)v_l\varphi_l\rangle + \text{complex conjugate} \\
&= \sum_{k>0}\sum_{l<0} \langle u_k\varphi_k, (\partial_x - B)v_l\varphi_{-l}\rangle + \text{c.c.} \\
&= \sum_{k,l>0} \langle u_k\varphi_k, (\partial_x - \ell_l)v_{-l}\varphi_l\rangle + \text{c.c.} \\
&= \sum_{k>0} \langle u_k, (\partial_x - \ell_k)v_{-k}\rangle + \text{c.c.}\ .
\end{aligned}$$

Hence, we can rewrite the expectation value of A as

$$\begin{aligned}
&\frac{\sum_{k>0}\langle u_k, (\partial_x - \ell_k)v_{-k}\rangle + \text{c.c.}}{\|u\|^2 + \|v\|^2} \\
&= \sum_{k>0:u_k\neq 0 \text{ or } v_{-k}\neq 0} \frac{\|u_k\|^2 + \|v_{-k}\|^2}{\|u\|^2 + \|v\|^2} \frac{\langle u_k, (\partial_x - \ell_k)v_{-k}\rangle + \text{c.c.}}{\|u_k\|^2 + \|v_{-k}\|^2}
\end{aligned}$$

which we can identify with an arithmetic mean of expectation values for single k's weighted according to their norm. Clearly, this expression is bounded

by

$$\sup_{k>0:u_k\neq 0 \text{ or } v_{-k}\neq 0} \frac{\langle u_k, (\partial_x - \ell_k)v_{-k}\rangle + \text{c.c.}}{\|u_k\|^2 + \|v_{-k}\|^2} \ .$$

Taking the supremum over $v \in F_-$ finally gives an upper bound

$$\sup_{v\in F_-} \frac{\langle u+v, A(u+v)\rangle}{\|u\|^2 + \|v\|^2} \leq \sup_{k>0} \sup_{v_{-k}\neq 0} \frac{\langle u_k, (\partial_x - \ell_k)v_{-k}\rangle + \text{c.c.}}{\|u_k\|^2 + \|v_{-k}\|^2}$$

and indeed, equality holds since we can always find a sequence of v approximating the right-hand-side. Clearly, we find the supremum over v_{-k} by choosing $v_{-k} = \lambda(-\partial_x - \ell_k)u_k$ for some real number λ. Maximising over all values of λ we find $\lambda = \frac{\|u_k\|}{\|(-\partial_x - \ell_k)u_k\|}$ and hence

$$\sup_{v\in F_-} \frac{\langle u+v, A(u+v)\rangle}{\|u\|^2 + \|v\|^2} = \sup_{\substack{k>0:\\ u_k\neq 0}} \frac{\|(-\partial_x - \ell_k)u_k\|}{\|u_k\|} \ .$$

Note that the left-hand side is precisely $E(u)$ as defined in Lemma 12. By construction the supremum is achieved at $v = L_{E(u)}u$ which coincides with the relation $v_{-k} = \lambda(-\partial_x - \ell_k)u_k$ above for $\lambda = E(u)^{-1}$.

Taking the infimum over all $u \neq 0$ we then obtain

$$\lambda_1 = \inf_{\substack{u\in \mathcal{C}_0^1((0,1);\mathcal{K}_+\cap D(B))\\ u\neq 0}} \sup_{\substack{k>0:\\ u_k\neq 0}} \left[\ell_k^2 + \frac{\|-\partial_x u_k\|^2}{\|u_k\|^2}\right]^{\frac{1}{2}} = \pi > 0 \qquad (21)$$

if there is an $\ell_k = 0$. We have used that by the variational principle for the Friedrichs extension of the Laplace operator the last term gives the lowest eigenvalue of the Dirichlet Laplacian on $L^2((0,1))$. If $\ker B = \{0\}$, then $\lambda_1 > \pi$. In both cases, A is a gapped operator and all assumptions for constructing a self-adjoint extension as in Theorem 1 are satisfied.

For comparison with the APS-extension we are interested in the domain of A_F, that is in particular in how the Hilbert space $\mathcal{F}_+$ appears in this setting. Recall that $\mathcal{F}_+$ is the closure of F_+ in the norm $\|\cdot\|_{F_+,E}$ constructed from the quadratic form q_E and the graph norm of the operator L_E

$$\|u\|^2_{F_+,E} = q_E(u,u) + \kappa_E \|L_E u\|^2_{\mathcal{H}} \ ,$$

where in our setting $q_E(u,u) = -E\|u\|^2 + \frac{1}{E}(\|\partial_x u\|^2 + \|Bu\|^2)$ and $\|L_E u\|^2 = \frac{1}{E^2}(\|\partial_x u\|^2 + \|Bu\|^2)$. Using that

$$\lambda_1^2 \|u\|^2 \leq \|\partial_x u\|^2 + \|Bu\|^2 \, ,$$

which follows from comparison with (21), it is directly seen that $q_E(u,u) \geq (\lambda_1 - E)\|u\|^2 + (\frac{1}{E} - \frac{1}{\lambda_1})(\|\partial_x u\|^2 + \|Bu\|^2)$ and hence $\|\cdot\|_{F_+,E}$ is equivalent to the sum of norms $\|\cdot\| + \|\partial_x \cdot\| + \|B\cdot\|$ as long as $E < \lambda_1$. Closing F_+ in this norm gives the Hilbert space $L^2((0,1); D(B)\cap\mathcal{K}_+)\cap H_0^1((0,1);\mathcal{K}_+)$, as can be seen from the following argument. Given an $f \in L^2((0,1); D(B)\cap\mathcal{K}_+)\cap H_0^1((0,1);\mathcal{K}_+)$, we use the standard approximation by smooth, compactly supported functions by mollification of a function in $H_0^1((0,1);\mathcal{K}_+)$, which allows to construct a sequence of $f_n \in \mathcal{C}_0^1((0,1); D(B)\cap\mathcal{K}_+)$ that converges

to f in the norm $\|\cdot\| + \|\partial_x \cdot\|$, see e.g. [15, Chapter 5.5]. Since $Bf \in L^2((0,1);\mathcal{K}_+)$, it may be approximated in the same fashion such that the sequence $\{f_n\}$ will also converge in $\|\cdot\| + \|B\cdot\|$. These considerations show that indeed $D(A_{APS}) \subset \mathcal{F}_+ \oplus \mathcal{H}_-$ and thus A_{APS} coincides with A_F by the uniqueness property proved in Theorem 1.

Appendix A. Essential Self-Adjointness of the Brown–Ravenhall Operator

Let H_0 be the self-adjoint free Dirac operator with domain $H^1(\mathbb{R}^3;\mathbb{C}^4) \subset L^2(\mathbb{R}^3;\mathbb{C}^4)$ and denote by $\Lambda_\pm$ the projections onto the positive/negative spectral subspace of H_0. For $\gamma \in \mathbb{R}$ the Brown–Ravenhall operator [5] is defined as

$$B_\gamma = \Lambda_+(H_0 - \gamma/|x|)\Lambda_+ = \Lambda_+(\sqrt{1-\Delta} - \gamma/|x|)\Lambda_+$$

on the Hilbert space $\Lambda_+ L^2(\mathbb{R}^3;\mathbb{C}^4)$. For a comprehensive review we refer to the textbook of Balinsky and Evans [2]. While the physically relevant case is $\gamma > 0$, we are interested in the case where $\gamma = -\nu \in [-1, 0]$. For $\gamma < 3/4$ the operator B_γ was proved to be self-adjoint on $\Lambda_+ H^1(\mathbb{R}^3;\mathbb{C}^4)$ by Tix [32]. Since $\Lambda_+ H^1(\mathbb{R}^3;\mathbb{C}^4) \subset H^1(\mathbb{R}^3;\mathbb{C}^4)$ we obtain from Hardy's inequality

$$\left\| \Lambda_+ \frac{\gamma}{|x|} \Lambda_+ \psi \right\|_{L^2(\mathbb{R}^3;\mathbb{C}^4)} \leq 2\gamma \left\| \frac{1}{2|x|} \Lambda_+ \psi \right\|_{L^2(\mathbb{R}^3;\mathbb{C}^4)} \leq 2\gamma \|\nabla \Lambda_+ \psi\|_{L^2(\mathbb{R}^3;\mathbb{C}^4)}$$

for any $\psi \in H^1(\mathbb{R}^3;\mathbb{C}^4)$. To prove that $B_{-\nu}$ is essentially self-adjoint on $\Lambda_+ \mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$, it thus suffices to prove the statement for B_0. This is an immediate consequence of the fact that the free Dirac operator H_0 is essentially self-adjoint on $\mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$.

We can also conclude that the operator $\Lambda_-(\sqrt{1-\Delta} + \nu/|x|)\Lambda_-$ is essentially self-adjoint on $\Lambda_- \mathcal{C}_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$ since it is unitarily equivalent to the Brown–Ravenhall operator $B_{-\nu}$ via the transform $U : L^2(\mathbb{R}^3;\mathbb{C}^4) \to L^2(\mathbb{R}^3;\mathbb{C}^4)$

$$\left[U \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \right](x) = \begin{pmatrix} \psi_2(-x) \\ \psi_1(-x) \end{pmatrix}.$$

References

[1] M. F. Atiyah, V. K. Patodi and I. M. Singer, *Spectral asymmetry and Riemannian Geometry. I*, Math. Proc. Camb. Phil. Soc. (1975), **77**, 43

[2] A. A. Balinsky and W. D. Evans, *Spectral analysis of relativistic operators*, Imperial College Press, London, 2011

[3] M. Sh. Birman and M. Z. Solomjak, *Spectral theory of selfadjoint operators in Hilbert space.* Translated from the 1980 Russian original by S. Khrush and V. Peller. Mathematics and its Applications (Soviet Series). D. Reidel Publishing Co., Dordrecht, 1987.

[4] J. G. Brasche and H. Neidhardt, *Some remarks on Kreĭn's extension theory*, Math. Nachr. **165** (1994), 159–181.

[5] G. E. Brown and D. G. Ravenhall. O *On the interaction of two electrons*, Proc. Roy. Soc. London Ser. A., **208** (1951), 552–559.

[6] S. N. Datta and G. Deviah, *The minimax technique in relativistic Hartree-Fock calculations*, Pramana **30**(5) (1988), 387–405

[7] E. B. Davies *Spectral theory and differential operators*, Cambridge Studies in Advanced Mathematics, 42. Cambridge University Press, Cambridge, 1995.

[8] D. A. Deckert and M. Oelker, *Distinguished self-adjoint extension of the two-body Dirac operator with Coulomb interaction*, Ann. Henri Poincaré **20** (2019), no. 7, 2407–2445.

[9] J. Dolbeault, M. J. Esteban, M. Loss, and L. Vega, *An analytical proof of Hardy-like inequalities related to the Dirac operator*, J. Funct. Anal. **216** (2004), no. 1, 1–21.

[10] J. Dolbeault, M. J. Esteban and E. Séré, *On the Eigenvalues of Operators with Gaps. Application to Dirac Operators*, J. Funct. Anal. **174** (2000) 208–226.

[11] R. G. Douglas and K. P. Wojciechowski, *Adiabatic Limit of the η-Invariants. The Odd-Dimensional Atiyah-Patodi-Singer Problem*, Commun. Math. Phys. 142 (1991), 139–168.

[12] M. J. Esteban, M. Lewin and E. Séré, *Domains for Dirac-Coulomb min–max levels*, Rev. Mat. Iberoam **35** (2019), no./ 3, 877–924.

[13] M. J. Esteban and M. Loss, *Self-adjointness for Dirac operators via Hardy-Dirac inequalities*, J. Math. Phys. **48** (2007), no. 11, 112107.

[14] M. J. Esteban and M. Loss, *Self-adjointness via partial Hardy-like inequalities*, Mathematical results in quantum mechanics, World Sci. Publ., Hackensack, NJ, 2008, 41–47.

[15] L. C. Evans, *Partial Differential Equations*, Second Edition, Graduate Studies in Mathematics, Vol. **19**, American Mathematical Society, Providence, Rhode Island.

[16] K. Furutani, *Atiyah–Patodi–Singer boundary condition and a splitting formula of a spectral flow* J. Geom. Phys. **56** (2006), 310–321.

[17] K. Friedrichs, *Spektraltheorie halbbeschränkter Operatoren und Anwendung auf die Spektralzerlegung von Differentialoperatoren*, Math. Ann. **109** (1934), no. 1, 465–487.

[18] M. Gallone and A. Michelangeli, *Self-adjoint realisations of the Dirac-Coulomb Hamiltonian for heavy nuclei*, Anal. Math. Phys. **9** (2019), no. 1, 585–616.

[19] M. Gallone and A. Michelangeli, *Discrete spectra for critical Dirac-Coulomb Hamiltonians*. J. Math. Phys. **59** (2018), no. 6, 062108.

[20] M. Griesemer and H. Siedentop, *A minimax principle for the eigenvalues in spectral gaps*, J. London Math. Soc. (2) **60** (1999), no. 2, 490-500.

[21] M. Klaus and R. Wüst, *Characterization and uniqueness of distinguished selfadjoint extensions of Dirac operators*, Comm. Math. Phys. **64** (1978/79), no. 2, 171–176.

[22] M. Kraus, M. Langer and C. Tretter, *Variational principles and eigenvalue estimates for unbounded block operator matrices and applications*, J. Comput. Appl. Math. **171** (2004), no. 1-2, 311–334.

[23] M. G. Krein, *The Theory of Self-adjoint Extensions of Semibounded Hermitian Operators and its Applications*, I. Mat. Sbornik **20** (1947), no. 3, 431–495 (in Russian).

[24] D. Krejčířik and Z. Lu, *Location of the essential spectrum in curved quantum layers*, J. Math. Phys. **55** (2014), no. 8, 083520.

[25] S. Morozov and D. Müller, *On the minimax principle for Coulomb-Dirac operators*, Math. Z., **280** (2015), 733–747.

[26] D. Müller, *Minimax principles, Hardy-Dirac inequalities and operator cores for two and three dimensional Coulomb-Dirac operators*, Doc. Math. **21** (2016), 1151–1169.

[27] G. Nenciu, *Self-adjointness and invariance of the essential spectrum for Dirac operators defined as quadratic forms*, Comm. Math. Phys. **48** (1976), no. 3, 235–247.

[28] M. Reed and B. Simon, *Methods of modern mathematical physics. I. Functional analysis.* Academic Press, New York-London, 1972.

[29] M. Reed and B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness.* Academic Press, New York-London, 1975.

[30] U.-W. Schmincke, *Essential selfadjointness of Dirac operators with a strongly singular potential*, Math. Z. **126** (1972), 71–81.
[31] J. D. Talman, *Minimax principle for the Dirac equation*, Phys. Rev. Lett. **57** (1986), no. 9, 1091–1094.
[32] C. Tix, *Self-adjointness and spectral properties of a pseudo-relativistic Hamiltonian due to Brown and Ravenhall*, Preprint, mp-arc: 97-441, 1997.
[33] C. Tretter, *Spectral theory of block operator matrices and applications*, Imperial College Press, London, 2008.
[34] R. Wüst, *Distinguished self-adjoint extensions of Dirac operators constructed by means of cut-off potentials*, Math. Z. **141** (1975), 93–98.
[35] R. Wüst, *Dirac operations with strongly singular potentials. Distinguished self-adjoint extensions constructed with a spectral gap theorem and cut-off potentials*, Math. Z. **152** (1977), no. 3, 259–271.

QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
Email address: schimmer@math.ku.dk

QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
Email address: solovej@math.ku.dk

QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
Email address: sabiha.tokus@math.ku.dk

[1]Corresponding author

Part II

A variational approximation to the Lieb–Liniger model

Estimation of the ground state energy

3 | Preliminaries

3.1 Introduction

Exact solvability of a quantum mechanical many-body model is rather an exception than the rule. This is one of the reasons for the importance of approximation models in mathematical physics theory.

A large part of the mathematical analysis of Bose systems is based on Bogoliubov's famous approximation theory that dates back to the year 1947. In our project we are concerned with a variational formulation of Bogoliubov's approximation that was proposed by Napiórkowski, Reuvers and Solovej in [NRS18a] and [NRS18b]. We study this approximation scheme by application to the Lieb–Liniger model, a one-dimensional model for bosonic particles, solved by Lieb and Liniger in [LL63]. We derive the ground state energy for the limiting cases of weak and strong coupling and compare it to the exact solution.

In this introductory chapter we give a short overview, introducing the Lieb–Liniger model as well as the Bogoliubov variational principle. In Chapter 4 we derive our results for the ground state energy. The chapter is structured such that we first present some derivations for general coupling and subsequently deal with first the case of strong and then weak coupling.

Our main results are found in Theorem 3 and Theorem 4.

3.2 The Lieb–Liniger model

Lieb and Liniger [LL63] introduced in 1963 a model for a gas of Bose particles in one dimension interacting via a two-body delta-potential. The gas is modelled by the Schrödinger equation for N particles in a one-dimensional box of length L with a Hamiltonian given by

$$H = -\sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + 2c \sum_{1\leq i<j\leq N} \delta(x_i - x_j) \tag{3.1}$$

with coupling constant $c \geq 0$. This model has been solved exactly in the sense that all eigenfunctions can be calculated explicitly and corresponding energy eigenvalues are uniquely determined by a transcendental equation [LL63, Lie63]. This is a rare property for

a quantum many body problem to have, making it ideal for testing approximation schemes, as was also pointed out in [LL63].

The density of the gas is denoted by $\varrho = \frac{N}{L}$. Rescaling the system in length by a factor of ϱ, that is viewing wave functions as functions of a variable $z = \varrho x$, we find that the Hamiltonian transforms to $\widetilde{H}$ with

$$\widetilde{H}\varrho^{-2} = -\sum_{i=1}^{N} \frac{\partial^2}{\partial z_i^2} + 2\frac{c}{\varrho} \sum_{1 \le i < j \le N} \delta(z_i - z_j)\,.$$

The system is thus really dependent only on one parameter, the dimensionless quantity $\frac{c}{\varrho} =: \xi$.

Similar dimensional arguments show that the total energy of the system scales with $N\varrho^2$ and thus the ground state energy E_0 of the system is of the form $E_0 = N\varrho^2 e(\xi)$ where $e(\xi)$ is the dimensionless ground state energy per particle of the operator $\widetilde{H}\varrho^{-2}$, depending only on the parameter ξ (cf. also [LL63]).

For the thermodynamic limit, where $N, L \to \infty$ while $\varrho = \frac{N}{L}$ is fixed, Lieb and Liniger present an equation that determines $e(\xi)$ uniquely (equation (3.19) in [LL63]; the parameter γ is exactly ξ). Moreover, they show that $e(\xi)$ is an analytic function of ξ except at the point $\xi = 0$. The expression determining $e(\xi)$ does however only give an implicit formula for $e(\xi)$ which in the absence of an explicit solution has been studied numerically.

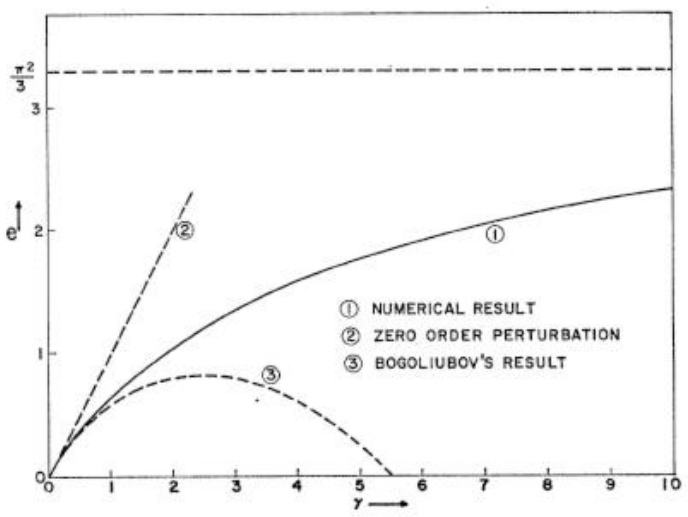


Figure 3.1: The ground state energy per particle $e(\xi)$ as a function of ξ which here is called γ. Curve 1 shows $e(\xi)$ as calculated numerically. Curve 2 gives the zero-order perturbation theory, $e(\xi) = \xi$ and curve 3 is $e(\xi)$ according to Bogoliubov's theory.[1]

Figure 3.1 illustrates the numerical solution of $e(\xi)$ in curve 2. It is compared to the zeroth order perturbation, $e(\xi) = \xi$ and an approximation by Bogoliubov's theory.

The dashed horizontal line marks the asymptotic limit of $e(\xi)$ at $\frac{\pi^2}{3}$ for $\xi \to \infty$. This limit coincides with the ground state energy of a general one-dimensional system with a hard-core potential as computed by Girardeau in [Gir60]. For such systems of impenetrable bosons in one dimension, Girardeau computed the ground state energy and showed that the corresponding bosonic ground state is identical to that of a non-interacting fermionic system.[2] Clearly, the two-body potential of the Hamiltonian H of the Lieb–Liniger model becomes hard-core in the limit $\xi \to \infty$.

3.3 A Bogoliubov variational principle

In his work on the theory of superfluidity, Bogoliubov [Bog47] introduced in 1947 an approximation scheme for the low-excitation spectrum of a gas of interacting bosonic particles, which since then has found many applications in the study of many-body systems of Bosons.

The idea behind his method is the assumption that at zero temperature and for small coupling constants most of the particles in the gas will occupy the lowest one-particle energy state, thereby forming a condensate with macroscopic occupation number $N_0 \gg 1$. This may be used to approximate the theory: for weak interactions Bogoliubov argues that one may apply an approximative Hamiltonian only containing quadratic terms in the bosonic field operators to describe the system.

Bogoliubov's approximation scheme is accomplished in two steps. We recall that a generic translation-invariant Hamiltonian of a non-relativistic gas on a d-dimensional torus with two-point interaction may in second quantisation be written as

$$H = \sum_p (\varepsilon(p) - \mu) a_p^* a_p + \frac{1}{2L^d} \sum_{k,p,q} v(p) a_{k+p}^* a_{q-p}^* a_k a_q \tag{3.2}$$

with $\varepsilon(p) = \frac{\hbar^2}{2m} p^2$ the free one-particle energy eigenvalues and $v(p)$ being a two-body potential function. We chose the grand canonical version including the chemical potential μ. Furthermore, for brevity we write $a_p^{(*)} = a^{(*)}(\phi_p)$, where $\{\phi_p\}_p$ is the orthonormal basis $\phi_p = \frac{e^{ipx}}{\sqrt{L^d}}$ of momentum eigenstates in the one-particle Hilbert space $\mathcal{H}$, which has a discrete index set as long as the system is of finite size. As usual, a and a^* are the operators mapping elements from $\mathcal{H}$ to the algebra of bosonic annihilation and creation operators on the Fock space $\mathcal{F}(\mathcal{H})$, which satisfy the canonical commutation relations.

We now introduce the approximations Bogoliubov suggested. The first observation he makes is that due to the large number N_0 of particles in the ground state, here the

[1] Figure 3.1 was published by user Elieb at `https://en.wikipedia.org/w/index.php?curid=24952249` under a CC BY-SA 3.0 license. It is also found in [LL63].

[2] In fact, Girardeau showed this not only for the ground state but for the whole spectrum.

$p = 0$ state, we have $N_0 + 1 \approx N_0$. Hence it is reasonable to neglect the commutator $a_0^* a_0 - a_0 a_0^* = 1$ and replace the operators a_0 and a_0^* by a number, setting $a_0 = a_0^* = \sqrt{N_0}$.

Secondly, since we assume a weak interaction with only few excitations, it is valid to ignore the contribution of terms in the Hamiltonians that are of power larger than two in the operators a_k or a_k^* with $k \neq 0$. What one is left with after some calculations is a Hamiltonian that is quadratic in the $a_k^{(*)}$ operators for $k \neq 0$. We call this approximated Hamiltonian the *Bogoliubov Hamiltonian* H_{Bog}.

In [NRS18a] and [NRS18b], Napiórkowski, Reuvers and Solovej introduced and analysed a variational reformulation of Bogoliubov's approximation theory in three dimensions. The idea behind this approach is to exploit the fact that when determining the ground state of a Bogoliubov Hamiltonian variationally it is sufficient to restrict the variation over all states to a variation over a certain class of Bogoliubov variational states only. This means that if E_0^{Bog} is the ground state of a Bogoliubov Hamiltonain H_{Bog} then we have

$$\begin{aligned} E_0^{\text{Bog}} &:= \inf \left\{ \frac{\langle \psi, H_{\text{Bog}} \psi \rangle}{\|\psi\|^2} \,\middle|\, \psi \in \mathcal{F}(\mathcal{H}) \right\} \\ &= \inf \left\{ \frac{\langle \psi, H_{\text{Bog}} \psi \rangle}{\|\psi\|^2} \,\middle|\, \psi \text{ is a Bogoliubov variational state in } \mathcal{F}(\mathcal{H}) \right\}. \end{aligned}$$

The approach the authors in [NRS18a] and [NRS18b] choose is to formulate a variational principle for the ground state energy of a Bose gas by leaving the Hamiltonian of the system unapproximated but vary over the Bogoliubov variational states only. The ground state energy may then be calculated by minimising an energy functional over the set of Bogoliubov variational states. We call this approximation theory the *Bogoliubov variational principle*.

Before we formulate this variational principle we first give a basic description of the class of Bogoliubov states. It is generally known that the ground state of a purely quadratic Hamiltonian is a *quasi-free state* (see e.g. Theorem 11.5 in [Sol14]), where by a quasi-free state we mean a state in the bosonic Fock state for which Wick's theorem holds. This means that we can find the ground state of a quadratic Hamiltonian by a variation over quasi-free states. Quasi-free states are completely characterised by their corresponding *generalised one-particle density matrix*. More specifically, if $\psi \in \mathcal{F}(\mathcal{H})$ is a quasi-free state in the bosonic Fock space over the one-particle Hilbert space $\mathcal{H}$ then we call the linear operator $\Gamma_\psi : \mathcal{H} \bigoplus \mathcal{H}' \to \mathcal{H} \bigoplus \mathcal{H}'$, with $\mathcal{H}'$ denoting the dual space of $\mathcal{H}$,

$$\Gamma_\psi = \begin{pmatrix} \gamma_\psi & \alpha_\psi \\ \alpha_\psi^* & 1 + J\gamma_\psi J^* \end{pmatrix}$$

with $\gamma_\psi (\gamma_\psi + 1) \geq \alpha_\psi \alpha_\psi^*$ the generalised one-particle density matrix of ψ. Here $J : \mathcal{H} \to \mathcal{H}'$ is the operator defined by $[J(g)](h) = \langle g, h \rangle_{\mathcal{H}}$ for all $g, h \in \mathcal{H}$ and γ_ψ and α_ψ are the *one-particle density matrix* and the *pairing operator* corresponding to ψ, respectively. The

one-particle density matrix is the operator $\gamma_\psi : \mathcal{H} \to \mathcal{H}$ which is defined by

$$\langle h, \gamma_\psi\, g\rangle_{\mathcal{H}} = \langle \psi, a^*(g)a(h)\psi\rangle_{\mathcal{H}} \qquad \forall g, h \in \mathcal{H}\,.$$

Similarly, the pairing operator $\alpha_\psi : \mathcal{H}' \to \mathcal{H}$ is the operator for which

$$\langle h, \alpha_\psi Jg\rangle_{\mathcal{H}} = \langle \psi, a(g)a(h)\psi\rangle_{\mathcal{H}} \qquad \forall g, h \in \mathcal{H}\,.$$

It holds $\alpha_\psi^* = J\alpha J$. Recalling our definition of quasi-free states, it is clear that the generalised one-particle density matrix of a state Γ_ψ of a quasi-free state $\psi \in \mathcal{F}(\mathcal{H})$ completely specifies the state ψ. In fact, there is a one-to-one correspondence between generalised one-particle density matrices and quasi-free states (cf. e.g. [Nam20], Chapter 6.5).

Quasi-free states are suitable variational states for quadratic Hamiltonians and thus for the excitation part of the Bogoliubov Hamiltonian. However, they are not alone sufficient to specify the ground state of a gas described by a Bogoliubov Hamiltonian. One additionally needs to take into account the fraction of particles in the condensate at $k \neq 0$, and thereby Bogoliubov's c-number substitution.

This is done by exploiting the fact that for every $g \in \mathcal{H}$ one can find a unitary transformation $U_g : \mathcal{F}(\mathcal{H}) \to \mathcal{F}(\mathcal{H})$ such that

$$U_g^* a(f) U_g = a(f) + \langle f, g\rangle_{\mathcal{H}} \qquad \forall f \in \mathcal{H}\,.$$

It becomes clear how this relates to the particle condensate when we now choose $g \in \mathcal{H}$ to be the one-particle ground state wave function, $g = \sqrt{N}_0\phi_0$. Then for $a_p = a(\phi_p)$ we have

$$U_{N_0}^* a_p U_{N_0} = a_p + \sqrt{N_0}\delta_{p,0}\,.$$

This means that this transformation shifts a_0 by $\sqrt{N_0}$ but leaves all a_p for $p \neq 0$ fixed. The full class of Bogoliubov variational states is then defined to be the set

$$\left\{\psi = U_g\omega_{\gamma,\alpha} \in \mathcal{F}(\mathcal{H}) \;\middle|\; \omega_{\gamma,\alpha} \text{ is a quasi-free state},\ g = \sqrt{N_0}\phi_0\right\}\,.$$

We can now formulate the Bogoliubov variational principle of [NRS18a] for a translation-invariant system of bosonic particles in three dimensions defined on a cube with side-length L with periodic boundary conditions. The one-particle space $\mathcal{H}$ is the space of symmetric square-integrable wave-functions and we choose the plane wave-basis for $\mathcal{H}$ such that $a_p = a(L^{-3/2}e^{ipx})$ with momenta $p \in \frac{2\pi}{L}\mathbb{Z}^3$.

In this case, the one-particle ground state is the state with $p = 0$ and the correct condensate function is the constant function $g \equiv L^{-3/2}\sqrt{N_0}$, where by N_0 we denote the number of particles in the condensate. Furthermore, the operators γ and α specifying a

translation-invariant quasi-free state ψ are defined by the functions

$$\gamma(k) \coloneqq \langle \psi, a_k^* a_k \psi \rangle \tag{3.3}$$

and

$$\alpha(k) \coloneqq \langle \psi, a_k a_{-k} \psi \rangle \,. \tag{3.4}$$

Hence, a Bogoliubov variational state is determined by a triple $(\gamma, \alpha, \varrho_0)$ with functions γ and α as in (3.3) and (3.4) and $\varrho_0 \coloneqq \frac{N_0}{L^3}$, describing the momentum distribution of the particles, pairing in the system and the condensate density, respectively. By evaluating the expectation value of a Hamiltonian as in (3.2) in such a Bogoliubov state, we obtain in the limit $L \to \infty$ the energy per unit volume.

The free energy per volume in a translation-invariant quasi-free state for general temperature $T \geq 0$ in the limit $L \to \infty$ is then given by the *Bogoliubov free energy functional* which we cite from [NRS18a]

$$\begin{aligned} \mathcal{F}(\gamma, \alpha, \varrho_0) = &\frac{1}{(2\pi)^3} \int p^2 \gamma(p) \mathrm{d}p - \mu \varrho - TS(\gamma, \alpha) + \frac{\hat{V}(0)}{2} \varrho^2 \\ &+ \frac{1}{2(2\pi)^6} \int \int \hat{V}(p-q)(\alpha(p)\alpha(q) + \gamma(p)\gamma(q)) \mathrm{d}p \mathrm{d}q \\ &+ \varrho_0 \frac{1}{(2\pi)^3} \int \hat{V}(p)(\gamma(p) + \alpha(p)) \mathrm{d}p \,. \end{aligned} \tag{3.5}$$

Here, S is the entropy which can be explicitly expressed in terms of the functions γ and α, cf. [NRS18a, p. 3]. The function $\hat{V}$ in momentum space is the Fourier transform of a radial two-body potential $V(x)$. Without loss of generality it was restricted to real-valued functions γ and α. The functional $\mathcal{F}$ is defined on the domain

$$\mathcal{D} = \left\{ (\gamma, \alpha, \varrho_0) \;\middle|\; \gamma \in L^1((1+p^2)\mathrm{d}p), \gamma(p) \geq 0, \alpha(p)^2 \leq \gamma(p)(1+\gamma(p)), 0 \leq \varrho_0 \leq \varrho \right\} . \tag{3.6}$$

The total density ϱ of the system is given by $\varrho = \varrho_0 + \frac{1}{(2\pi)^3} \int \gamma(p) \mathrm{d}p$. For positive temperatures $T \geq 0$ the condition $\alpha^2(p) \leq \gamma(p)(1+\gamma(p))$ is strictly speaking not needed to be specified explicitly since it is in this case already ensured by the entropy-term.

3.4 Bogoliubov variational approach for the Lieb–Liniger model

In our project we are aiming at transferring the Bogoliubov variational approach from [NRS18a] to one dimension in general and the Lieb–Liniger gas in particular. More specifically, we calculate the ground state energy of the Lieb–Liniger model by the Bogoliubov variational principle and compare it to the exact solution of the model.

In order to achieve this, we need to minimise the one-dimensional equivalent of the functional $\mathcal{F}$ in (3.5) for the special case of V being a δ-potential, more specifically $V(x_i - x_j) = 2c\delta(x_i - x_j)$. In addition, we here consider a canonical ensemble at temperature $T = 0$, where the density ϱ is fixed and $\mu = 0$.

In this situation, the functional $\mathcal{F}$ simplifies in various ways. Without the entropy and chemical potential terms we are left with a functional where due to the special form of the potential, with $\hat{V} \equiv 2c$, the potential energy does not vary with the specific form of γ and α but only depends on the values of the integrals of γ and α, respectively. Additionally, since we keep the total density ϱ fixed, a state $(\gamma, \alpha, \varrho_0) \in \mathcal{D}$ is entirely specified by a tuple (γ, α) satisfying the conditions in $\mathcal{D}$, and we can always derive ϱ_0 from γ by the relation $\varrho_0 = \varrho - \frac{1}{2\pi}\int \gamma$. At this stage we keep ϱ_0 as a notational aid, although it might appear like an 'overspecification' of a state, and we may freely interchange ϱ_0 and $\varrho - \frac{1}{2\pi}\int \gamma$.

The Bogoliubov free energy functional (3.5) adapted to the one-dimensional Lieb–Liniger model at zero temperature and fixed density then becomes

$$\begin{aligned}\mathcal{F}(\gamma, \alpha, \varrho_0) = \frac{1}{2\pi}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + c\Bigg(\varrho^2 + \frac{1}{\pi}\varrho_0\int_{\mathbb{R}}\gamma(p)\mathrm{d}p + \frac{1}{4\pi^2}\left(\int_{\mathbb{R}}\gamma(p)\mathrm{d}p\right)^2 \\ + \frac{1}{\pi}\varrho_0\int_{\mathbb{R}}\alpha(p)\mathrm{d}p + \frac{1}{4\pi^2}\left(\int_{\mathbb{R}}\alpha(p)\mathrm{d}p\right)^2\Bigg)\,.\end{aligned} \tag{3.7}$$

According to the Bogoliubov variational principle, if the ground state energy per unit volume $E_0 L^{-1}$ exists, then it is given by the infimum of (3.7) over all states in $\mathcal{D}$ with $\varrho_0 = \varrho - \frac{1}{2\pi}\int\gamma$. That is, we vary (3.7) over all γ and α satisfying the conditions in $\mathcal{D}$.

After replacing all $\frac{1}{2\pi}\int\gamma$ by $\varrho - \varrho_0$, we obtain an expression for the ground state energy per particle given in dimensionless units,

$$\begin{aligned}e(\xi) = \frac{E_0}{N\varrho^2} = \inf_{\substack{(\gamma,\alpha,\varrho_0)\in\mathcal{D}:\\ \int\gamma = 2\pi(\varrho-\varrho_0)}} \frac{1}{\varrho^3}\mathcal{F}(\gamma,\alpha,\varrho_0) = \inf_{\substack{(\gamma,\alpha,\varrho_0)\in\mathcal{D}:\\ \int\gamma = 2\pi(\varrho-\varrho_0)}} \frac{1}{\varrho^3}\frac{1}{2\pi}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + \\ \xi\left(2 - \left(\frac{\varrho_0}{\varrho}\right)^2 + \frac{1}{\pi}\frac{\varrho_0}{\varrho^2}\int_{\mathbb{R}}\alpha(p)\mathrm{d}p + \frac{1}{4\pi^2\varrho^2}\left(\int_{\mathbb{R}}\alpha(p)\mathrm{d}p\right)^2\right)\,.\end{aligned} \tag{3.8}$$

This is the starting point for an estimation of the ground state energy.

4 | Estimation of the ground state energy

A calculation of the ground state energy of the Lieb–Liniger model by the Boguliubov variational approach means to compute the infimum in (3.8). Eventually we will see that an exact computation is not feasible in a straightforward way. We can, however, expand the ground state energy for large and small values of ξ.

The proof of these expansions is the aim of this chapter. In the first section, we start by substantially simplifying (3.8). This will be done in the first section of this chapter. Subsequently we find the expansions for large ξ and then small ξ.

4.1 Observations for general ξ

4.1.1 Minimisation in α

We start from (3.8) and calculate. The minimisation of (3.8) in α can be carried out directly as soon as we notice that the energy is only dependent on the integral of α and that the strictly convex function $x^2 + 2\varrho_0 x$ has a unique minimiser at $x = -\varrho_0$.

Hence, without any further restrictions, any α that has $\int \alpha = -2\pi\varrho_0$ would be a minimiser in (3.8). However, since α is subject to the condition $\alpha(p)^2 \leq \gamma(p)(\gamma(p)+1)$ $\forall p \in \mathbb{R}$, the lowest energy we can get is when finding an α that satisfies

$$\int \alpha = -\min\left\{2\pi\varrho_0, \int \sqrt{\gamma(\gamma+1)}\right\}. \tag{4.1}$$

In fact, there is always a $0 \leq r_0 < \infty$ such that the function defined by

$$\alpha_{r_0}(p) \coloneqq -\min\left\{2\pi r_0, \sqrt{\gamma(p)(\gamma(p)+1)}\right\}$$

has an integral satisfying (4.1) and satisfies the condition $\alpha^2(p) \leq \gamma(p)(\gamma(p)+1)$ by definition. To show that the integral of α_{r_0} satisfies (4.1), we notice that for $r = 0$, we have $\int \alpha_{r=0} = 0$, whereas $\int \alpha_r = -\int \min\left\{2\pi r, \sqrt{\gamma(\gamma+1)}\right\}$ is a continuous and monotone

function of r and tends to $-\int \sqrt{\gamma(\gamma+1)}$ for $r \to \infty$. Hence, by the intermediate value theorem, it exists a r_0 such that for $\alpha = \alpha_{r_0}$ equation (4.1) holds.

This means that the ground state energy per particle may be calculated as

$$\begin{aligned}\frac{E_0}{N\varrho^2} &= \frac{1}{\varrho^3} \inf_{\substack{(\gamma,\alpha,\varrho_0)\in\mathcal{D}:\\ \int\gamma=2\pi(\varrho-\varrho_0)}} \mathcal{F}(\gamma,\alpha,\varrho_0) = \frac{1}{\varrho^3} \inf_{\substack{(\gamma,\varrho_0):\\ (\gamma,\alpha_{r_0},\varrho_0)\in\mathcal{D},\\ \int\gamma=2\pi(\varrho-\varrho_0)}} \mathcal{F}(\gamma,\alpha_{r_0},\rho_0) \\ &= \inf_{\substack{(\gamma,\varrho_0):\\ (\gamma,\alpha_{r_0},\varrho_0)\in\mathcal{D},\\ \int\gamma=2\pi(\varrho-\varrho_0)}} \frac{1}{\varrho^3}\frac{1}{2\pi}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + \xi\left(2-\left(\frac{\varrho_0}{\varrho}\right)^2 - 2\frac{\varrho_0}{\varrho^2}\min\left\{\varrho_0, \frac{1}{2\pi}\int\sqrt{\gamma(\gamma+1)}\right\}\right. \\ &\qquad \left. + \frac{1}{\varrho^2}\left(\min\left\{\varrho_0, \frac{1}{4\pi^2}\int\sqrt{\gamma(\gamma+1)}\right\}\right)^2\right).\end{aligned} \tag{4.2}$$

Furthermore, introducing two abbreviations

$$\kappa \coloneqq \frac{\int\gamma}{\int\sqrt{\gamma(\gamma+1)}} \tag{4.3}$$

and

$$\delta \coloneqq \varrho - \varrho_0 \tag{4.4}$$

we may rewrite the expression in (4.2) as

$$\begin{aligned}\frac{E_0}{N\varrho^2} &= \inf_{(\gamma,\kappa,\delta)\in\mathcal{D}_0'} \mathcal{E}(\gamma,\kappa,\delta) \\ &\coloneqq \inf_{(\gamma,\kappa,\delta)\in\mathcal{D}_0'} \frac{1}{2\pi}\frac{1}{\varrho^3}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p \; + \xi\left(2-\left(\frac{\varrho-\delta}{\varrho}\right)^2 - 2\frac{\varrho-\delta}{\varrho^2}\min\left\{\varrho-\delta, \frac{\delta}{\kappa}\right\}\right. \\ &\qquad \left. + \frac{1}{\varrho^2}\left(\min\left\{\varrho-\delta, \frac{\delta}{\kappa}\right\}\right)^2\right),\end{aligned} \tag{4.5}$$

where we now minimise over (γ,κ,δ) in the domain

$$\begin{aligned}\mathcal{D}_0' = \{(\gamma,\kappa,\delta) \mid \gamma \in L^1((1+p^2)\mathrm{d}p), \gamma(p)\geq 0, \\ \int\gamma = 2\pi\delta, \int\sqrt{\gamma(\gamma+1)} = 2\pi\frac{\delta}{\kappa}, 0\leq\kappa<1, 0\leq\delta\leq\varrho\Big\}.\end{aligned}$$

That is, we minimise over all γ satisfying two additional integral constraints for a δ and a κ in the given range. We will occasionally refer to the conditions $\gamma \in L^1((1+p^2)dp)$ and $\gamma \geq 0$ as the integrability and positivity condition for γ, respectively.

4.1.2 Parametrisation of the minimising γ

Starting from (4.5), we can show that there exists a minimising γ and calculate it by the method of Lagrange multipliers. This allows us in the end to rewrite (4.5) as an infimum over two parameters ε and δ, where ε is the characterising parameter for the 'shape' of γ.

First of all, for fixed $0 \leq \delta \leq \varrho$ and $0 \leq \kappa < 1$ and any arbitrary paramaters μ and λ, it trivially holds that

$$\begin{aligned}
&\inf_{\substack{\gamma \in L^1((1+p^2)\mathrm{d}p),\gamma\geq 0:\\ \int\gamma = 2\pi\delta, \int\sqrt{\gamma(\gamma+1)} = 2\pi\delta/\kappa}} \int p^2\gamma \\
&= \inf_{\substack{\gamma \in L^1((1+p^2)\mathrm{d}p),\gamma\geq 0:\\ \int\gamma = 2\pi\delta, \int\sqrt{\gamma(\gamma+1)} = 2\pi\delta/\kappa}} \int p^2\gamma + \mu\int\gamma - \lambda\int\sqrt{\gamma(\gamma+1)} - 2\pi\mu\delta + 2\pi\lambda\frac{\delta}{\kappa}\,,
\end{aligned}$$

and consequently also

$$\begin{aligned}
&\inf_{\substack{\gamma \in L^1((1+p^2)\mathrm{d}p),\gamma\geq 0:\\ \int\gamma = 2\pi\delta, \int\sqrt{\gamma(\gamma+1)} = 2\pi\delta/\kappa}} \int p^2\gamma \\
&\geq \inf_{\gamma \in L^1((1+p^2)\mathrm{d}p),\gamma\geq 0} \int p^2\gamma + \mu\int\gamma - \lambda\int\sqrt{\gamma(\gamma+1)} - 2\pi\mu\delta + 2\pi\lambda\frac{\delta}{\kappa}\,.
\end{aligned} \tag{4.6}$$

Equality holds if there is a minimising $\gamma_{\min}$ for the right-hand side in (4.6) that additionally satisfies $\int\gamma = 2\pi\delta$ and $\int\sqrt{\gamma(\gamma+1)} = 2\pi\frac{\delta}{\kappa}$. We find this $\gamma_{\min}$, dependent on μ and λ, by solving the Euler-Lagrange equation for γ, which reads

$$p^2 + \mu - \lambda\frac{\gamma + \frac{1}{2}}{\sqrt{\gamma(\gamma+1)}} = 0\,. \tag{4.7}$$

For any given choice of parameters μ and λ, this equation uniquely determines $\gamma_{\min}$, while only certain choices μ and λ give a solution $\gamma_{\min}$ that also satisfies $\gamma_{\min} \in L^1((1+p^2)\mathrm{d}p$ and $\gamma_{\min} \geq 0$. Since the functional $\gamma \mapsto \int p^2\gamma + \mu\int\gamma - \lambda\int\sqrt{\gamma(\gamma+1)}$ is convex, it is also assured that the extremum derived from the Euler-Lagrange equation is indeed a minimum.

As will be seen, it is possible to pick a μ and a λ such that $\gamma_{\min}$ satisfies the positivity and integrability condition and we also have $\int\gamma_{\min} = 2\pi\delta$ and $\int\sqrt{\gamma(\gamma+1)} = 2\pi\frac{\delta}{\kappa}$. For such a choice of parameters, (4.6) is indeed an equality and $\gamma_{\min}$ is a minimiser for (4.5).

In a first step, we determine the possible range of μ and λ for γ to be non-negative and $\gamma \in L^1((1+p^2)\mathrm{d}p$. From (4.7) we derive that any λ of interest needs to be non-negative because $p^2 + \mu$ is necessarily positive for large values of p. Isolating γ in (4.7) gives

$$\gamma_{\min}(p) = \gamma_{\mu,\lambda}(p) \coloneqq \frac{1}{2}\left(\frac{p^2+\mu}{\sqrt{(p^2+\mu)^2 - \lambda^2}} - 1\right)\,, \tag{4.8}$$

where the non-negativity of λ assures that $\gamma_{\mu,\lambda}$ non-negative. Besides that we see that for $\lambda > 0$ we necessarily have $\mu \geq \lambda$ for a γ that is real-valued and positive. For a $\gamma_{\mu,\lambda}$ satisfying the integrability conditions in $\mathcal{D}_0'$ we finally need $\mu > \lambda$. The case $\lambda = 0$ corresponds to vanishing γ. It is not difficult to check that the condition $\mu > \lambda > 0$ or $\lambda = 0$ is also sufficient for the positivity and integrability condition on γ.

In a next step we would like to prove that for any given choice of $\delta \geq 0$ and $\kappa \geq 0$, there is a μ and a λ with $\mu > \lambda > 0$ or $\lambda = 0$ such that $\gamma_{\mu,\lambda}$ matches the positivity and integrability condition and it additionally holds that $\int \gamma_{\mu,\lambda} = 2\pi\delta$ and $\int \sqrt{\gamma_{\mu,\lambda}(\gamma_{\mu,\lambda}+1)} = \frac{2\pi\delta}{\kappa}$.

Indeed, we will be able to show that for any $\delta > 0$ and $\kappa > 0$ there exist such μ and λ, and in this case $\mu > \lambda > 0$. While the parameters μ and λ are by our arguments not necessarily uniquely determined by δ and κ. We still believe, although without being in the possession of any proof, that for any pair of κ and δ there is exactly one μ and exactly one λ.

The cases where either κ, δ or both are equal to 0 are slightly more involved but it is possible to show that they only occur as limiting cases of a strictly positive δ and κ. Therefore they may be left out in the determination of the infimum.

In the following, we will give the necessary arguments that support these claims. First, we set $\varepsilon = \frac{\lambda}{\mu}$ and then rewrite $\gamma_{\mu,\lambda}$ in terms of ε and μ. For $\mu > \lambda \geq 0$ the parameter ε is non-negative and strictly smaller than 1. We can express the relevant integrals as functions dependent on ε:

$$\int_{\mathbb{R}} \gamma_{\mu,\lambda}(p)\mathrm{d}p = \sqrt{\mu}\int_0^\infty \left(\frac{k^2+1}{\sqrt{(k^2+1)^2-\varepsilon^2}} - 1\right)\mathrm{d}k =: \sqrt{\mu} I_1(\varepsilon) \tag{4.9}$$

as well as

$$\int_{\mathbb{R}} \sqrt{\gamma_{\mu,\lambda}(p)(\gamma_{\mu,\lambda}(p)+1)}\mathrm{d}p = \sqrt{\mu}\int_0^\infty \frac{\varepsilon}{\sqrt{(k^2+1)^2-\varepsilon^2}}\mathrm{d}k =: \sqrt{\mu} I_2(\varepsilon) \tag{4.10}$$

and

$$\int_{\mathbb{R}} p^2\gamma_{\mu,\lambda}(p)\mathrm{d}p = \mu^{3/2}\int_0^\infty k^2\left(\frac{k^2+1}{\sqrt{(k^2+1)^2-\varepsilon^2}} - 1\right)\mathrm{d}k =: \mu^{3/2} I_3(\varepsilon)\,. \tag{4.11}$$

For $0 \leq \varepsilon < 1$ the integrals I_1, I_2 and I_3 all have finite values. They are also continuous functions of ε and they are 0 if and only if $\varepsilon = 0$. The ratio of $\int \gamma$ and $\int \sqrt{\gamma(\gamma+1)}$ only depends on ε. More precisely,

$$\kappa(\varepsilon) = \frac{I_1(\varepsilon)}{I_2(\varepsilon)}\,, \tag{4.12}$$

and κ is a continuous function of ε on the open interval $(0,1)$. From the properties of the

integrals I_1 and I_2, more specifically their expansions around 0, it is not difficult to derive that $\lim_{\varepsilon\to 0}\kappa = 0$, and clearly $\kappa < 1$ for all ε. Furthermore, it also holds that $\lim_{\varepsilon\to 1}\kappa = 1$. More detailed calculations which support these claims are also provided in the appendix.

By the intermediate value theorem we conclude that κ can be continuously extended to a surjective function that maps the interval $[0,1)$ onto $[0,1)$. Hence, for given $\delta > 0$ and $\kappa > 0$, there is an $\varepsilon > 0$, not necessarily unique, such that $\kappa(\varepsilon) = \kappa$ and the function $\gamma_{\mu,\lambda}$ with $\mu = \left(\frac{2\pi\delta}{I_1(\varepsilon)}\right)^2$ and $\lambda = \varepsilon\mu$ then satisfies $\int \gamma_{\mu,\lambda} = 2\pi\delta$ and $\int \sqrt{\gamma_{\mu,\lambda}(p)(\gamma_{\mu,\lambda}(p)+1)} = \frac{2\pi\delta}{\kappa}$.

The following proposition proves that the cases where $\delta = 0$ or $\kappa = 0$ do not need to be discussed separately.

Proposition 4. The variational problem (4.5) is equivalent to a variational problem where we replace $\mathcal{D}_0'$ by the domain $\mathcal{D}'$ with

$$\mathcal{D}' = \{(\gamma,\kappa,\delta) \mid \gamma \in L^1((1+p^2)\mathrm{d}p), \gamma(p) \geq 0, \\ \int \gamma = 2\pi\delta, \int \sqrt{\gamma(\gamma+1)} = \frac{2\pi\delta}{\kappa}, 0 < \kappa < 1, 0 < \delta \leq \varrho\Big\}.$$

Proof. First consider the case where $\delta = 0$ and $\kappa > 0$. Then we also have $\gamma \equiv 0$. We choose sequences $\{\delta_n\}$ with $\delta_n > 0$ and $\delta_n \to 0$ as well as $\{\kappa_n\}$ with $\kappa_n > 0$ and $\kappa_n \to \kappa$. By (4.12) we know there is a sequence $\{\varepsilon_n\}$ with $\varepsilon_n > 0$ such that $\kappa(\varepsilon_n) = \kappa_n$. The sequence $\{\varepsilon_n\}$ is not necessarily convergent if the relation $\kappa(\varepsilon)$ is not one-to-one. However, since $\kappa(\varepsilon)$ is continuous and furthermore $\kappa = 0$ if and only if $\varepsilon = 0$ we know that there is an $N \in \mathbb{N}$ such that $\varepsilon_n > c$ for all $n \geq N$ for some $c > 0$. Furthermore, we define a sequence $\{\gamma_n\}$, where $\gamma_n(p) = \gamma_{\mu_n,\lambda_n}(p)$ as given in (4.8) with $\mu_n = \left(\frac{2\pi\delta_n}{I_1(\varepsilon_n)}\right)^2$, $\mu_n > 0$ for all n, and $\lambda_n = \varepsilon_n\mu_n$. Both $\{\mu_n\}$ and $\{\lambda_n\}$ are convergent sequences with limit 0. Hence, $\{\gamma_n\}$ converges pointwise to $\gamma_{0,0} \equiv 0$ and we can see that the energy $\mathcal{E}(\gamma_n,\kappa_n,\delta_n)$ (cf. (4.5)) converges to $\mathcal{E}(0,0,\kappa) = \xi$.

The case $\delta = 0$ and $\kappa = 0$ is addressed in a very similar way. Here, we fix $\mu > 0$ and find a sequence of $\kappa_n > 0$ with $\kappa_n \to 0$, as well as a corresponding sequence $\{\varepsilon_n\}$, such that $\kappa(\varepsilon_n) = \kappa_n$. The ε_n's converge to 0 since $\kappa(\varepsilon) = 0$ if and only if $\varepsilon = 0$. Additionally, we define sequences $\{\delta_n\}$ and $\{\lambda_n\}$ by $\delta_n = \frac{1}{2\pi}\sqrt{\mu}I_1(\varepsilon_n)$ and $\lambda_n = \mu\varepsilon_n$. Both sequences converge to 0 and so does $\gamma_n = \gamma_{\mu,\lambda_n}$ pointwise. By the same reasoning as in the former case we then also find the limiting energy $\mathcal{E}(0,0,0) = \xi$.

If $\delta > 0$ but $\kappa = 0$, then, for any γ satisfying the conditions in $\mathcal{D}_0'$, we set $\gamma_n = \gamma\mathbb{1}_{\{p\leq n\}}$ and $\delta_n = \frac{1}{2\pi}\int\gamma_n$. By monotone (or dominated) convergence the δ_n's converge δ. Likewise, $\int\sqrt{\gamma_n(\gamma_n+1)}$ converges to $\int\sqrt{\gamma(\gamma+1)}$, and thus the sequence $\{\kappa_n\}$ with $\kappa_n \coloneqq 2\pi\delta_n\left(\int\sqrt{\gamma_n(\gamma_n+1)}\right)^{-1}$ converges to 0. The energy $\mathcal{E}(\gamma_n,\kappa_n,\delta_n)$ then converges to $\mathcal{E}(\gamma,0,\delta)$. □

We have thus been able to prove that there are numbers μ and λ, not necessarily unique,

such that the function $\gamma_{\mu,\lambda}$ is a minimiser for γ in (4.5), and we can finally conclude that

$$\begin{aligned}\frac{E_0}{N\varrho^2} &= \inf_{(\gamma,\kappa,\delta)\in\mathcal{D}'} \frac{1}{2\pi}\frac{1}{\varrho^3}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + \\ &\quad \xi\left(2-\left(\frac{\varrho-\delta}{\varrho}\right)^2 - 2\frac{\varrho-\delta}{\varrho^2}\min\left\{\varrho-\delta,\frac{\delta}{\kappa}\right\} + \frac{1}{\varrho^2}\left(\min\left\{\varrho-\delta,\frac{\delta}{\kappa}\right\}\right)^2\right) \\ &= \inf_{\substack{0<\varepsilon<1,\\ 0<\delta/\varrho\leq 1}} 4\pi^2\left(\frac{\delta}{\varrho}\right)^3\frac{I_3(\varepsilon)}{I_1(\varepsilon)^3} + \xi\Bigg(2-\left(1-\frac{\delta}{\varrho}\right)^2 - \\ &\quad 2\left(1-\frac{\delta}{\varrho}\right)\min\left\{1-\frac{\delta}{\varrho},\frac{1}{\kappa(\varepsilon)}\frac{\delta}{\varrho}\right\} + \left(\min\left\{1-\frac{\delta}{\varrho},\frac{1}{\kappa(\varepsilon)}\frac{\delta}{\varrho}\right\}\right)^2\Bigg) \\ &=: \inf_{\substack{0<\varepsilon<1,\\ 0<\delta/\varrho\leq 1}} E_k(\varepsilon,\delta/\varrho) + E_p(\varepsilon,\delta/\varrho) =: \inf_{\substack{0<\varepsilon<1,\\ 0<\delta/\varrho\leq 1}} E(\varepsilon,\delta/\varrho)\,,\end{aligned} \tag{4.13}$$

where κ is given by the relation (4.12). The notation E_k for the kinetic energy, comprising the first term in the above expression for the energy, and E_p for the potential energy, collecting all the remaining terms, was introduced here.

We make the following observation, which´ simplifies (4.13) additionally.

Proposition 5. For the variational problem stated in (4.13), it holds:

i) The infimum over ε is found for ε such that $\kappa(\varepsilon)$ lies in the sector $\kappa \geq \frac{\delta}{\varrho}\left(1-\frac{\delta}{\varrho}\right)^{-1}$.

ii) The infimum over δ is in fact a minimum and achieved for a $\delta > 0$.

For the ground state energy this means that (4.13) can be rewritten as follows:

$$\begin{aligned}\frac{E_0}{N\varrho^2} = \inf_{\substack{0<\delta/\rho\leq 1,\ 0<\varepsilon<1:\\ \kappa(\varepsilon)\geq\frac{\delta}{\varrho}\left(1-\frac{\delta}{\varrho}\right)^{-1}}} E(\varepsilon,\delta/\rho) &= \min_{0<\delta/\rho\leq 1}\ \inf_{\substack{0<\varepsilon<1:\\ \kappa(\varepsilon)\geq\frac{\delta}{\varrho}\left(1-\frac{\delta}{\varrho}\right)^{-1}}} 4\pi^2\left(\frac{\delta}{\varrho}\right)^3\frac{I_3(\varepsilon)}{I_1(\varepsilon)^3} \\ &\quad + \xi\left(1-2\frac{\delta}{\varrho}\left(\frac{1}{\kappa(\varepsilon)}-1\right)+\frac{\delta^2}{\varrho^2}\left(\frac{2}{\kappa(\varepsilon)}+\frac{1}{\kappa(\varepsilon)^2}-1\right)\right)\,.\end{aligned} \tag{4.14}$$

Proof. ad i): Clearly, we can exclude that the infimum is found for a $\delta > 0$ and $\varepsilon \to 0$. This is because $I_1(\varepsilon)^{-3}$ diverges as ε^{-6} whereas $I_3(\varepsilon)$ behaves like ε^2 for $\varepsilon \to 0$ as computed in the appendix. Thus, the kinetic energy diverges to infinity whereas the potential energy stays finite in this limiting case.

Assume now that the ground state energy is a minimum achieved by a $(\delta/\varrho)_{\min}$ and an $\varepsilon_{\min}$ such that $\kappa(\varepsilon_{\min}) < (\delta/\rho)_{\min}\left(1-(\delta/\rho)_{\min}\right)^{-1}$. Then, the kinetic energy in the ground state is

$$E_{\mathrm{k}}(\varepsilon_{\min},(\delta/\varrho)_{\min}) = 4\pi^2\left(\frac{\delta}{\varrho}\right)^3_{\min}\frac{I_3(\varepsilon_{\min})}{I_1^3(\varepsilon_{\min})}$$

and the potential energy is

$$E_{\mathrm{p}}(\varepsilon_{\min}, (\delta/\varrho)_{\min}) = 2\xi \left(\frac{\delta}{\varrho}\right)_{\min} \left(2 - \left(\frac{\delta}{\varrho}\right)_{\min}\right) .$$

However, by keeping $\varepsilon_{\min}$ fixed but choosing a $\delta/\rho < (\delta/\rho)_{\min}$ we can see that both the kinetic as well as the potential energy decrease.

ad ii): Assume the infimum was found for $(\gamma, \kappa, \delta) \in \mathcal{D}'$ with $\delta \to 0$, or equivalently for a $(\gamma, \kappa, \delta) \in \mathcal{D}'_0$ with $\delta = 0$. Then the total energy for this state is $\mathcal{E}(\gamma, \kappa, \delta) = \xi$ since we necessarily have $\gamma = 0$. However, it is easy to see that by choosing any other state with γ close to 0, we always can achieve a lower energy: let γ be any non-zero function in $L^1((1+p^2)\mathrm{d}p)$ for which $\sqrt{\gamma(\gamma+1)}$ is also integrable and $\sigma > 0$ be a positive real number. We now define $\gamma_\sigma = \sigma\gamma$. We can calculate the corresponding δ_σ and κ_σ directly from γ_σ, and find, for σ small enough,

$$\varrho - \delta_\sigma = \varrho - \frac{\sigma}{2\pi}\int \gamma > \frac{\sqrt{\sigma}}{2\pi}\int \sqrt{\gamma(\sigma\gamma+1)} = \frac{\delta_\sigma}{\kappa_\sigma} .$$

This allows us to compute the total energy according to (4.13) and we obtain

$$\begin{aligned}\mathcal{E}(\gamma_\sigma, \kappa_\sigma, \delta_\sigma) = \frac{1}{2\pi}\frac{1}{\varrho^3}\sigma\int p^2\gamma + \xi\Bigg(2 - \frac{1}{\varrho^2}\left(\varrho - \frac{\sigma}{2\pi}\int\gamma\right)^2 \\ - \frac{\sqrt{\sigma}}{\pi\varrho^2}\left(\varrho - \frac{\sigma}{2\pi}\int\gamma\right)\int\sqrt{\gamma(1+\sigma\gamma)} + \frac{1}{4\pi^2\varrho^2}\sigma\left(\int\sqrt{\gamma(1+\sigma\gamma)}\right)^2\Bigg) .\end{aligned}$$

If we let $\sigma \to 0$, we find

$$\mathcal{E}(\gamma_\sigma, \kappa_\sigma, \delta_\sigma) = \xi - \sqrt{\sigma}\frac{1}{\pi\varrho}\int\gamma + \mathcal{O}(\sigma)$$

and thus, for σ small enough, $\mathcal{E}(\gamma_\sigma, \kappa_\sigma, \delta_\sigma) < \xi$. □

The variational problem (4.14) is not trivial to solve, and some of the difficulty stems from the fact that we are lacking an explicit expression of ε in terms of κ since we are not able determine any explicit values for the integrals I_1 and I_2 for arbitrary ε.

It is, however, possible to consider the problem in some limiting cases. In the following we first address the situation where $\xi \to \infty$ and subsequently consider the behaviour of the ground state energy when $\xi \to 0$.

4.2 Ground state energy for large ξ

For large ξ, we find an expansion of the ground state energy E_0 as in the following theorem.

Theorem 3. In the limit of strong coupling or low density when $\xi \to \infty$, the ground state energy per particle for the Lieb–Liniger model calculated by the Bogoliubov variational principle and defined by (3.8) diverges and it holds

$$\frac{E_0}{N\varrho^2} = 4\xi^{1/2} + \mathcal{O}(1)\,.$$

We will start with deriving an a priori upper bound for the lowest energy. By means of this bound it will be possible to find the expansion of the infimum in (4.14).

4.2.1 Upper bound

Proposition 6. In the limit $\xi \to \infty$ the Lieb–Liniger ground state energy as defined in Theorem 3 can be estimated from above by:

$$\frac{E_0}{N\varrho^2} \leq 4\xi^{1/2} + \mathcal{O}(1)\,. \tag{4.15}$$

Proof. We derive the upper bound by constructing a suitable trial state $(\gamma, \kappa, \delta) \in \mathcal{D}'$ for the variational problem in (4.13). In this triple, we set

$$\delta(\xi) = \frac{1}{2}\varrho\xi^{-1/2}\,,$$

and

$$\kappa(\xi) = \frac{\delta}{\varrho}(\xi) = \frac{1}{2}\xi^{-1/2}\,.$$

Furthermore we choose $\gamma = \gamma_{\mu,\lambda}$ with $\mu = \left(\frac{2\pi\delta}{I_1(\varepsilon)}\right)^2$ and $\lambda = \mu\varepsilon$ for some ε satisfying the relation (4.12) for the given κ. While we cannot be sure this ε is unique, it is enough for us to know that there exists at least one. If there are several, we can arbitrarily choose one of them.

For sufficiently large ξ the value of κ will be close to zero and the same holds for the corresponding ε. We can find an expansion of κ around $\varepsilon = 0$ by using the relation (4.12). Both I_1 and I_2 may be expanded around $\varepsilon = 0$ and for $\varepsilon < 1$ we have (cf. the appendix)

$$\begin{aligned} I_1(\varepsilon) &= \frac{1}{2}\int \frac{1}{(k^2+1)^2}\varepsilon^2 + \mathcal{O}(\varepsilon^3) \\ I_2(\varepsilon) &= \int \frac{1}{k^2+1}\varepsilon + \mathcal{O}(\varepsilon^2)\,. \end{aligned}$$

This holds since both integrals are sufficiently regular as functions of ε, in particular not only continuous but also repeatedly continuously differentiable and the derivatives are bounded on any interval $[0, r]$ with $r < 1$. We can thus write

$$\kappa(\varepsilon) = \frac{1}{2} \frac{\int \frac{1}{(k^2+1)^2}}{\int \frac{1}{k^2+1}} \varepsilon + \mathcal{O}(\varepsilon^2) \,,$$

which also implies that κ has a non-zero derivative at $\varepsilon = 0$. By the inverse function theorem, we equally obtain an expansion of ε around $\kappa = 0$

$$\varepsilon(\kappa) = 2 \frac{\int \frac{1}{k^2+1}}{\int \frac{1}{(k^2+1)^2}} \kappa + \mathcal{O}(\kappa^2) \,. \tag{4.16}$$

This allows us now to calculate the total energy for the state $(\gamma_{\mu,\lambda}, \kappa(\xi), \delta(\xi))$ as an expansion in ξ for $\xi \to 0$. First, since $\frac{\delta}{\kappa} > \varrho - \delta$, the potential energy is

$$E_{\mathrm{p}}(\varepsilon, {}^{\delta}\!/_{\varrho}(\xi)) = 2\xi \frac{\delta}{\varrho}(\xi) \left(2 - \frac{\delta}{\varrho}(\xi)\right) = \xi^{1/2} \left(2 - \frac{1}{2}\xi^{-1/2}\right) = 2\xi^{1/2} - \frac{1}{2} \,.$$

For the kinetic energy we expand $I_3(\varepsilon)$ in a similar way as $I_1(\varepsilon)$ and $I_2(\varepsilon)$ to obtain

$$E_{\mathrm{k}}(\varepsilon, {}^{\delta}\!/_{\varrho}(\xi)) = 4\pi^2 \left(\frac{\delta}{\varrho}\right)^3 \frac{I_3(\varepsilon)}{I_1^3(\varepsilon)} = 4\pi^2 \left(\frac{\delta}{\varrho}\right)^3 \frac{1}{I_1^3(\varepsilon)} \left(\frac{1}{2} \int_0^\infty \frac{k^2}{(k^2+1)^2} \mathrm{d}k \; \varepsilon^2 + \mathcal{O}(\varepsilon^3)\right) . \tag{4.17}$$

Using (4.16) for rewriting both $I_3(\varepsilon)$ and $I_1(\varepsilon)$ as an expansion in κ, we find

$$I_1(\varepsilon(\kappa)) = 2 \frac{\left(\int \frac{1}{k^2+1}\right)^2}{\int \frac{1}{(k^2+1)^2}} \kappa^2 + \mathcal{O}(\kappa^3) \,, \tag{4.18}$$

and likewise

$$I_3(\varepsilon(\kappa)) = 2 \int_0^\infty \frac{k^2}{(k^2+1)^2} \mathrm{d}k \frac{\left(\int \frac{1}{k^2+1} \mathrm{d}k\right)^2}{\left(\int \frac{1}{(k^2+1)^2} \mathrm{d}k\right)^2} \kappa^2 + \mathcal{O}(\kappa^3) \,.$$

Inserting these expansions into (4.17) we obtain

$$\begin{aligned} E_{\mathrm{k}}(\varepsilon, \delta/\varrho(\xi)) &= 4\pi^2 \left(\frac{\delta}{\varrho}\right)^3 \left[2\int_0^\infty \frac{k^2}{(k^2+1)^2}\mathrm{d}k \frac{\left(\int \frac{1}{k^2+1}\mathrm{d}k\right)^2}{\left(\int \frac{1}{(k^2+1)^2}\mathrm{d}k\right)^2}\kappa^2 + \mathcal{O}(\kappa^3) \right] \\ &\qquad \times \left[2\frac{\left(\int \frac{1}{k^2+1}\right)^2}{\int \frac{1}{(k^2+1)^2}}\kappa^2 + \mathcal{O}(\kappa^3) \right]^{-3} \\ &= \pi^2 \left(\frac{\delta}{\varrho}\right)^3 \frac{1}{\kappa^4} \int_0^\infty \frac{k^2}{(k^2+1)^2}\mathrm{d}k \int_0^\infty \frac{1}{(k^2+1)^2}\mathrm{d}k \left(\int_0^\infty \frac{1}{k^2+1}\mathrm{d}k\right)^{-4} (1+\mathcal{O}(\kappa)) \\ &= \left(\frac{\delta}{\varrho}\right)^3 \frac{1}{\kappa^4}(1+\mathcal{O}(\kappa))\,. \end{aligned} \tag{4.19}$$

Finally, by our choices for $\frac{\delta}{\varrho}(\xi)$ and $\kappa(\xi)$, we find

$$E_{\mathrm{k}}(\varepsilon, \delta/\varrho(\xi)) = 2\xi^{1/2} + \mathcal{O}(1)\,.$$

This gives the claimed result.

□

4.2.2 Proof of Theorem 3

The upper bound from Proposition 6 can be used to give an estimate on δ and κ at the ground state in the limiting case of large ξ. In a first step, we will show that there is a minimising δ/ϱ for the variational problem in (4.13) and prove that it is $\mathcal{O}(\xi^{-1/2})$ for $\xi \to \infty$. Secondly, we derive an equivalent statement for κ. This information can be used to expand the energy in ξ and compute its minimum to leading order.

From (4.13), we see that for any fixed δ/ρ and ε the potential energy $E_{\mathrm{p}}(\varepsilon, \delta/\rho)$ is at least $2\xi\frac{\delta}{\varrho}\left(2-\frac{\delta}{\varrho}\right)$, the minimal value the potential energy can take for any given δ/ρ. This happens when ε is such that $\min\{\varrho-\delta, \frac{\delta}{\kappa(\varepsilon)}\} = \varrho - \delta$. Hence, in summary we have for $\delta/\rho = (\delta/\rho)_{\min}$,

$$2\xi\left(\frac{\delta}{\rho}\right)_{\min}\left(2-\left(\frac{\delta}{\rho}\right)_{\min}\right) \leq E_{\mathrm{pot}}(\varepsilon, (\delta/\rho)_{\min}) \leq \frac{E_0}{N\varrho^2}$$

for any ε. Using the upper bound (4.15) we may then estimate

$$\left(\frac{\delta}{\rho}\right)_{\min} \leq \xi^{-1/2} + \mathcal{O}(\xi^{-1})\,. \tag{4.20}$$

Proposition 5 allows us now to also estimate κ. We know that the minimal energy has to be

looked for in the sector $\kappa \geq \frac{\delta}{\varrho}\left(1-\frac{\delta}{\varrho}\right)^{-1}$. At the same time we may, by the upper bound (4.15), exclude the possibility that the minimal energy will be found for $\varepsilon \to 1$ for which also $\kappa \to 1$ since the potential energy alone would exceed the upper bound in this case. Therefore we know that there is an $\varepsilon_{\min}$ for which the total energy takes a minimum, and we choose the notation $\kappa_{\min} = \kappa(\varepsilon_{\min})$. Taking this a step further, we can also use the upper bound (4.15) and the fact that the kinetic energy is always positive to find

$$\begin{aligned}&\xi\left(2-\left(1-\left(\frac{\delta}{\rho}\right)_{\min}\right)^2-2\left(1-\left(\frac{\delta}{\rho}\right)_{\min}\right)\left(\frac{\delta}{\rho}\right)_{\min}\frac{1}{\kappa_{\min}}+\left(\frac{\delta}{\rho}\right)^2_{\min}\frac{1}{\kappa^2_{\min}}\right)\\&\leq 2\xi^{1/2}+\mathcal{O}(1)\,.\end{aligned}$$

This allows us to conclude, omitting some positive terms on the left hand side,

$$\xi\left(1-2\left(\frac{\delta}{\varrho}\right)_{\min}\frac{1}{\kappa_{\min}}\right)\leq 2\xi^{1/2}+\mathcal{O}(1)\,,$$

and by using (4.20), we finally find

$$\kappa_{\min}\leq 2\xi^{-1/2}+\mathcal{O}(\xi^{-1})\,. \tag{4.21}$$

Hence, for $\xi \to \infty$ we are sure that the minimising κ tends to 0 and we can therefore expand the kinetic energy as previously in (4.19),

$$E_{\mathrm{k}}(\varepsilon,\delta/\rho)=\left(\frac{\delta}{\varrho}\right)^3\frac{1}{\kappa^4}\left(1+\mathcal{O}(\xi^{-1/2})\right) \tag{4.22}$$

and additionally for the potential energy, derived from (4.14), taking (4.20) and (4.21) into account,

$$\begin{aligned}E_{\mathrm{p}}(\varepsilon,\delta/\rho)&=\xi\left(\left(1-\frac{\delta}{\varrho}\frac{1}{\kappa}\right)^2+2\frac{\delta}{\varrho}\left(1+\frac{\delta}{\varrho}\frac{1}{\kappa}\right)-\frac{\delta^2}{\varrho^2}\right)\\&=\xi\left(\left(1-\frac{\delta}{\rho}\frac{1}{\kappa}\right)^2+2\frac{\delta}{\varrho}\left(1+\frac{\delta}{\varrho}\frac{1}{\kappa}\right)+\mathcal{O}(\xi^{-1})\right)\,,\end{aligned} \tag{4.23}$$

where in both cases κ is always understood as $\kappa(\varepsilon)$. Combining both to the total energy $E(\varepsilon,\delta/\rho) = E_k(\varepsilon,\delta/\rho) + E_p(\varepsilon,\delta/\rho)$ we see that the kinetic energy is of smaller order than the leading order term in the potential energy. Hence, when minimising in κ to leading order, the kinetic energy does not influence the result and we immediately find

$$\kappa_{\min}=\left(\frac{\delta}{\varrho}\right)_{\min}+\mathcal{O}(\xi^{-1})\,.$$

Inserting this into (4.22) and (4.23), we find

$$E = \left(\frac{\varrho}{\delta} + 4\xi\frac{\delta}{\varrho}\right)\left(1 + \mathcal{O}(\xi^{-1/2})\right)$$

from which we can extract

$$\left(\frac{\delta}{\varrho}\right)_{\min} = \frac{1}{2}\xi^{-1/2} + \mathcal{O}(\xi^{-1})\,.$$

Finally, inserting $\delta_{\min}$ and $\kappa_{\min}$, we find the minimal energy to be

$$\frac{E_0}{N\varrho^2} = 4\xi^{1/2}(1 + \mathcal{O}(\xi^{-1/2}))\,.$$

This proves Theorem 3.

4.3 Ground state energy for small ξ

As in the case for large ξ, we find an expansion of the ground state energy in the Lieb–Liniger model for the limiting case of $\xi \to 0$:

Theorem 4. In the limit of weak coupling or high density when $\xi \to 0$, the ground state energy per particle of the Lieb–Liniger model according the Bogoliubov approximation scheme has an expansion

$$\frac{E_0}{N\varrho^2} = \xi - \frac{4}{3\pi}\xi^{3/2} + \frac{1}{32\pi^2}\xi^2(\ln\xi)^2 + \mathcal{O}(\xi^2 \ln\xi \ln(-\ln\xi))\,.$$

Similarly to the previous section, we first derive a priori bounds, a lower and an upper bound, which allow us to then compute the expansion of the energy as stated in Theorem 4.

4.3.1 Lower bound

The lower bound on the ground state energy we present here is valid for all ξ but mainly of use for the case of small ξ which we are considering in this section.

Proposition 7. For the ground state energy of the Lieb–Liniger model in the Bogoliubov approximation scheme we find the lower bound

$$\frac{E_0}{N\varrho^2} \geq \xi - \frac{4}{3\pi}\xi^{3/2}\,. \tag{4.24}$$

Proof. The ground state energy $\frac{E_0}{N\varrho^2}$ in the Bogoliubov approximation scheme is computed as an infimum of (3.7) over all states in $\mathcal{D}$, see (3.6). We find a lower bound on this energy in two steps. Firstly, we omit all terms quadratic in γ and α in the energy functional $\varrho^{-3}\mathcal{F}$ in (3.7). The resulting functional we denote by $\varrho^{-3}\tilde{\mathcal{F}}$. Secondly, we relax the integrability condition on γ in $\mathcal{D}$. This results in taking an infimum

$$\begin{aligned}&\inf_{(\gamma,\alpha,\varrho_0)\in\tilde{\mathcal{D}}} \frac{1}{\varrho^3}\tilde{\mathcal{F}}(\gamma,\alpha,\varrho_0)\\ &= \inf_{(\gamma,\alpha,\varrho_0)\in\tilde{\mathcal{D}}} \frac{1}{\varrho^3}\frac{1}{2\pi}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + \xi\left(1 + \frac{1}{\pi}\frac{\varrho_0}{\varrho^2}\left(\int_{\mathbb{R}}\gamma(p)\mathrm{d}p + \int_{\mathbb{R}}\alpha(p)\mathrm{d}p\right)\right)\,,\end{aligned}$$

over a set $\tilde{\mathcal{D}} = \{(\gamma,\alpha,\varrho_0) \mid \gamma(p) \geq 0, \alpha(p)^2 \leq \gamma(p)(1+\gamma(p)), 0 \leq \varrho_0 \leq \varrho\}$. The infimum of $\tilde{\mathcal{F}}$ over $\tilde{\mathcal{D}}$ is a lower bound on $\frac{E_0}{N\varrho^2}$. We compute it in the following.

In a first step, the functional $\varrho^{-3}\tilde{\mathcal{F}}$ can be minimised in α. The minimum is found for α

being the function for which $\alpha(p) = -(\gamma(p)(\gamma(p)+1))^{1/2}$ for all $p \in \mathbb{R}$, thus we have

$$\inf_{(\gamma,\alpha,\varrho_0)\in\tilde{\mathcal{D}}} \frac{1}{\varrho^3}\tilde{\mathcal{F}}(\gamma,\alpha,\varrho_0) = \inf_{\substack{(\gamma,\varrho_0):\\ \gamma(p)\geq 0,\ 0\leq\varrho_0\leq\varrho}} \frac{1}{\varrho^3}\frac{1}{2\pi}\int_{\mathbb{R}} p^2\gamma(p)\mathrm{d}p + \xi\left(1+\frac{1}{\pi}\frac{\varrho_0}{\rho^2}\left(\int_{\mathbb{R}}\gamma(p)\mathrm{d}p - \int_{\mathbb{R}}\sqrt{\gamma(p)(\gamma(p)+1)}\mathrm{d}p\right)\right).$$

Evaluating the Euler-Lagrange equations for γ reveals – by similar calculations as before – a minimising γ given by

$$\gamma(p) = \frac{1}{2}\left(\frac{p^2+2\varrho_0 c}{\sqrt{(p^2+2\varrho_0 c)^2 - 4\varrho_0^2 c^2}} - 1\right)$$

which is just $\gamma_{\mu,\lambda}$ as in (4.8) with $\mu = \lambda = 2\varrho_0 c$. Note that the function $\gamma_{2\varrho_0 c, 2\varrho_0 c}$ is still integrable when multiplied with a factor of p^2 and thus the kinetic energy corresponding to this function is finite. Taken alone, it is, however, not integrable, which is why the integrability condition on γ was relaxed. Inserting the solution $\gamma_{2\varrho c, 2\varrho_0 c}$, one finally finds

$$\inf_{(\gamma,\alpha,\varrho)\in\tilde{\mathcal{D}}} \frac{1}{\varrho^3}\tilde{\mathcal{F}}(\gamma,\alpha,\varrho_0) = \inf_{0\leq\varrho_0\leq\varrho} \xi - \frac{4}{3\pi}\left(\frac{\varrho_0}{\varrho}\right)^{3/2}\xi^{3/2},$$

where it was used that $I_3(1) = \frac{\sqrt{2}}{3}$ and $(I_2 - I_1)(1) = \sqrt{2}$. This is clearly minimised when $\varrho_0 = \varrho$ and we thus obtain

$$\inf_{(\gamma,\alpha,\varrho)\in\tilde{\mathcal{D}}} \frac{1}{\varrho^3}\tilde{\mathcal{F}}(\gamma,\alpha,\varrho_0) = \xi - \frac{4}{3\pi}\xi^{3/2}. \tag{4.25}$$

□

4.3.2 Upper bound

As in the case of small ξ, we also derive an a priori upper bound and use it later together with the lower bound from Proposition 7 to derive the minimal energy.

Proposition 8. In the limit $\xi \to 0$ the ground state energy of the Lieb–Liniger model as defined in Theorem 4 is bounded from above by

$$\frac{E_0}{N\varrho^2} \leq \xi - \frac{4}{3\pi}\xi^{3/2} + \frac{1}{32\pi^2}\xi^2(\ln\xi)^2 + \mathcal{O}(\xi^2\ln\xi\ln(-\ln\xi)). \tag{4.26}$$

Proof. We can find an upper bound in the case $\xi \to 0$ by choosing a suitable δ/ρ and ϵ

dependent on the value of ξ in the expression (4.14). We start therefore by setting

$$\frac{\delta}{\varrho}(\xi) = \frac{1}{2\pi}\sqrt{\frac{2}{3}}\xi^{1/2}\left(\frac{1}{\kappa(\varepsilon)} - 1\right)^{1/2}\left(\frac{I_1^3(\varepsilon)}{I_3(\varepsilon)}\right)^{1/2},$$

which is easily rewritten into

$$\frac{\delta}{\varrho}(\xi) = \frac{1}{2\pi}\sqrt{\frac{2}{3}}\xi^{1/2}\left(\frac{I_2(\varepsilon) - I_1(\varepsilon)}{I_3(\varepsilon)}\right)^{1/2} I_1(\varepsilon)\,.$$

This choice of δ/ρ leads to an expression the total energy $E_k(\varepsilon, \delta/\rho) + E_p(\varepsilon, \delta/\rho)$ which depends on ε and ξ:

$$E_k(\varepsilon, \delta/\varrho(\xi)) + E_p(\varepsilon, \delta/\varrho(\xi)) = \frac{1}{\pi}\frac{\sqrt{2}}{3\sqrt{3}}\xi^{3/2}\frac{(I_2(\varepsilon) - I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}} + \xi\Bigg(1-$$
$$\frac{1}{\pi}\sqrt{\frac{2}{3}}\xi^{1/2}\frac{(I_2(\varepsilon) - I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}} + \frac{1}{(2\pi)^2}\frac{2}{3}\xi\left(\frac{I_2(\varepsilon) - I_1(\varepsilon)}{I_3(\varepsilon)}\right) I_1^2(\varepsilon)\left(\frac{1}{\kappa(\varepsilon)^2} + \frac{2}{\kappa(\varepsilon)} - 1\right)\Bigg).$$

The expression can be sorted according to powers of ξ:

$$\begin{aligned} E_k(\varepsilon, \delta/\rho(\xi)) + E_p(\varepsilon, \delta/\rho(\xi)) = \xi &- \frac{2}{3\pi}\sqrt{\frac{2}{3}}\frac{(I_2(\varepsilon) - I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}}\xi^{3/2} \\ &+ \frac{1}{6\pi^2}\left(\frac{1}{\kappa(\varepsilon)^2} + \frac{2}{\kappa(\varepsilon)} - 1\right)\left(\frac{I_2(\varepsilon) - I_1(\varepsilon)}{I_3(\varepsilon)}\right) I_1^2(\varepsilon)\xi^2\,, \end{aligned} \tag{4.27}$$

where it was assumed that the ε-dependent terms do not change the ordering for our choice of $\varepsilon(\xi)$. This will be confirmed by computations in the next paragraphs.

To find an upper bound for this ground state energy, we choose ϵ dependent of ξ to be $\varepsilon(\xi) = 1 - \nu(\xi)$ with

$$\nu(\xi) = -\frac{1}{4\pi}\frac{1}{(3 - 2\sqrt{2})}\xi^{1/2}\ln\xi\,. \tag{4.28}$$

For very small ξ this choice of $\nu(\xi)$ is close to 0 and in the limiting case we have $\lim_{\xi\to 0}\nu(\xi) = 0$. When inserting (4.28) into (4.27) we cannot give a precise expression of the energy as a function of ξ, but we may expand the energy for small values of ξ around 0.

To this end, we first expand the expression (4.27) for ε close to 1. The detailed computations leading to this expansion can be found in the appendix, Section A.2. We find that for

$\varepsilon = 1 - \nu$ and $\nu \to 0$

$$E_k(\varepsilon, \delta/\rho) + E_p(\varepsilon, \delta/\rho) = \xi - \frac{4}{3\pi}\xi^{3/2}\left(1 - \frac{3}{8}\nu(3-2\sqrt{2}) + \mathcal{O}(\nu^2 \ln \nu)\right) + \frac{1}{8\pi^2}\xi^2\left((\ln \nu)^2 + \mathcal{O}(1)\right) . \tag{4.29}$$

For our choice of $\nu(\xi)$ in (4.28), we then obtain for $\xi \to 0$

$$E_k(\varepsilon, \delta/\rho(\xi)) + E_p(\varepsilon, \delta/\rho(\xi)) = \xi - \frac{4}{3\pi}\xi^{3/2} + \frac{1}{32\pi^2}\xi^2(\ln \xi)^2 + \mathcal{O}(\xi^2 \ln \xi \ln(-\ln \xi)) ,$$

which is the claimed upper bound for the minimal energy for small ξ. □

4.3.3 Proof of Theorem 4

In the following we show that in the limit $\xi \to 0$, the minimal energy can be given as an expansion in ξ, coinciding with the upper bound (4.26). As a starting point, we find in the variational problem (4.14) the exact minimising δ/ρ as a function of ξ and ε and expand it for small ξ. In a second step the minimising ε can be found by solving its minimising equation to leading order in ξ.

From (4.14) we find that a minimising δ/ρ must necessarily satisfy

$$12\pi^2\left(\frac{\delta}{\varrho}\right)^2 \frac{I_3(\varepsilon)}{I_1^3(\varepsilon)} + 2\frac{\delta}{\varrho}\left(\frac{1}{\kappa^2} + \frac{2}{\kappa} - 1\right)\xi - 2\left(\frac{1}{\kappa} - 1\right)\xi = 0 .$$

Solving this for $\frac{\delta}{\varrho}$ we obtain

$$\left(\frac{\delta}{\rho}\right)_{\min} = -\frac{1}{12\pi^2}\xi\frac{I_1^3(\varepsilon)}{I_3(\varepsilon)}\left(\frac{1}{\kappa^2} + \frac{2}{\kappa} - 1\right) + \frac{1}{2\pi}\left[\frac{2}{3}\xi\frac{I_1^3(\varepsilon)}{I_3(\varepsilon)}\left(\frac{1}{\kappa} - 1\right) + \frac{1}{36\pi^2}\xi^2\frac{I_1^6(\varepsilon)}{I_3^2(\varepsilon)}\left(\frac{1}{\kappa^2} + \frac{2}{\kappa} - 1\right)^2\right]^{1/2} . \tag{4.30}$$

Since the energy is a strictly convex function of δ/ρ, we conclude that $(\delta/\rho)_{\min}$ is indeed a minimiser. We can also see that for finite ξ, letting ε tend to 1, we would obtain $(\delta/\rho)_{\min} = 0$. At the same time, from Proposition 5 we also conclude, that we will not find the energy minimised by $\varepsilon \to 0$ unless also $(\delta/\rho)_{\min} = 0$. Thus in both cases, for $\varepsilon \to 0$ or $\varepsilon \to 1$, this would put the total energy to the value of ξ, which was excluded to be the ground state energy by general arguments, cf. Proposition 5.2, and also contradicts the upper bound (4.26) for a finite but small enough ξ. We conclude that for $\xi > 0$ but small enough, if the infimum (4.14) exists, it is really a minimum with an $0 < \varepsilon_{\min} < 1$.

The logical next step would be to calculate the energy for $\delta/\rho = (\delta/\rho)_{\min}$, leaving us with an expression of the energy only dependent on ε, which we subsequently should

minimise for ε. However, this procedure cannot be carried out in its most simple form since the energy will depend on ε indirectly via the integrals I_1, I_2 and I_3 which does not allow us to solve for a minimising $\varepsilon_{\min}$ immediately. Instead, we first need to draw some conclusions about the properties of $\varepsilon_{\min}$. This will give us the possibility to expand the relevant quantities in powers of ξ for small ξ.

We are able to prove the following lemma:

Lemma 2. Let $\varepsilon_{\min}(\xi)$ be the minimising ε in (4.14) dependent on ξ. When $\xi \to 0$ we find that $\varepsilon_{\min}(\xi) \to 1$ and

$$|\ln(1-\varepsilon_{\min}(\xi))| = \mathcal{O}(|\ln \xi|)\,.$$

Proof. The proof of this lemma relies heavily on repeatedly using knowledge about the structure of the minimisers of the functional $\widetilde{\mathcal{F}}$ which was introduced earlier in the proof of the lower bound on the ground state energy (4.24).

First of all, we note that

$$\inf_{(\gamma,\alpha,\varrho_0)\in\widetilde{\mathcal{D}}} \frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma,\alpha,\varrho_0) = \inf_{\substack{\mu,\lambda:\ \mu\geq\lambda\geq 0,\\ \alpha:\alpha^2\leq\gamma_{\mu,\lambda}(\gamma_{\mu,\lambda}+1),0\leq\varrho_0\leq\varrho}} \frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma_{\mu,\lambda},\alpha,\varrho_0)$$

since we have seen that $\widetilde{\mathcal{F}}$ has a unique minimiser of the form $\gamma_{\mu,\lambda}$ with $\mu=\lambda=2\varrho_0 c$. Using that the minimising α is $\widetilde{\alpha}_{\min} = -(\gamma_{\mu,\lambda}(\gamma_{\mu,\lambda}+1))^{1/2}$, and replacing $\lambda=\mu\varepsilon$, we thus find

$$\begin{aligned}
\inf_{(\gamma,\alpha,\varrho_0)\in\widetilde{\mathcal{D}}} \frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma,\alpha,\varrho_0) &= \inf_{\substack{\mu\geq 0, 0\leq\varepsilon\leq 1\\ 0\leq\varrho_0\leq\varrho}} \xi + \frac{1}{2\pi}\frac{1}{\varrho^3}\mu^{3/2} I_3(\varepsilon) - \frac{1}{\pi}\mu^{1/2}\frac{\varrho_0}{\varrho^2}\xi(I_2(\varepsilon)-I_1(\varepsilon)) \\
&= \inf_{\substack{0\leq\varepsilon\leq 1\\ 0\leq\varrho_0\leq\varrho}} \xi - \frac{2}{3\pi}\sqrt{\frac{2}{3}}\left(\frac{\varrho_0}{\varrho}\right)^{3/2}\xi^{3/2}\frac{(I_2(\varepsilon)-I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}} \\
&= \inf_{0\leq\varepsilon\leq 1} \xi - \frac{2}{3\pi}\sqrt{\frac{2}{3}}\xi^{3/2}\frac{(I_2(\varepsilon)-I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}}
\end{aligned} \tag{4.31}$$

where the expression on the right-hand side was first minimised in μ and subsequently in ϱ_0, finding that $\tilde{\mu}_{\min} = \frac{2}{3}\varrho_0\varrho\xi I_3(\varepsilon)^{-1}(I_2(\varepsilon)-I_1(\varepsilon))$ and $\widetilde{(\varrho_0)}_{\min} = \varrho$. Using our knowledge that the expression (4.31) has a unique minimum with value $\xi - \frac{4}{3\pi}\xi^{3/2}$, cf. (4.25), we conclude that

$$\max_{0\leq\varepsilon\leq 1}\frac{(I_2(\varepsilon)-I_1(\varepsilon))^{3/2}}{I_3(\varepsilon)^{1/2}} = \left.\frac{(I_2(\varepsilon)-I_1(\varepsilon))^{3/2}}{I_3^{1/2}(\varepsilon)}\right|_{\varepsilon=1} = \frac{(\sqrt{2})^{3/2}}{(\sqrt{2}/3)^{1/2}} = \sqrt{6}$$

and that this maximum is unique.

We now derive properties of the minimising parameters for $\mathcal{F}$. We denote by $\mathrm{x}_{\min}$, where x could be any of the quantities or parameters $\mathcal{F}$ depends on, the value for x that minimises $\mathcal{F}$ and likewise for $\tilde{\mathrm{x}}$ and $\widetilde{\mathcal{F}}$, respectively. All of the $\mathrm{x}_{\min}$ depend on ξ. The following chain of inequalities holds

$$\begin{aligned}\mathcal{F}\left(\gamma_{\min}=\gamma_{\mu_{\min},\lambda_{\min}},\alpha_{\min},\varrho_{0,\min}=\varrho-\delta_{\min}\right)&\geq\widetilde{\mathcal{F}}\left(\gamma_{\min},\alpha_{\min},\varrho_{0,\min}\right)\\&\geq\widetilde{\mathcal{F}}\left(\gamma_{\widetilde{\mu}_{\min},\lambda_{\min}},\widetilde{\alpha}_{\min},\varrho_{0,\min}\right).\end{aligned}$$

Inserting $\widetilde{\alpha}_{\min}$ and $\widetilde{\mu}_{\min}$, we thus find, according to the upper bound in (4.26)

$$\begin{aligned}\frac{1}{\varrho^3}\widetilde{\mathcal{F}}\left(\gamma_{\widetilde{\mu}_{\min},\lambda_{\min}},\widetilde{\alpha}_{\min},\varrho_{0,\min}\right)&=\xi-\frac{2}{3\pi}\sqrt{\frac{2}{3}}\left(\frac{\varrho_{0,\min}}{\varrho}\right)^{3/2}\xi^{3/2}\frac{(I_2-I_1)^{3/2}}{I_3^{1/2}}(\varepsilon_{\min})\\&\leq\xi-\frac{4}{3\pi}\xi^{3/2}+\frac{1}{32\pi^2}\xi^2(\ln\xi)^2+\mathcal{O}(\xi^2\ln\xi\ln(-\ln\xi)).\end{aligned}\tag{4.32}$$

Assume now that there exists a $\sigma_\varrho>0$ and a $\sigma_\varepsilon>0$ such that for ξ arbitrarily small, we always have $\varrho-\varrho_{0,\min}(\xi)>\sigma_\varrho$ and $1-\varepsilon_{\min}(\xi)>\sigma_\varepsilon$. Since $\widetilde{\mathcal{F}}$ has a unique minimum equal to $\xi-\frac{4}{3\pi}\xi^{3/2}$ at $\varepsilon=1$ and $\varrho_0=\varrho$, this implies in particular

$$\frac{1}{\varrho^3}\widetilde{\mathcal{F}}\left(\gamma_{\widetilde{\mu}_{\min},\lambda_{\min}},\widetilde{\alpha}_{\min},\varrho_{0,\min}\right)>\xi-\frac{4}{3\pi}\xi^{3/2}.$$

This inequality contradicts (4.32) when ξ is very small. Hence, we conclude that for $\xi\to0$

$$\varepsilon_{\min}(\xi)\to1$$

and

$$\varrho_{0,\min}(\xi)\to\varrho$$

by which the first part the statement in Lemma 2 is proven. In a similar manner we can draw conclusions about $\mu_{\min}$. Again using the upper bound (4.26), we start this time with the following inequality

$$\begin{aligned}&\frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma_{\mu_{\min},\varepsilon_{\min}},\widetilde{\alpha}_{\min},\varrho_{0,\min})\\&\quad=\xi+\frac{1}{2\pi}\left(\frac{\sqrt{\mu_{\min}}}{\varrho}\right)^3I_3(\varepsilon_{\min})-\frac{1}{\pi}\frac{\sqrt{\mu_{\min}}}{\varrho}\frac{\varrho_{0,\min}}{\varrho}\xi(I_2(\varepsilon)-I_1(\varepsilon))\\&\quad\leq\xi-\frac{4}{3\pi}\xi^{3/2}+\frac{1}{32\pi^2}\xi^2(\ln\xi)^2\end{aligned}\tag{4.33}$$

which holds for all ξ. Again, we use our knowledge that $\widetilde{\mathcal{F}}$ is uniquely minimised in μ by $\widetilde{\mu}_{\min}=\frac{2}{3}\varrho_{0,\min}\varrho\xi I_3(\varepsilon_{\min})^{-1}(I_2(\varepsilon_{\min})-I_1(\varepsilon_{\min}))$, which converges to $\mu_{\min}=2\varrho^2\xi$

when $\xi \to 0$. We claim that indeed also

$$\lim_{\xi\to 0} \mu_{\min}^{1/2} \varrho^{-1} \xi^{-1/2} = \sqrt{2}\,. \tag{4.34}$$

Suppose this was not the case. Then similarly we would have

$$\frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma_{\mu_{\min},\lambda_{\min}}, \widetilde{\alpha}_{\min}, \varrho_{0,\min}) > \xi - \frac{4}{3\pi}\xi^{3/2}$$

for all ξ, which now contradicts (4.33) when $\xi \to 0$.

With this knowledge we are able to prove the second part of the Lemma. By the upper (4.26) and lower bound (4.24) for the ground state energy, we have

$$\xi - \frac{4}{3\pi}\xi^{3/2} \leq \frac{1}{\varrho^3}\mathcal{F}(\gamma_{\mu_{\min},\lambda_{\min}}, \alpha_{\min}, \varrho_{0,\min}) \leq \xi - \frac{4}{3\pi}\xi^{3/2} + \frac{1}{32\pi^2}\xi^2(\ln\xi)^2\,.$$

Rewriting

$$\begin{aligned}&\frac{1}{\varrho^3}\mathcal{F}(\gamma_{\mu_{\min},\lambda_{\min}}, \alpha_{\min}, \varrho_{0,\min}) = \\ &\frac{1}{\varrho^3}\widetilde{\mathcal{F}}(\gamma_{\mu_{\min},\lambda_{\min}}, \alpha_{\min}, \varrho_{0,\min}) + \frac{1}{\varrho^2}\xi\left(\left(\int \gamma_{\mu_{\min},\lambda_{\min}}\right)^2 + \left(\int \alpha_{\min}\right)^2\right)\end{aligned}$$

and using that $\widetilde{\mathcal{F}}$ has a unique minimum, corresponding to the lower bound (4.24), we can conclude

$$\xi\left(\left(\int \gamma_{\mu_{\min},\lambda_{\min}}\right)^2 + \left(\int \alpha_{\min}\right)^2\right) \leq \frac{1}{32\pi^2}\xi^2\varrho^2(\ln\xi)^2$$

and thus

$$I_1(\varepsilon_{\min}(\xi)) \leq \frac{1}{4\sqrt{2}\pi}\frac{\varrho\xi^{1/2}}{\sqrt{\mu_{\min}}}|ln\xi|\,.$$

Hence, for all $\xi \leq \xi_0$, with $\xi_0 > 0$ small enough, by applying (4.34), we find

$$I_1(\varepsilon_{\min}(\xi)) \leq \frac{1}{4\sqrt{2}\pi}|\ln\xi|\,, \tag{4.35}$$

and thus

$$|\ln(1-\varepsilon_{\min}(\xi))| < |\ln\xi|\,.$$

□

We can now sort and expand $(\delta/\rho)_{\min}$ (4.30) in powers of ξ for ξ close to 0. Here, we use the knowledge that $I_3(\varepsilon)$ and $I_2(\varepsilon) - I_1(\varepsilon)$ converge as functions of ε for $\varepsilon \to 1$, cf. the appendix. Therefore, the only divergent parts in $(\delta/\rho)_{\min}$ are terms involving a free-standing $I_1(\varepsilon)$ which according to Lemma 2, cf. (4.35), do not grow faster than logarithmically. Hence,

we have for $\xi \to 0$

$$\left(\frac{\delta}{\rho}\right)_{\min} = \frac{1}{2\pi}\sqrt{\frac{2}{3}}\left(\frac{I_2(\varepsilon)-I_1(\varepsilon)}{I_3(\varepsilon)}\right)^{1/2} I_1(\varepsilon)\xi^{1/2} - \frac{1}{12\pi^2}\xi\frac{I_1^3(\varepsilon)}{I_3(\varepsilon)}\left(\frac{1}{\kappa^2}+\frac{2}{\kappa}-1\right)+\mathcal{O}(\xi^2)\,.$$

Inserting this into the energy, we find

$$\begin{aligned} E_k(\delta_{\min},\varepsilon)+E_p(\delta_{\min},\varepsilon) = \xi - \frac{2}{3\pi}\sqrt{\frac{2}{3}}\frac{(I_2(\varepsilon)-I_1(\varepsilon))^{3/2}}{I_3^{1/2}}\xi^{3/2} \\ +\frac{1}{6\pi^2}\left(\frac{1}{\kappa^2}+\frac{2}{\kappa}-1\right)\frac{I_2(\varepsilon)-I_1(\varepsilon)}{I_3(\varepsilon)}I_1^2(\varepsilon)\xi^2+\mathcal{O}(\xi^{5/2})\,. \end{aligned}$$

As before, cf. (4.29), we expand this expression of the energy also for $\varepsilon \to 1$, writing $\varepsilon = 1-\nu$. Taking the derivative of this expansion and bearing in mind the statements of Lemma 2, we find the minisming equation for ν to leading order:

$$\frac{1}{2\pi}(3-2\sqrt{2})\xi^{3/2}+\frac{1}{4\pi^2}\xi^2\frac{1}{\nu}\ln\nu = 0$$

or equivalently

$$\ln\nu = -2\pi(3-2\sqrt{2})\xi^{-1/2}\nu\,. \tag{4.36}$$

This equation is difficult to solve exactly, but it can be solved approximately. We substitute ν by the variable $\eta := \nu\xi^{-1/2}$, and rewrite (4.36) by

$$\ln\eta + 2\pi(3-2\sqrt{2})\eta = -\frac{1}{2}\ln\xi\,. \tag{4.37}$$

As may directly be concluded from (4.36), η diverges for $\xi \to 0$. This means, $\ln\eta << \eta$ and hence we choose to ignore $\ln\eta$ in (4.37), which can then easily be solved and we find

$$\nu = -\frac{1}{4\pi(3-2\sqrt{2})}\xi^{1/2}\ln\xi\,.$$

This is the same expression for ν which was used to derive the upper bound. Hence, we obtain a minimal energy for $\xi \to 0$, coinciding with (4.26), which proves Theorem 4.

4.4 Conclusion and perspectives

In this project, we have analysed a one-parameter model of a Bose gas with delta-potential in one dimension as presented by Lieb and Liniger in [LL63] by a Bogoliubov variational principle. More precisely, the ground state energy of the Lieb–Liniger model was determined as an expansion in the relevant parameter ξ for the cases $\xi \to \infty$ and $\xi \to 0$ in Theorem 3 and Theorem 4, respectively.

We may compare our results to the literature. Lieb and Liniger were able to solve the model completely, however their solution for the function $e(\xi) = \frac{E_0}{N\varrho^2}$ is an implicit integral equation, . To our knowledge, there is until today no explicit expression for the function $e(\xi)$ available.

For $\xi \to \infty$, Lieb and Liniger find by numerical computations that $\lim_{\xi\to\infty} e(\xi) = \frac{\pi^2}{3}$, which coincides with Girardeau's results [Gir60] for general one-dimensional Bose gases with a hard-core potential. When applying the Bogoliubov variational principle we find a diverging ground state energy for $\xi \to \infty$ according to Theorem 3. This shows clearly that the Bogoliubov approximation scheme is not a suitable approach for large ξ. A closer look at Girardeau's solution explains why: when in the limit $\xi \to \infty$ the Lieb–Liniger gas becomes a gas of 'impenetrable bosons' the ground state is effectively given by a fermionic state. Since we determine the ground state energy by a variation over bosonic states, the Bogoliubov variational approach cannot match the fermionic solution.

For ξ close to 0, we find an expansion as given in Theorem 4. The first two terms coincide with the value one obtains by applying Bogoliubov's approximation method as has been explained in [LL63]. Since the 70's there had been conjectures (see [TW16] for a short review) that the first three terms in the expansion of the Lieb–Liniger ground state energy should be

$$e(\xi) = \xi - \frac{4}{3\pi}\xi^{3/2} + \left[\frac{1}{6} - \frac{1}{\pi^2}\right]\xi^2 + \mathcal{O}(\xi^2)\,. \tag{4.38}$$

This was verified by Tracy and Widom [TW16] in 2016 by what they call the Leppington–Levine method. Our expansion for small ξ agrees with equation (4.38) for the first terms up to order $\xi^{3/2}$. The order of the third term differs by a factor $(\ln(\xi))^2$.

Summing up, we find that the Bogoliubov variational approach is for large ξ not a good approximation for the Lieb–Liniger model due to the fermionic nature of the ground state in this case. The expansion for $\xi \to 0$ is correct in the first two terms but deviates in the third term by a logarithmic factor.

A question for further research could be to investigate models with more general potentials by an application of the Bogoliubov variational principle, in particular to find expansions for the ground state energy for a small coupling-density ratio.

A | Computations of integrals

In this appendix, properties of the integrals I_1, I_2, I_3 defined in (4.9) – (4.11) and correlated quantities, which are used throughout the previous chapters, are derived.

A.1 Properties of I_1, I_2 and I_3 for $\varepsilon \to 0$

Clearly, all integrals I_1, I_2 and I_3 are 0 when $\varepsilon = 0$. Their first derivatives can easily be calculated. Since all integrands are multiple times continuously differentiable, as well as bounded by their own value for $\varepsilon = \frac{1}{2}$ (or any other $\varepsilon < 1$), which furthermore is integrable, we find that all I_i, $i = 1, 2, 3$, are continuously differentiable. Furthermore, their derivatives can be bounded. More precisely, we have

- $$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I_1(\varepsilon) = \int_0^\infty \left(\frac{k^2+1}{\sqrt{(k^2+1)^2 - \varepsilon^2}} - 1 \right) \mathrm{d}k = \varepsilon \int_0^\infty \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{3/2}} \, \mathrm{d}k$$
  $$\leq \frac{1}{2} \int_0^\infty \frac{k^2+1}{((k^2+1)^2 - \frac{1}{4})^{3/2}} \, \mathrm{d}k \ ,$$
- $$\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2} I_1(\varepsilon) = \int_0^\infty \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{3/2}} + 3\varepsilon^2 \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{5/2}} \, \mathrm{d}k \ ,$$
- $$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I_2(\varepsilon) = \int_0^\infty \frac{\varepsilon}{\sqrt{(k^2+1)^2 - \varepsilon^2}} \mathrm{d}k$$
  $$= \int_0^\infty \frac{1}{\sqrt{(k^2+1)^2 - \varepsilon^2}} + \frac{\varepsilon^2}{((k^2+1)^2 - \varepsilon^2)^{3/2}} \, \mathrm{d}k \ ,$$
- $$\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2} I_2(\varepsilon) = \int_0^\infty \frac{\varepsilon}{\sqrt{(k^2+1)^2 - \varepsilon^2}} \, \mathrm{d}k$$
  $$= \int_0^\infty \frac{3\varepsilon}{((k^2+1)^2 - \varepsilon^2)^{3/2}} + \frac{3\varepsilon^2}{((k^2+1)^2 - \varepsilon^2)^{5/2}} \, \mathrm{d}k \ ,$$
- $$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I_3(\varepsilon) = \int_0^\infty k^2 \left(\frac{k^2+1}{\sqrt{(k^2+1)^2 - \varepsilon^2}} - 1 \right) \mathrm{d}k = \varepsilon \int_0^\infty k^2 \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{3/2}} \, \mathrm{d}k \ ,$$
- $$\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2} I_3(\varepsilon) = \int_0^\infty k^2 \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{3/2}} + 3\varepsilon^2 k^2 \frac{k^2+1}{((k^2+1)^2 - \varepsilon^2)^{5/2}} \, \mathrm{d}k \ .$$

Bounds on the derivatives are found equivalently as for $\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I_1(\varepsilon)$ and higher derivatives are computed analogously. By Taylor's theorem, we therefore find the following expansions for the integrals for $\varepsilon \to 0$:

- $I_1(\varepsilon) = \dfrac{1}{2}\displaystyle\int_0^\infty \frac{1}{(k^2+1)^2}\mathrm{d}k\, \varepsilon^2 + \mathcal{O}(\varepsilon^3)$,
- $I_2(\varepsilon) = \displaystyle\int_0^\infty \frac{1}{k^2+1}\mathrm{d}k\, \varepsilon + \mathcal{O}(\varepsilon^2)$,
- $I_3(\varepsilon) = \dfrac{1}{2}\displaystyle\int_0^\infty \frac{k^2}{(k^2+1)^2}\mathrm{d}k\, \varepsilon^2 + \mathcal{O}(\varepsilon^3)$.

From the above results, an expansion for κ can immediately be derived:

$$\kappa(\varepsilon) = \frac{I_1(\varepsilon)}{I_2(\varepsilon)} = \frac{1}{2}\frac{\int \frac{1}{(k^2+1)^2}}{\int \frac{1}{k^2+1}}\varepsilon + \mathcal{O}(\varepsilon^2)\,,$$

and hence $\kappa \to 0$ for $\varepsilon \to 0$.

A.2 Properties of I_1, I_2 and I_3 for $\varepsilon \to 1$

This section is relevant for a derivation of (4.29). We set $\varepsilon = 1 - \nu$.

A.2.1 The integral $I_1(1-\nu)$ for $\nu \to 0$

Lemma 3. For $\nu \to 0$, we have

$$I_1(1-\nu) = -\frac{1}{2\sqrt{2}}\ln\nu + \text{const} + \mathcal{O}(1)\,.$$

Proof. First, we have

$$\begin{aligned}
I_1(1-\nu) &= \int_0^\infty \left(\frac{k^2+1}{\sqrt{(k^2+1)^2-(1-\nu)^2}} - 1\right)\mathrm{d}k \\
&= \int_0^1 \left(\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\right)\mathrm{d}k + \int_0^1 \left(\frac{k^2}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - 1\right)\mathrm{d}k \\
&+ \int_1^\infty \left(\frac{k^2+1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - 1\right)\mathrm{d}k \\
&= \int_0^1 \left(\frac{1}{\sqrt{2k^2+\nu(2-\nu)}}\right)\mathrm{d}k \\
&+ \int_0^1 \left(\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - \frac{1}{\sqrt{2k^2+\nu(2-\nu)}}\right)\mathrm{d}k
\end{aligned}$$

$$+\int_0^1\left(\frac{k^2}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k+\int_1^2\left(\frac{k^2+1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k$$
$$+\int_2^\infty\left(\frac{k^2+1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k\,.$$

The last four integrals on the right hand side can all be estimated, such that for any $0\leq\nu\leq 1$:

$$\int_2^\infty\left(\frac{k^2+1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k=\int_2^\infty\frac{1+\frac{1}{k^2}-\sqrt{1+\frac{2}{k^2}+\frac{\nu(2-\nu)}{k^4}}}{\sqrt{1+\frac{2}{k^2}+\frac{\nu(2-\nu)}{k^4}}}\mathrm{d}k$$
$$\leq\int_2^\infty 1+\frac{1}{k^2}-\sqrt{1+\frac{2}{k^2}}\mathrm{d}k\leq C=\frac{C}{24}\int_2^\infty\frac{1}{k^4}\mathrm{d}k$$

for some constant $C>1$. Furthermore, we find

$$\int_1^2\left(\frac{k^2+1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k\leq\int_1^2\left(\frac{k^2+1}{\sqrt{k^4+2k^2}}-1\right)\mathrm{d}k$$
$$=\sqrt{3}(\sqrt{2}-1)-1+\frac{1}{\sqrt{2}}(\ln\Big(2+\sqrt{6}\Big)-\ln\Big(1+\sqrt{3}\Big))\,.$$

For the third last integral we estimate

$$\int_0^1\left(\frac{k^2}{\sqrt{k^4+2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k\leq\int_0^1\left(\frac{k^2}{\sqrt{k^4+2k^2}}-1\right)\mathrm{d}k$$
$$=\int_0^1\left(\frac{k}{\sqrt{k^2+2}}-1\right)\mathrm{d}k=\Big(\sqrt{k^2+2}-k\Big)\Big|_0^1=\sqrt{3}-\sqrt{2}-1\,.$$

It remains the second integral, for which we find

$$\left|\int_0^1\left(\frac{1}{\sqrt{2k^2+\nu(2-\nu)}}-\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\right)\mathrm{d}k\right|$$
$$=\int_0^1\left(\frac{1}{\sqrt{2k^2+\nu(2-\nu)}}-\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\right)\mathrm{d}k$$
$$=\int_0^1\frac{1}{\sqrt{2k^2+\nu(2-\nu)}\cdot\sqrt{k^4+2k^2+\nu(2-\nu)}}\times$$
$$\left(\sqrt{k^4+2k^2+\nu(2-\nu)}-\sqrt{2k^2+\nu(2-\nu)}\right)\mathrm{d}k$$
$$=\int_0^1\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\left(\sqrt{1+\frac{k^4}{2k^2+\nu(2-\nu)}}-1\right)\mathrm{d}k$$

$$\leq \frac{1}{2}\int_0^1 \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\left(\frac{k^4}{2k^2+\nu(2-\nu)}\right)\mathrm{d}k$$
$$\leq \frac{1}{4}\int_0^1 \frac{k}{\sqrt{k^2+2}} = \frac{1}{4}\left(\sqrt{3}-\sqrt{2}\right).$$

The only divergent term comes from the first integral:

$$\int_0^1 \left(\frac{1}{\sqrt{2k^2+\nu(2-\nu)}}\right)\mathrm{d}k = \int_0^{1/\sqrt{\nu}} \left(\frac{1}{\sqrt{2k^2+(2-\nu)}}\right)\mathrm{d}k$$
$$= \frac{1}{\sqrt{2}}\ln\left(k+\sqrt{k^2+1-\frac{\nu}{2}}\right)\Big|_0^{1/\sqrt{\nu}} = \frac{1}{\sqrt{2}}\left(\ln\left(\frac{1}{\sqrt{\nu}}+\sqrt{1+\frac{1}{\nu}-\frac{\nu}{2}}\right)-\ln\left(\sqrt{1-\frac{\nu}{2}}\right)\right)$$
$$= \frac{1}{2\sqrt{2}}\left(-\ln(\nu)+2\ln\left(1+\sqrt{1+\nu-\frac{\nu^2}{2}}\right)-\ln\left(1-\frac{\nu}{2}\right)\right).$$

Here, the last two terms converge to a constant when $\nu \to 0$ and leave us with the leading term $-\frac{1}{2\sqrt{2}}\ln\nu$, which was claimed. □

A.2.2 The integral $(I_2 - I_1)(1-\nu)$ for $\nu \to 0$

Lemma 4. For $\nu \to 0$, we have

$$(I_2-I_1)(1-\nu) = \sqrt{2}+\frac{1}{4\sqrt{2}}\nu\ln\nu+\frac{\nu}{2\sqrt{2}}\left(-\frac{7}{2}+2\sqrt{2}-\frac{5}{2}\ln 2\right)+\mathcal{O}(\nu).$$

Proof. By definition

$$(I_2-I_1)(1-\nu) = \int_0^\infty 1-\frac{\nu+k^2}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\mathrm{d}k.$$

For $\nu = 0$ the integral is easily calculated:

$$(I_2-I_1)(1) = \int_0^\infty 1-\frac{k}{\sqrt{k^2+2}}\mathrm{d}k = \sqrt{2}.$$

We find the next terms in the expansion therefore by computing

$$(I_2-I_1)(1-\nu)-(I_2-I_1)(1) = \int_0^\infty k^2\left(\frac{1}{\sqrt{k^4+2k^2}}-\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\right)\mathrm{d}k$$
$$-\nu\int_0^\infty \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\mathrm{d}k.$$

We split this integral into two parts, from 0 to $\sqrt{\nu}$ and from $\sqrt{\nu}$ to ∞, denoting them by

$J_{21}^s(\nu)$ and $J_{21}^l(\nu)$, respectively. The former then becomes

$$\begin{aligned}
J_{21}^s(\nu) &= \int_0^{\sqrt{\nu}} k^2 \left(\frac{1}{\sqrt{k^4+2k^2}} - \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} \right) \mathrm{d}k \\
&\quad - \nu \int_0^{\sqrt{\nu}} \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\, \mathrm{d}k \\
&= \nu \int_0^1 k^2 \left(\frac{1}{\sqrt{\nu k^4+2k^2}} - \frac{1}{\sqrt{\nu k^4+2k^2+(2-\nu)}} \right) \mathrm{d}k \\
&\quad - \nu \int_0^1 \frac{1}{\sqrt{\nu k^4+2k^2+(2-\nu)}}\, \mathrm{d}k \\
&= \nu \int_0^1 k^2 \left(\frac{1}{\sqrt{2k^2}} - \frac{1}{\sqrt{2k^2+2}} \right) - \frac{1}{\sqrt{2k^2+2}}\, \mathrm{d}k + \mathcal{O}(\nu^2) \\
&= \frac{\nu}{\sqrt{2}} \int_0^1 \left(k - \sqrt{k^2+1} \right) \mathrm{d}k + \mathcal{O}(\nu^2) = \frac{\nu}{2\sqrt{2}} \left(\sqrt{2} - 1 - \ln\left(\sqrt{2}+1 \right) \right) + \mathcal{O}(\nu^2)\,.
\end{aligned}$$

For the other part,we have

$$\begin{aligned}
J_{21}^l(\nu) &= \int_{\sqrt{\nu}}^{\infty} k^2 \left(\frac{1}{\sqrt{k^4+2k^2}} - \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} \right) \mathrm{d}k \\
&\quad - \nu \int_{\sqrt{\nu}}^{\infty} \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\, \mathrm{d}k \\
&= \int_{\sqrt{\nu}}^{\infty} \frac{k^2}{\sqrt{k^4+2k^2}} \left(1 - \frac{1}{\sqrt{1+\frac{\nu(2-\nu)}{k^4+2k^2}}} \right) \mathrm{d}k - \nu \int_{\sqrt{\nu}}^{\infty} \frac{1}{\sqrt{k^4+2k^2}} \frac{1}{\sqrt{1+\frac{\nu(2-\nu)}{k^4+2k^2}}} \mathrm{d}k\,.
\end{aligned} \tag{A.1}$$

We can expand these terms by using the power series expansion for $(1+x)^{-1/2}$, which converges for $|x|<1$:

$$\begin{aligned}
\frac{1}{\sqrt{1+x}} &= 1 - \frac{1}{2}x + \frac{\frac{1}{2}\left(\frac{1}{2}+1\right)}{2!}x^2 + \ldots + (-1)^n \frac{\frac{1}{2}\left(\frac{1}{2}+1\right)\ldots\left(\frac{1}{2}+n-1\right)}{n!}x^n \\
&= 1 - \frac{1}{2}x + \frac{1\cdot 3}{2\cdot 4}x^2 - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}x^3 + \ldots + (-1)^n \frac{1\cdot 3\cdot \ldots \cdot (2n-1)}{2\cdot 4\cdot 6\cdot \ldots \cdot 2n}x^n\,.
\end{aligned} \tag{A.2}$$

Hence, we may continue to compute (A.1),

$$\begin{aligned}
J_{21}^l(\nu) = \int_{\sqrt{\nu}}^{\infty} \frac{k^2}{\sqrt{k^4+2k^2}} \left(1 - 1 + \frac{\nu}{2}\frac{(2-\nu)}{k^4+2k^2} - \frac{3}{8}\left(\frac{\nu(2-\nu)}{k^4+2k^2} \right)^2 + \ldots \right) \mathrm{d}k \\
- \nu \int_{\sqrt{\nu}}^{\infty} \frac{1}{\sqrt{k^4+2k^2}} \left(1 - \frac{\nu}{2}\frac{(2-\nu)}{k^4+2k^2} + \frac{3}{8}\left(\frac{\nu(2-\nu)}{k^4+2k^2} \right)^2 + \ldots \right) \mathrm{d}k
\end{aligned}$$

$$
\begin{aligned}
&= \nu \int_{\sqrt{\nu}}^{\infty} \frac{k^2}{(k^4+2k^2)^{3/2}} - \frac{1}{\sqrt{k^4+2k^2}}\,\mathrm{d}k + \mathcal{R}_{21}(\nu) \\
&= -\nu \int_{\sqrt{\nu}}^{\infty} \frac{1}{k}\frac{k^2+1}{(k^2+2)^{3/2}}\mathrm{d}k + \mathcal{R}_{21}(\nu) \\
&= -\frac{\nu}{2\sqrt{2}}\left(-\ln\sqrt{\nu} + \ln\left(\sqrt{\nu+2}+\sqrt{2}\right) + \frac{\sqrt{2}}{\sqrt{\nu+2}}\right) + \mathcal{R}_{21}(\nu) \\
&= \frac{1}{4\sqrt{2}}\nu\ln\nu - \frac{\nu}{2\sqrt{2}}\left(1+\frac{3}{2}\ln 2\right) + \mathcal{O}(\nu^2) + \mathcal{R}_{21}(\nu)\,,
\end{aligned}
$$

where we expand the terms in powers of ν and where

$$
\begin{aligned}
\mathcal{R}_{21}(\nu) = &\int_{\sqrt{\nu}}^{\infty} k^2\left(\frac{1}{\sqrt{k^4+2k^2}} - \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\right)\mathrm{d}k \\
&- \nu\int_{\sqrt{\nu}}^{\infty} \frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}}\,\mathrm{d}k + \nu\int_{\sqrt{\nu}}^{\infty}\frac{1}{k}\frac{k^2+1}{(k^2+2)^{3/2}}\mathrm{d}k \\
&= \nu\int_1^{\infty}\left(\frac{k^2}{\sqrt{\nu k^4+2k^2}} - \frac{k^2+1}{\sqrt{\nu k^4+2k^2+(2-\nu)}}\right) + \frac{1}{k}\frac{\nu k^2+1}{(\nu k^2+2)^{3/2}}\,\mathrm{d}k \\
&=: \nu F_{21}(\nu)\,.
\end{aligned}
\tag{A.3}
$$

For $\nu = 0$, we find

$$
\begin{aligned}
F_{21}(0) &= \frac{1}{\sqrt{2}}\int_1^{\infty}\mathrm{d}k\left(k + \frac{1}{2k} - \sqrt{k^2+1}\right) \\
&= \frac{1}{2\sqrt{2}}\left[k^2 - \sqrt{k^2+1}k + \ln k - \ln\left(k+\sqrt{k^2+1}\right)\right]\Big|_1^{\infty} \\
&= \frac{1}{2\sqrt{2}}\left(\sqrt{2} - \frac{3}{2} + \ln\left(\frac{\sqrt{2}+1}{2}\right)\right)\,.
\end{aligned}
$$

For a general ν, by Taylor's theorem and the mean-value theorem, we find

$$
\frac{1}{\sqrt{1+x}} \le 1 - \frac{1}{2}x + \frac{3}{8}x^2 \qquad \text{for } |x| < 1, \tag{A.4}
$$

which may be used to estimate the integrand of (A.3) from below. The lower bound is then derived by

$$
\begin{aligned}
&k^2\left(\frac{1}{\sqrt{\nu k^4+2k^2}} - \frac{1}{\sqrt{\nu k^4+2k^2+(2-\nu)}}\right) \\
&- \frac{1}{\sqrt{\nu k^4+2k^2+(2-\nu)}} + \frac{1}{k}\frac{\nu k^2+1}{(\nu k^2+2)^{3/2}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(k^2 - k^2 \frac{1}{\sqrt{1 + \frac{2-\nu}{\nu k^4 + 2k^2}}} - \frac{1}{\sqrt{1 + \frac{2-\nu}{\nu k^4 + 2k^2}}} + \frac{1 + \nu k^2}{2 + \nu k^2} \right) \\
&\geq \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(k^2 - k^2 \left(1 - \frac{1}{2} \frac{2-\nu}{\nu k^4 + 2k^2} + \frac{3}{8} \frac{(2-\nu)^2}{(\nu k^4 + 2k^2)^2} \right) \right. \\
&\left. - \left(1 - \frac{1}{2} \frac{2-\nu}{\nu k^4 + 2k^2} + \frac{3}{8} \frac{(2-\nu)^2}{(\nu k^4 + 2k^2)^2} \right) + \frac{1 + \nu k^2}{2 + \nu k^2} \right) \\
&= \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(-\frac{\nu}{2} \frac{1}{\nu k^2 + 2} - \frac{3}{8} \left(k^2 + 1 \right) \frac{(2-\nu)^2}{(\nu k^4 + 2k^2)^2} + \frac{1}{2} \frac{2-\nu}{\nu k^4 + 2k^2} \right) \\
&\geq \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(-\frac{\nu}{2} \frac{1}{\nu k^2 + 2} - \frac{3}{8} \frac{1}{\nu k^2 + 2} \frac{(2-\nu)^2}{\nu k^4 + 2k^2} - \frac{3}{8} \frac{(2-\nu)^2}{(\nu k^4 + 2k^2)^2} \right) \\
&\geq -\frac{1}{2\sqrt{2}} \frac{1}{k^3} - \frac{3}{32\sqrt{2}} (2-\nu)^2 \left(\frac{1}{k^3} + \frac{1}{k^4} \right) \\
&\geq -\frac{1}{8\sqrt{2}} \left(\frac{7}{k^3} + \frac{3}{k^4} \right) ,
\end{aligned}
$$

where the last line represents an integrable function. At the same time we can also establish an upper bound, by analogously using

$$
\frac{1}{\sqrt{1+x}} \geq 1 - \frac{1}{2} x \, , \tag{A.5}
$$

by which we can bound the integrand of (A.3) such that

$$
\begin{aligned}
& k^2 \left(\frac{1}{\sqrt{\nu k^4 + 2k^2}} - \frac{1}{\sqrt{\nu k^4 + 2k^2 + (2-\nu)}} \right) \\
& - \frac{1}{\sqrt{\nu k^4 + 2k^2 + (2-\nu)}} + \frac{1}{k} \frac{\nu k^2 + 1}{(\nu k^2 + 2)^{3/2}} \\
& \leq \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(k^2 + k^2 \left(-1 + \frac{1}{2} \frac{2-\nu}{\nu k^4 + 2k^2} \right) - 1 + \frac{1}{2} \frac{2-\nu}{\nu k^4 + 2k^2} + \frac{\nu k^2 + 1}{\nu k^2 + 2} \right) \\
& = \frac{1}{\sqrt{\nu k^4 + 2k^2}} \left(\frac{1}{\nu k^4 + 2k^2} - \frac{\nu}{2} \frac{k^2 + 1}{\nu k^4 + 2k^2} \right) \\
& \leq \frac{1}{2\sqrt{2}} \left(\frac{2}{k^3} + \frac{1}{k^4} \right) ,
\end{aligned}
$$

which likewise is integrable.

Hence, by dominated convergence, we can conclude that $\mathcal{R}_{21}(\nu)$ is continuous as a function of ν and

$$
\mathcal{R}_{21}(\nu) = \nu F_{21}(0) + o(\nu) \, ,
$$

which concludes the proof of Lemma 4. □

A.2.3 The integral $I_3(1-\nu)$ for $\nu \to 0$

Lemma 5. For $\nu \to 0$, we have

$$I_3(1-\nu) = \frac{\sqrt{2}}{3} + \frac{1}{4\sqrt{2}}\nu \ln \nu + \frac{\nu}{2\sqrt{2}}(-\frac{1}{2} - \frac{5}{2}\ln 2) + \mathcal{O}(\nu).$$

Proof. From the definition of I_3, (4.11), we find

$$I_3(1) = \frac{\sqrt{2}}{3}$$

and

$$I_3(1-\nu) - I_3(1) = \int_0^\infty k^2(k^2+1)\left(\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - \frac{1}{\sqrt{k^4+2k^2}}\right)\mathrm{d}k\,.$$

For a computation of $I_3(1-\nu) - I_3(1)$, we split this integral into two parts for small and large k, $J_3^s(\nu)$ and $J_3^l(\nu)$, respectively. We begin with small k.

$$\begin{aligned}
J_3^s(\nu) &= \int_0^{\sqrt{\nu}} k^2(k^2+1)\left(\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - \frac{1}{\sqrt{k^4+2k^2}}\right)\mathrm{d}k \\
&= \nu\int_0^1 k^2(1+\nu k^2)\left(\frac{1}{\sqrt{\nu k^4+2k^2+2-\nu}} - \frac{1}{\sqrt{\nu k^4+2k^2}}\right)\mathrm{d}k \\
&= \frac{\nu}{\sqrt{2}}\int_0^1 k^2\left(\frac{1}{\sqrt{k^2+1}} - \frac{1}{k}\right)\mathrm{d}k + \mathcal{O}(\nu^2) \\
&= \frac{\nu}{2\sqrt{2}}\left(k\sqrt{k^2+1} - k^2 - \ln\Big(k+\sqrt{k^2+1}\Big)\right)\Big|_0^1 + \mathcal{O}(\nu^2) \\
&= \frac{\nu}{2\sqrt{2}}\left(\sqrt{2} - 1 - \ln\Big(1+\sqrt{2}\Big)\right) + \mathcal{O}(\nu^2)
\end{aligned}$$

by Taylor's theorem and dominated convergence. For large k we find

$$\begin{aligned}
J_3^l(\nu) &= \int_0^{\sqrt{\nu}} k^2(k^2+1)\left(\frac{1}{\sqrt{k^4+2k^2+\nu(2-\nu)}} - \frac{1}{\sqrt{k^4+2k^2}}\right)\mathrm{d}k \\
&= \int_{\sqrt{\nu}}^\infty \frac{k^2(k^2+1)}{\sqrt{k^4+2k^2}}\left(\frac{1}{\sqrt{1+\frac{\nu(2-\nu)}{k^4+2k^2}}} - 1\right)\mathrm{d}k\,.
\end{aligned}$$

By using the series expansion in (A.2), this can be written as

$$\begin{aligned}
J_3^l(\nu) &= \int_\nu^\infty \frac{k^2(k^2+1)}{\sqrt{k^4+2k^2}} \left(-\frac{1}{2}\left(\frac{\nu(2-\nu)}{k^4+2k^2}\right) + \frac{3}{8}\left(\frac{\nu(2-\nu)}{k^4+2k^2}\right)^2 + \dots \right) \mathrm{d}k \\
&= -\nu \int_{\sqrt{\nu}}^\infty \frac{1}{k} \frac{k^2+1}{(k^2+2)^{3/2}}\, \mathrm{d}k + \mathcal{R}_3(\nu) \\
&= \frac{\nu}{2}\left(\frac{1}{\sqrt{k^2+2}} + \frac{1}{\sqrt{2}}\left(\ln\left(\sqrt{2}\sqrt{k^2+1}+2\right) - \ln k \right) \right) \Bigg|_{\sqrt{\nu}}^\infty + \mathcal{R}_3(\nu) \\
&= \frac{1}{4\sqrt{2}} \nu \ln \nu - \frac{\nu}{2\sqrt{2}}(1+\frac{3}{2}\ln 2) + \mathcal{R}_3(\nu)
\end{aligned}$$

with

$$\begin{aligned}
\mathcal{R}_3(\nu) &= \int_{\sqrt{\nu}}^\infty \frac{k^2(k^2+1)}{\sqrt{k^4+2k^2}} \left(\frac{1}{\sqrt{1+\frac{\nu(2-\nu)}{k^4+2k^2}}} - 1 + \nu \frac{1}{k^4+2k^2} \right) \mathrm{d}k \\
&= \nu \int_1^\infty \frac{k^2(\nu k^2+1)}{\sqrt{\nu k^4+2k^2}} \left(\frac{1}{\sqrt{1+\frac{2-\nu}{\nu k^4+2k^2}}} - 1 + \frac{1}{\nu k^4+2k^2} \right) \mathrm{d}k \\
&=: \nu F_3(\nu)\,.
\end{aligned} \tag{A.6}$$

For $\nu = 0$, we find

$$\begin{aligned}
F_3(0) &= \frac{1}{\sqrt{2}} \int_1^\infty k \left(\frac{k}{k^2+1} - 1 + \frac{1}{2k^2} \right) \mathrm{d}k \\
&= \frac{1}{2\sqrt{2}} \left(-k^2 + \sqrt{k^2+1}k + \ln k - \ln\left(\sqrt{k^2+1}+k\right) \right) \Bigg|_1^\infty \\
&= \frac{1}{2\sqrt{2}} \left(\frac{3}{2} - \sqrt{2} + \ln\left(\frac{\sqrt{2}+1}{2}\right) \right)\,.
\end{aligned}$$

Again, we can prove by dominated convergence that $F_3(\nu)$ is a function continuous in ν. First, we see that the integrand in (A.6) is positive for any $0 \le \nu \le 1$:

$$\begin{aligned}
&\frac{k^2(\nu k^2+1)}{\sqrt{\nu k^4+2k^2}} \left(\frac{1}{\sqrt{1+\frac{2-\nu}{\nu k^4+2k^2}}} - 1 + \frac{1}{\nu k^4+2k^2} \right) \\
\ge &\frac{k^2(\nu k^2+1)}{\sqrt{\nu k^4+2k^2}} \left(1 - \frac{1}{2}\frac{2-\nu}{\nu k^4+2k^2} - 1 + \frac{1}{\nu k^4+2k^2} \right) \\
\ge &\frac{\nu}{2} \frac{k^2(1+\nu k^2}{(\nu k^4+2k^2)^{3/2}} \ge 0\,,
\end{aligned}$$

where once more (A.5) was used. At the same time, the integrand is bounded from above

by an integrable function for any $0 \leq \nu \leq 1$ as the following computation shows:

$$\begin{aligned}
&\frac{k^2(\nu k^2+1)}{\sqrt{\nu k^4+2k^2}}\left(\frac{1}{\sqrt{1+\frac{2-\nu}{\nu k^4+2k^2}}}-1+\frac{1}{\nu k^4+2k^2}\right)\\
&\leq \frac{k^2(\nu k^2+1)}{\sqrt{\nu k^4+2k^2}}\left(1-\frac{1}{2}\frac{2-\nu}{\nu k^4+2k^2}+\frac{3}{8}\frac{(2-\nu)^2}{(\nu k^4+2k^2)^2}-1+\frac{1}{\nu k^4+2k^2}\right)\\
&\leq \frac{k^2(1+\nu k^2)}{\nu k^4+2k^2}\left(\frac{\nu}{2}\frac{1}{(\nu k^4+2k^2)^{1/2}}+\frac{3}{8}\frac{(2-\nu)^2}{(\nu k^4+2k^2)^{3/2}}\right)\\
&\leq \frac{\nu^{1/2}}{2k^2}+\frac{3}{8}\frac{(2-\nu)^2}{(2k^2)^{3/2}}\\
&\leq \frac{1}{2k^2}+\frac{3}{16\sqrt{2}k^3}\,.
\end{aligned}$$

Hence, $F_3(\nu)$ is indeed continuous and we have

$$\mathcal{R}_3(\nu) = F_3(0) + \mathcal{O}(\nu)\,,$$

which finishes the proof of Lemma 5. □

A.2.4 Computation of the ratio $\frac{(I_2-I_1)^{3/2}}{I_3^{1/2}}$ for $\nu \to 0$

By combining Lemma 4 and Lemma 5, we find for $\nu \to 0$

$$\begin{aligned}
&\frac{(I_2-I_1)^{3/2}(1-\nu)}{I_3^{1/2}(1-\nu)}\\
&=\frac{(I_2-I_1)^{3/2}(1)}{I_3^{1/2}(1)}\left[1+\frac{3}{2}\frac{1}{(I_2-I_1)(1)}\left(\frac{1}{4\sqrt{2}}\nu\ln\nu+\frac{\nu}{2\sqrt{2}}\left(-\frac{7}{2}+2\sqrt{2}-\frac{5}{2}\ln 2\right)\right)\right.\\
&\left.+\mathcal{O}(\nu)\right]\times\left[1-\frac{1}{2}\frac{1}{I_3(1)}\left(\frac{1}{4\sqrt{2}}\nu\ln\nu+\frac{\nu}{2\sqrt{2}}\left(-\frac{1}{2}-\frac{5}{2}\ln 2\right)\right)+\mathcal{O}(\nu)\right]\\
&=\frac{(I_2-I_1)^{3/2}(1)}{I_3^{1/2}(1)}\left[1+\frac{1}{8\sqrt{2}}\nu\ln\nu\left(\frac{3}{(I_2-I_1)(1)}-\frac{1}{I_3(1)}\right)\right.\\
&\left.+\frac{\nu}{4\sqrt{2}}\left(\frac{3}{(I_2-I_1)(1)}\left(-\frac{7}{2}+2\sqrt{2}-\frac{5}{2}\ln 2\right)-\frac{1}{I_3(1)}\left(-\frac{1}{2}-\frac{5}{2}\ln 2\right)\right)+\mathcal{O}(\nu)\right]\,.
\end{aligned}$$

Using that $I_3(1) = \sqrt{2}/3$ and $(I_2-I_1)(1) = 3I_3(1) = \sqrt{2}$, this now reads

$$\frac{(I_2-I_1)^{3/2}(1-\nu)}{I_3^{1/2}(1-\nu)} = \sqrt{6}\left(1+\frac{3}{8}\nu\left(2\sqrt{2}-3\right)\right)\,.$$

Bibliography II

[Bog47] Nikolay N. Bogoliubov, *On the theory of superfluidity*, Journal of Physics – USSR **11** (1947), no. 1.

[Gir60] Marvin D. Girardeau, *Relationship between Systems of Impenetrable Bosons and Fermions in One Dimension*, Journal of Mathematical Physics **1** (1960), no. 6, 516.

[LL63] Elliott H. Lieb and Werner Liniger, *Exact Analysis of an Interacting Bose Gas. I. The General Solution and the Ground State*, Physical Review **130** (1963), no. 4, 1605–1616.

[Lie63] Elliott H. Lieb, *Exact Analysis of an Interacting Bose Gas. II. The Excitation Spectrum*, Physical Review **130** (1963), no. 4, 1616–1624.

[Nam20] Phan Thàn Nam, *Mathematical Quantum Mechanics II, Lecture notes*, 2020. available at `https://www.math.lmu.de/~nam/LectureNotesMQM2020.pdf`.

[NRS18a] Marcin Napiórkowski, Robin Reuvers, and Jan Philip Solovej, *The Bogoliubov Free Energy Functional I: Existence of Minimizers and Phase Diagram*, Archive for Rational Mechanics and Analysis **229** (2018), no. 3, 1037–1090.

[NRS18b] ———, *The Bogoliubov free energy functional II: The dilute limit*, Communications in Mathematical Physics **360** (2018), no. 1, 347–403.

[Sol14] Jan Philip Solovej, *Many Body Quantum Mechanics*, 2014. lecture notes, draft, with corrections and additions of P. T. Nam. An earlier version was published at `http://web.math.ku.dk/~solovej/mbnotes`.

[Sol06] ———, *Upper bounds to the ground state energies of the one- and two-component charged Bose gases*, Communications in Mathematical Physics **266** (2006), 797–818.

[TW16] Craig A. Tracy and Harold Widom, *On the ground state energy of the δ-function Bose gas*, Journal of Physics A: Mathematical and Theoretical **49** (2016), 294001.