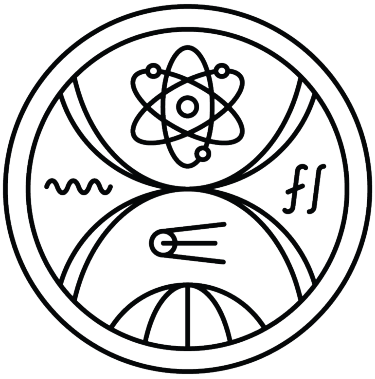
COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

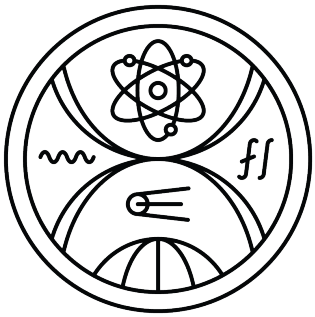


Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model

ACADEMIC DISSERTATION

2025 **Lukáš Holka**

COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND
INFORMATICS



Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model

ACADEMIC DISSERTATION

Study programme:	**Nuclear and Subnuclear Physics**
Field of study:	**13. - Physics**
External educational institution:	**Institute of Physics, Slovak Academy of Sciences**
Department:	**Department of Theoretical Physics**
Supervisor:	**Mgr. Erik Bartoš, PhD.**
Consultant:	**RNDr. Stanislav Dubnička, DrSc.**

2025 **Lukáš Holka**

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Mgr. Lukáš Holka
Study programme:	Nuclear and Subnuclear Physics (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	Physics
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model

Annotation: The problem is related to the experimental measurement of the differential effective cross section of the elastic scattering of muons on atomic electrons Be or C, in the COMPASS experiment at CERN as a function of the square of the transferred momentum and the subsequent extraction of the behavior of the running coupling constant. It consists of contributions from leptons, five light quarks with a mass less than 5 GeV and a heavy "top" quark. In order to extract the hadron contribution from light quarks determining the LO of the hadron contribution to the anomalous magnetic moment of the muon, it is necessary to know the behavior of the lepton contributions in the space-like region. If in such a case it has the same logarithmic dependence as in the time-like region, since the square of the transferred momentum is negative, the complex value LO reaches the anomalous magnetic moment of the muon, which contradicts its classical evaluation using the dispersion relation over experimental values or R(s).

Aim: The reconstruction of the behavior of the lepton contribution to the running fine-structure constant in the time-like region and the finding its explicit shape in the space-like region.

Tutor:	Mgr. Erik Bartoš, PhD.
Consultant:	RNDr. Stanislav Dubnička, DrSc.
Department:	FMFI.KJFB - Department of Nuclear Physics and Biophysics
Head of department:	prof. RNDr. Jozef Masarik, DrSc.
Assigned:	03.05.2021

Approved: 15.07.2021 prof. RNDr. Jozef Masarik, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Mgr. Lukáš Holka
Študijný program: jadrová a subjadrová fyzika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: fyzika
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model
Správanie bežiacej konštanty jemnej štruktúry v priestorupodobnej a časupodobnej oblasti a iné aplikácie Unitárneho a analytického modelu

Anotácia: Problém súvisí s experimentálnym meraním diferenciálneho účinného prierezu pružného rozptylu muónov na atomových elektrónoch Be, prípadne C, na experimente COMPASS v CERN ako funkcie kvadrátu prenesenej hybnosti a následného vytiahnutia správania sa bežiacej väzbovej konštanty. Tá pozostáva z príspevkov od leptónov, piatich ľahkých kvarkov s hmotnosťou menšou ako 5 GeV a ťažkého „top“ kvarku. Aby sa z nej mohol vytiahnuť hadrónový príspevok od ľahkých kvarkov, určujúci LO hadrónového príspevku do anomálneho magnetického momentu miónu, potrebuje sa poznať správanie sa leptónových príspevkov v priestorupodobnej oblasti. Ak v takom prípade má rovnakú logaritmickú závislosť ako v časupodobnej oblasti, nakoľko kvadrát prenesenej hybnosti je záporný, dostáva sa komplexná hodnota LO k anomálnemu magnetickému momentu miónu, čo protirečí s jeho klasickým vyhodnotením pomocou disperzného vzťahu cez experimentálne hodnoty R(s).

Cieľ: Rekonštrukcia správania sa leptónového príspevku do bežiacej konštanty jemnej štruktúry v časupodobnej oblasti a nájdenie jej explicitného tvaru v priestorupodobnej oblasti.

Školiteľ: Mgr. Erik Bartoš, PhD.
Konzultant: RNDr. Stanislav Dubnička, DrSc.
Katedra: FMFI.KJFB - Katedra jadrovej fyziky a biofyziky
Vedúci katedry: prof. RNDr. Jozef Masarik, DrSc.

Dátum zadania: 03.05.2021

Dátum schválenia: 15.07.2021 prof. RNDr. Jozef Masarik, DrSc.
garant študijného programu

.....
študent

.....
školiteľ

Declaration of Authorship

I, Lukáš Holka, declare that this thesis, titled "*Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model*", and the work presented in it are my own. I confirm that

- This work was done whole or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed:

Date:

Acknowledgment

I would like to express my gratitude to Erik Bartoš, Stanislav Dubnička, Anna Zuzana Dubničková, and Andrej Liptaj for all their help and support during my work on this thesis. I am thankful to John Collins for his insightful comments on various topics during our correspondence. I am also grateful to my family, especially my wife Gaziza, for their patience and support.

Lukáš

Comenius University Bratislava
Faculty of Mathematics, Physics, and Informatics

Abstract

Philosophiae Doctor

Behavior of the running fine-structure constant in the space-like and time-like regions and other applications of the Unitary and Analytic model

Lukáš Holka

The first part of this thesis presents an introduction to some topics of the analytic structure of scattering amplitudes and related objects in quantum field theory. These techniques are necessary both in the construction of the Unitary and Analytic model, and in the study of the running of the fine structure constant. Then, in the second part, electromagnetic form factors and their Unitary and Analytic model are introduced. The last part of the thesis consists of a review of certain properties of the running fine structure constant and a summary of results that were achieved during my doctoral study. A detailed proof of the reality of the fine structure constant in the space-like region is given, and the relation between the imaginary part of that running constant in the time-like region and the total cross section for the electron–positron annihilation is reviewed. The thesis concludes with the summary of several results that I have achieved in collaboration with my colleagues. In particular, we have described the hadronic contribution to the running of the fine structure constant using the Unitary and Analytic model, and used that description to evaluate the leading hadronic contribution to the muon's magnetic anomaly. I also present our investigation of the damped oscillatory structures that were observed in the effective form factor of the proton, and of the anomaly in the charged to neutral yield ratio for the decay of the $\phi(1020)$ resonance into kaon pairs.

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky, a informatiky

Abstrakt

Philosophiae Doctor

Správanie bežiacej konštanty jemnej štruktúry v priestorupodobnej a časupodobnej oblasti a iné aplikácie Unitárneho a analytického modelu

Lukáš Holka

Táto dizertačná práca začína úvodom do analytickej štruktúry amplitúd rozptylu a príbuzných veličín v kvantovej teórii poľa. Tieto techniky sú potrebné ako na uvedenie Unitárneho a analytického modelu, tak i na štúdium bežania konštanty jemnej štruktúry. V druhej časti práce sú potom prezentované elektromagnetické form faktory a ich Unitárny a analytický model. Posledná časť práce obsahuje prehľad vybraných vlastností bežiacej konštanty jemnej štruktúry a súhrn výsledkov dosiahnutých počas môjho doktorského štúdia. Je v nej uvedený podrobný dôkaz toho, že v priestorupodobnej oblasti je bežiaca konštanta jemnej štruktúry reálna. Tiež je prezentovaný vzťah medzi imaginárnou časťou bežiacej konštanty v časupodobnej oblasti a totálnym účinným prierezom elektrón pozitrónovej anihilácie. Práca je zakončená zhrnutím niekoľkých výsledkov, ktoré som dosiahol v spolupráci so svojimi kolegami. Predovšetkým sa nám podarilo popísať hadrónový príspevok do bežania konštanty jemnej štruktúry pomocou Unitárneho a analytického modelu, a za použitia tohoto popisu vyhodnotiť vedúci príspevok hadrónov do magnetickej anomálie miónu. Tiež je prezentovaná naša štúdia tlmených oscilácií, ktoré boli pozorované v efektívnom form faktore protónu, a anomálie v pomere rozpadových šírok rezonancie $\phi(1020)$ na nabité a neutrálne kaóny.

Contents

Preface

The main aim of this thesis is to introduce the Unitary and Analytic (U&A) model of electromagnetic form factors of hadrons, as well as some of its applications on which I have worked on during my doctoral study. The U&A model is based on particular analytic properties which form factors are believed to possess. Before we introduce the model, we will therefore need to first discuss some topics of analyticity in particle physics.

The thesis is divided into three parts. In the first part, which is comprised of chapters 1 and 2, we will discuss what appears to be the most important way in which analyticity enters into particle physics. From the point of view of this thesis, this is important in order to introduce the U&A model, but also to establish some results which are going to be needed later in the final, third part of the thesis. These aims have to a large extent informed the selection of topics that are covered in this first part of the thesis. But analytic properties of amplitudes constitute an important and fascinating topic in high energy physics and their implications reach far beyond any single phenomenological model, such as the U&A model, or specific applications, such as those presented in this thesis. For this reason I have allowed myself to mention also some topics which are only adjacent to our present goals, but which I still found important to introduce from a more general point of view. In fact, I hope that the first part of this thesis can serve quite independently from the rest of the work as a gentle introduction to the topic of analyticity in high energy physics.

The second part consists of chapters 3 and 4. We will introduce the general concept of form factor and mention the analytic properties that it is usually assumed to posses in the chapter 3, which in turn will allow us to introduce the U&A model in the chapter 4. The main purpose of the chapter 3 is to make the thesis more self-contained. The U&A model describes electromagnetic form factors of hadrons, and it is therefore appropriate to present first the general notion of form factor. This will also come handy later when we will discuss the magnetic anomaly of the muon. Considering, however, that the topic of form factors is quite large, in order to keep the volume of the thesis manageable I decided to present this topic in a manner that is somewhat less comprehensive than the first part of the thesis. The chapter 4, on the other hand, is very thorough and introduces the U&A

model in quite a detail. Even though the model itself is not new, the style of the presentation to large extent is. I have tried to clearly describe the structure and principles behind the construction of the model, and I hope that this chapter will prove useful to any researcher who wishes to understand how the model works, and how to build or modify it for their own purposes.

The last, third part of the thesis is then concerned with some of the results that I have obtained in collaboration with my colleagues during my doctoral study. In the chapter 5 I discuss some aspects of the running of the fine structure constant, including its significance for the evaluation of the magnetic anomaly of the muon. I describe how its imaginary part can be obtained from measurements of the total cross section in the time-like region, and I present a proof that the constant is real in the space-like region. Afterwards I also mention some other, more particular results, including the contribution of the hadronic vacuum polarization into the magnetic anomaly of the muon. The thesis is then concluded by chapter 6, in which I describe two other applications of the U&A model on which I collaborated.

Lukáš Holka

Part I

An introduction to the study of analyticity in quantum field theory

This first part of the thesis discusses some features of the analytic structure of scattering amplitudes. My aim was mainly to introduce the concepts and explain why it might be reasonable to assume certain analytic properties of various quantities, and I did not attempt to provide rigorous proofs or a complete treatment of this topic. For practical reasons I have partly limited the scope towards issues that were necessary for the introduction of the U&A model, but I have also tried to provide quite a general overview of this field, or, at least, of some of its main principles. I hope that Part 1 will be useful also for a reader who is not interested in the remaining two parts of the thesis. Nonetheless, for a more thorough treatment the reader is advised to consult other literature, and I will now briefly recommend some sources that the reader may find interesting.

A classic textbook for these topics is *The Analytic S-Matrix* written by Eden, Landshoff, Olive, and Polkinghorne [1]. I do not know of any single modern treatment that would fully substitute for this book. However, one disadvantage of the book is that at the time of its writing some issues covered there were not yet properly understood and the book therefore contains many subtle mistakes.

A good review of the analyticity in the context of strong coupling physics was published by S. Mandelstam in 1962 [2]. This review was published during the times when the S-matrix program was very popular, and provides some insight into that historical period. Similarly to [1] it is not fully up-to-date, but it is very well-written and offers a lot of value even today. Let me, however, add one comment. At the time when [2] was written it was generally assumed that the condition of renormalizability constraints the physically admissible set of field theories. It was believed that non-renormalizable theories should be discarded as unphysical. This played a role in the plausibility argument for the main goal of the S-matrix program, which was to fully determine the S-matrix just from a small set of mathematical properties (causality, unitarity, and crossing-symmetry) up only to a finite number of constants, the values of which could be measured by experiments. Today the view on non-renormalizability is more complex.

There is an excellent treatment by H. Nussenzveig: *Causality and dispersion relations* [3]. It is mathematically much more precise than the above-mentioned references. Its limitation is that it does not venture beyond non-relativistic quantum mechanics. Nevertheless, I would highly recommend this book to anyone who wishes to study the related mathematics, as well as some of the physics, to a greater depth.

I also highly recommend the recently published review by S. Mizera [4], which is a modern and very readable introduction to the subject of analyticity in quantum physics. Although it is meant to serve as an introduction to the subject and is therefore by design not very detailed or complete, it covers a larger ground than most other sources and also mentions some modern topics and recent results.

A very authoritative source for an important subset of the results is the classic book *PCT, Spin and Statistics, and All That* by R. Streater and A. Wightman [5]. One can find there a precise mathematical formulation of the relation between distributions whose support is limited to a cone and analyticity and boundedness properties of their Laplace transforms. This relation is the main pillar which supports much of what we are going discuss in this thesis.

Chapter 1

Causality, analyticity, and dispersion relations

Our first aim is to introduce the relation between causality and analyticity in one of its most elementary forms. Roughly speaking, the mechanism is as follows. We consider some function, or more generally a tempered distribution, that vanishes on a suitable subset of its domain. A simple case of this would be a function f of a single variable x that is zero for all $x < 0$. As we will see later, this can be further generalized to a distribution whose support vanishes outside of some cone. The property that $f(x) = 0$ for $x < 0$ then translates into particular constraints on the Fourier transform of f. Those constraints are that the Fourier transform is the boundary value of an analytic function with certain asymptotic properties. Later, we will present this statement in a greater detail, but for now our goal is only to introduce the concept and build some basic intuition. For that reason we will begin by considering the simple example of a harmonic oscillator.

1.1 Classical harmonic oscillator

A traditional introductory example (see e.g. [3, 4]) is that of a classical damped harmonic oscillator subjected to an external force. The position x of such an oscillator, when considered as a function of time t, satisfies the following differential equation,

$$\ddot{x}(t) + 2\lambda\dot{x}(t) + \omega_0^2 x(t) = f(t). \tag{1.1}$$

As usual, the dots above x correspond to time derivatives, and the parameter λ governs the damping of oscillations. The set of physical values of λ is $\lambda \geq 0$. Such values correspond to the oscillations being damped rather than amplified. The corresponding homogeneous equation has two independent solutions, in the familiar form $x(t) \propto \exp(-i\omega t)$. After we plug this into the

homogeneous version of (1.1) we find the following two solutions for ω:

$$\omega_{\pm} = -i\lambda \pm \sqrt{\omega_0^2 - \lambda^2}. \tag{1.2}$$

In order to find a solution of the inhomogeneous equation (1.1) it is convenient to inspect the Fourier transform of the equation.

First, however, we need to fix our notation. For a function $h(t)$ we will define its Fourier transform $\tilde{h}(\omega)$ as

$$\tilde{h}(\omega) = \mathcal{F}[h](\omega) = \int_{-\infty}^{+\infty} dt\; h(t)\mathrm{e}^{i\omega t}. \tag{1.3}$$

The inverse Fourier transform is then

$$h(t) = \mathcal{F}^{-1}[\tilde{h}](t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega\; \tilde{h}(\omega)\mathrm{e}^{-i\omega t}. \tag{1.4}$$

Now we can express $x(t)$ and $f(t)$ as inverse Fourier transforms

$$x(t) = \mathcal{F}^{-1}[\tilde{x}](t) = \frac{1}{2\pi}\int d\omega\; \tilde{x}(\omega)\mathrm{e}^{-i\omega t}, \tag{1.5}$$

$$f(t) = \mathcal{F}^{-1}[\tilde{f}](t) = \frac{1}{2\pi}\int d\omega\; \tilde{f}(\omega)\mathrm{e}^{-i\omega t}. \tag{1.6}$$

The equation (1.1) is then equivalent to the condition

$$\left(-\omega^2 - 2i\lambda\omega + \omega_0^2\right)\tilde{x}(\omega) = -\left(\omega - \omega_+\right)\left(\omega - \omega_-\right)\tilde{x}(\omega) = \tilde{f}(\omega). \tag{1.7}$$

The solution is therefore

$$\tilde{x}(\omega) = \frac{-1}{\left(\omega - \omega_+\right)\left(\omega - \omega_-\right)}\tilde{f}(\omega) = \tilde{g}(\omega)\tilde{f}(\omega), \tag{1.8}$$

where we denoted $\tilde{g}(\omega) = -1/\left(\omega - \omega_+\right)\left(\omega - \omega_-\right)$. The full solution is then $\mathcal{F}^{-1}[\tilde{x}(\omega)](t)$, where $\tilde{x}$ is given by (1.8), plus a linear combination of the two independent solutions $\exp(-i\omega_+ t)$ and $\exp(-i\omega_- t)$ of the homogeneous equation. Any solution of the homogeneous equation, however, blows up for $t \to -\infty$. Such solutions are therefore sensible only when considering the evolution on a time interval that is bounded from below. This is not our case. In fact, we are interested only in the response of the oscillator to the external force. Therefore, for our purposes, the full solution is just the inverse Fourier transform of (1.8).

Let us now take a step back and consider how causality and analyticity enter into this picture. Our system is that of an oscillator reacting to an external force. A reasonable criterion of causality therefore seem to be to demand that the oscillator, if being at rest initially, remains at rest until the force starts acting on it. In other words, its response to the external force must not precede the action of the force.

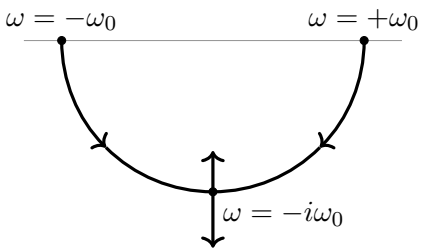


Figure 1.1: The poles of $\tilde{g}(\omega)$ as a function of λ. From (1.2) we see that for $\lambda = 0$ the two poles are at the points $\omega = -\omega_0$ and $\omega = +\omega_0$ on the real axis and as λ increases they move below the real axis along the circle $|\omega| = \omega_0$. They meet at $\omega = -i\omega_0$ for $\lambda = \omega_0$. Afterwards, one of the poles continues along the imaginary axis downwards to the infinity while the other one continues along the imaginary axis upwards towards $\omega = 0$.

Does our solution satisfy this property? Notice, that if we define

$$g(t) = \mathcal{F}^{-1}[\tilde{g}](t) = \frac{1}{2\pi}\int d\omega\ \tilde{g}(\omega)\mathrm{e}^{-i\omega t}, \tag{1.9}$$

then we can write

$$\begin{aligned} x(t) &= \frac{1}{2\pi}\int d\omega\ \tilde{g}(\omega)\tilde{f}(\omega)\mathrm{e}^{-i\omega t} \\ &= \frac{1}{2\pi}\int d\omega\ \tilde{g}(\omega)\left[\int dt' f(t')\mathrm{e}^{i\omega t'}\right]\mathrm{e}^{-i\omega t} = \int dt' g(t-t')f(t'), \end{aligned} \tag{1.10}$$

where in the second equality we wrote $\tilde{f}(\omega)$ as the Fourier transform of $f(t)$, and then assumed that it is possible to change the order of integration. From this equation we see that the response of the oscillator to the force acting at time t_0 is given by the Green's function g as $g(t-t_0)$. (Formally, if $f(t) = \delta(t-t_0)$ then $x(t) = g(t-t_0)$.) Hence, the above-mentioned criterion of causality[1] translates to the condition

$$g(t) = 0 \text{ for } t < 0. \tag{1.11}$$

Notice now, that for $\lambda > 0$ the function

$$\tilde{g}(\omega) = \frac{-1}{(\omega-\omega_+)\,(\omega-\omega_-)} \tag{1.12}$$

has poles only in the lower half-plane of ω. [The poles are at ω_+ and ω_- and from (1.2) we can calculate their positions as a function of λ. See also Fig. 1.1.]

[1]Formally, the criterion stated that if $f(t) = 0$ for $t < t_0$ then $x(t) = 0$ for $t < t_0$.

This allows us to use the following argument to show that $g(t) = 0$ for $t < 0$. If $t < 0$ the exponent in (1.9) is $+i\omega|t|$, which means that the factor $\exp(-i\omega t)$ provides an exponential suppression of the integrand when $\Im(\omega) > 0$. In other words, there is an exponential suppression when ω is in the upper half-plane. We can therefore close the contour of integration using a large semicircle in the upper half-plane, and use the asymptotic properties of $|\tilde{g}|$ for large $|\omega|$ to show that when the contour is stretched all the way to infinity the integral over this whole contour has the same value as the original integral (1.9). However, since $\tilde{g}$ is analytic in the upper half-plane, the integral over the whole contour vanishes by the Cauchy's theorem, which implies that (1.9) is also zero. Hence, if $t < 0$ then $g(t) = 0$.

[Since a similar technique will be employed over and over again throughout this work, let us take a time and spell out this argument more carefully. Instead of (1.9) consider

$$g(t) = \frac{1}{2\pi} \int_{-R}^{+R} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t}, \tag{1.13}$$

for some $R > 0$. The integral (1.9) is then recovered in the limit of $R \to +\infty$. Now, for a fixed R define the contour of integration $C(R)$ as going along the real axis from $-R$ to $+R$ and then upwards along a semicircle in the upper half-plane until we close the contour at the point $-R$ (see Fig. 1.2). Now, since $\tilde{g}$ is analytic in the whole upper half-plane, it is analytic inside the region enclosed by $C(R)$ for any $R > 0$ and the Cauchy's theorem implies that

$$\frac{1}{2\pi} \int_{C(R)} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t} = 0 \tag{1.14}$$

for any $R > 0$. Next, let us denote by $U(R)$ the semicircle part of $C(R)$. In other words, $U(R) = C(R) - [-R, +R]$. The integral

$$\frac{1}{2\pi} \int_{U(R)} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t} \tag{1.15}$$

satisfies the inequality

$$\left| \frac{1}{2\pi} \int_{U(R)} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t} \right| \leq \frac{1}{2\pi} \int_{U(R)} |d\omega| \; |\tilde{g}(\omega)|, \tag{1.16}$$

because in the upper half-plane the modulus of $\exp(-i\omega t)$ is less than 1. Hence, we can use the property that $|\tilde{g}(\omega)| = \mathcal{O}\left(|\omega|^{-2}\right)$ as $|\omega| \to \infty$ to show that the absolute value of the integral over $U(R)$ is $\mathcal{O}\left(R^{-1}\right)$ as $R \to \infty$. Sending $R \to \infty$ in (1.14) leads to the equation

$$0 = \lim_{R \to \infty} \left(\int_{-R}^{+R} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t} + \int_{U(R)} d\omega \; \tilde{g}(\omega) \mathrm{e}^{-i\omega t} \right), \tag{1.17}$$

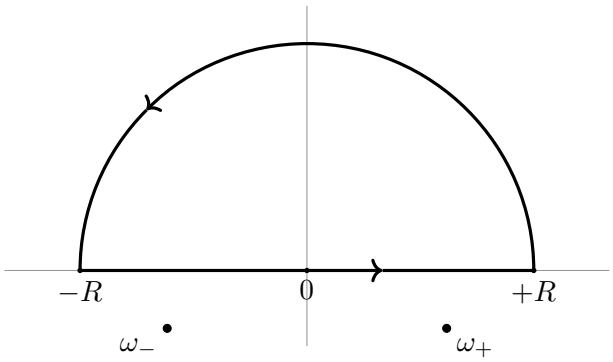


Figure 1.2: To show that $g(t) = 0$ for $t < 0$ one can consider the integral (1.14) along the contour $C(R)$ depicted above. The contour starts at $-R$, continues along the real axis to $+R$, and then it returns back to $-R$ along the semicircle in the upper half-plane. The singularities of $\tilde{g}(\omega)$ at $\omega = \omega_+$ and $\omega = \omega_-$ are restricted to the lower half-plane. By the Cauchy's theorem the integral over the whole $C(R)$ vanishes for all $R > 0$. In the limit $R \to \infty$ the integral over the semicircle goes to zero and the integral from $-R$ to $+R$ approaches (1.9).

and since the second integral goes to zero, so does the first one. This shows that the integral (1.9) vanishes when $t < 0$.

Let us note that since we had $|\tilde{g}(\omega)| \propto |\omega|^{-2}$ for $|\omega| \to \infty$ it was easy to show that the integral over $U(R)$ vanishes. We could show that the integral over $U(R)$ vanishes even if only a weaker bound $|\tilde{g}(\omega)| = \mathcal{O}\left(|\omega|^{-1}\right)$ for large $|\omega|$ was available, but a more detailed argument would then be necessary. In such cases one needs to make use of the fact that there is an exponential suppression in the upper half-plane, instead of relying only on the boundedness of the exponential factor. It is then necessary to suitably divide the path $U(R)$ into several segments and consider them separately. For example, we could divide $U(R)$ into the path starting at $-R$ and going upwards along the contour for a distance proportional to $\sqrt{R}$, a similar path but starting at $+R$, and the rest of $U(R)$. The absolute value of the integral over the two segments attached to $+R$ and $-R$ goes as $\mathcal{O}(R^{-1/2})$ for $R \to \infty$, and the absolute value of the integral over the rest of $U(R)$ decreases at least as $\exp(-\sqrt{R}|t|)$ for large R.]

We can use a similar technique to calculate $g(t)$ for $t > 0$. The exponential factor now becomes large for $\Im(\omega) > 0$, and we need to close the contour in the lower half-plane. Using the same argument as before we can again show that the integral over the large semicircle goes to zero as the contour is stretched to infinity. This time, however, the function $\tilde{g}$ is not analytic in the region enclosed by the contour: there are two poles at $\omega = \omega_+$ and

$\omega = \omega_-$ in that region. The theorem of residues then yields

$$\begin{aligned} g(t) &= -2\pi i \sum_{p\in\{\omega_+,\omega_-\}} \mathrm{Res}\left\{\frac{1}{2\pi}\tilde{g}(\omega)\mathrm{e}^{-i\omega t}; p\right\} \\ &= i\left(\frac{\mathrm{e}^{-i\omega_+ t}}{\omega_+ - \omega_-} + \frac{\mathrm{e}^{-i\omega_- t}}{\omega_- - \omega_+}\right) = \mathrm{e}^{-\lambda t}\frac{\sin\left(\sqrt{\omega_0^2-\lambda^2}\,t\right)}{\sqrt{\omega_0^2-\lambda^2}}, \end{aligned} \tag{1.18}$$

for $t > 0$. Here, $\mathrm{Res}\{f(z), a\}$ stands for the residue of $f(z)$ at $z = a$, and the sum of residues is multiplied by $-2\pi i$ because we integrate along the contour in the clockwise direction. Altogether, we have obtained the result

$$g(t) = \frac{\sin\left(\sqrt{\omega_0^2-\lambda^2}\,t\right)}{\sqrt{\omega_0^2-\lambda^2}}\mathrm{e}^{-\lambda t}\theta(t), \tag{1.19}$$

where θ stands for the Heaviside step function, which equals zero for negative values of its argument and one for positive values of the argument,

$$\theta(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 & \text{if } t > 0. \end{cases} \tag{1.20}$$

As we can see from (1.19) the oscillator responds to a brief impulse of force by damped oscillations with lifetime $1/\lambda$. The step function factor in (1.19) ensures that the response obeys our causality criterion: oscillations appear only after the force acts, no earlier.

Let us now take a step back and discuss our findings. In the context of the harmonic oscillator we formulated the causality condition in terms of the Green's function $g(t - t_0)$ as the condition that $g(t - t_0) = 0$ if $t < t_0$. To show that $g(t)$ satisfies this property we started from considerations of its Fourier transform $\tilde{g}(\omega)$. As we saw, the causality condition followed from the fact that $\tilde{g}$ was analytic in the upper half-plane and that it had suitable asymptotic properties for large $|\omega|$.

In fact, this argument can be reversed: if $g(t) = 0$ for $t < 0$ then $\tilde{g}(\omega)$ is analytic in the upper half-plane and has certain asymptotic properties. This is because for such functions $g(t)$ we have

$$\tilde{g}(\omega) = \int_{-\infty}^{+\infty} dt\; g(t)\mathrm{e}^{i\omega t} = \int_{0}^{+\infty} dt\; g(t)\mathrm{e}^{i\omega t}, \tag{1.21}$$

since for $t < 0$ the integrand vanishes. But for positive values of t the exponential factor $\exp(i\omega t) = \exp(i\Re\{\omega\}t)\exp(-\Im\{\omega\}t)$ serves to improve the integral's convergence when $\Im\{\omega\} > 0$, and as a consequence the function $\tilde{g}(\omega)$ is analytic in the upper half-plane. In following sections we will slightly generalize our findings and state the results in a more precise form.

Let us mention one last thing in relation to the harmonic oscillator. The observation that $\tilde{g}(\omega)$ was analytic in the upper half-plane relied on the condition that $\lambda > 0$. It was necessary that the oscillations were damped, or, in other words, that the system was dissipating energy. But what about the cases when $\lambda < 0$? When $\lambda < 0$, the system on its own (absent the external force) is unstable. If the external force does not act, any non-zero value of x at time t_0 results in arbitrarily large oscillations at times $t > t_0$. To obtain a finite solution the force must therefore act to suppress those oscillations. For example, if the external force was a short delta-function-like impulse, the finite solution would consist of exponentially increasing oscillations followed by the short burst of force acting to set the oscillator to rest. In other words, the force follows the oscillations, not the other way around. A simple way to think about this is to notice that the solutions for $\lambda < 0$ are the same as those for $-\lambda$ after we reverse the flow of time, $t \to -t$.

1.2 A simple generalization

The discussion of the previous section applies to a much larger set of systems. Let us now present a very simple generalization so that we can more clearly see which aspects do not depend on the specific form of the differential equation for the harmonic oscillator.

Consider an abstract system which reacts to an input $f(t)$ with a response $x(t)$, corresponding respectively to the external force and the position of the oscillator in the previous section. Let us now make the following assumptions about this system:

1. Let the response of the system be linear in the input. This corresponds to the property that

$$x(t) = \int dt_i \, g(t, t_i) f(t_i), \tag{1.22}$$

   where $g(t, t_i)$ is the response at t to the input of the form $f(t) = \delta(t - t_i)$.

2. Let the system be time-translation invariant. This implies that $g(t, t_i)$ depends only on the difference between t and t_i:

$$g(t, t_i) = g(t - t_i). \tag{1.23}$$

In this case we find that the response of the system is the convolution of the input f and the Green's function g

$$x(t) = (g * f)(t) = \int dt_i \, g(t - t_i) f(t_i). \tag{1.24}$$

Recall now that the Fourier transform of a convolution $a * b$ is the product of the Fourier transforms $\mathcal{F}[a]$ and $\mathcal{F}[b]$:

$$\begin{aligned}\mathcal{F}\left[a * b\right](\omega) &= \int dt\ (a * b)(t)\mathrm{e}^{i\omega t} = \int dt \int dt'\ a(t-t')b(t')\mathrm{e}^{i\omega t} \\ &= \int dt' \int dt \left[a(t-t')\mathrm{e}^{i\omega(t-t')}\right]\left[b(t')\mathrm{e}^{i\omega t'}\right] \\ &= \int dt' \int d(t-t') \left[b(t')\mathrm{e}^{i\omega t'}\right]\left[a(t-t')\mathrm{e}^{i\omega(t-t')}\right] \\ &= \int dt'\ b(t')\mathrm{e}^{i\omega t'} \int dt\ a(t)\mathrm{e}^{i\omega t} = \mathcal{F}\left[a\right](\omega)\ \mathcal{F}\left[b\right](\omega). \quad (1.25)\end{aligned}$$

Therefore, in terms of $\tilde{x} = \mathcal{F}[x]$, $\tilde{f} = \mathcal{F}[f]$, and $\tilde{g} = \mathcal{F}[g]$, the equation (1.24) reads

$$\tilde{x}(\omega) = \tilde{g}(\omega)\ \tilde{f}(\omega). \quad (1.26)$$

Thus, we see that for any linear time-translation-invariant system we recover the general structure of the previous section.

Just as before, a reasonable condition of causality is

$$g(t) = 0 \text{ for } t < 0. \quad (1.27)$$

This condition is sometimes called the *primitive causality* condition. Minor modifications are often necessary, as for example requiring that $g(t) = 0$ for all $t < t_0$ for some constant t_0 instead of (1.27), especially if the input and the response are measured at different locations in space.

The condition (1.27) on g then translates to the property that its Fourier transform $\tilde{g}$ is analytic in the upper half-plane and satisfies certain asymptotic properties. Such functions are sometimes called causal transforms and in the following sections we will present this statement in greater detail. We will also introduce another useful property that characterizes causal transforms: they satisfy equations known as dispersion relations.

1.3 Analytic continuation from a set of measured data

Let us now approach the same problem from another direction. Recall the following result from complex analysis. Consider a region G and two analytic functions f, g defined on G. Then, if the set $A = \{z \in G : f(z) = g(z)\}$ contains a limit point, the two functions are identical everywhere on G, that is, $A = G$. As a consequence, if a function is defined and analytic on G, it is fully determined by its values on any subset of G that contains a limit point, and in particular by its values on any interval[2]. Hence, if we

[2]To avoid any confusion, one should not forget that this kind of continuation is unique only as long as G remains fixed. It may be possible to further extend the definition domain

knew that some quantity $f(t)$ is an analytic function on the real axis (or, more precisely, in some neighborhood of the real axis), we could attempt to determine f everywhere just from measurements in some small interval of t. At first sight, this may appear to be a sensible plan, considering the uniqueness of such a continuation. However, that is wrong. The problem is that realistic measurements have only finite precision, and the analytic continuation is generally unstable. As an example, consider some analytic function $f(t)$ and another one defined by $g(t) = f(t)\exp(iat)$, where the parameter a is real and very small. For any fixed interval (t_1, t_2) and a fixed measurement precision there exists a sufficiently small value of a such that the functions g and f are indistinguishable on that interval given the measurement precision. However, when continued to the lower half-plane, the factor $\exp(iat)$ grows exponentially with $-\Im(t)$, and the two functions eventually become measurably different.

There are ways to address this issue and devise methods of analytic continuation that are more suitable for real world applications. Consider the situation where $f(z)$ is analytic in some simply connected region G and continuous on its closure. (At this point it becomes useful to specify what is meant under the term 'region'. We define a region to be an open connected subset of the complex plane. A simply connected region is a region without any holes: any closed path in such a region can be contracted continuously to a single point. Here we invoked the simple connectedness just for the sake of simplicity.) For any $z_0 \in G$ the Cauchy's integral formula yields

$$f(z_0) = \frac{1}{2\pi i}\int_{\partial G} dz \frac{f(z)}{z - z_0}, \tag{1.28}$$

where we integrate over the boundary ∂G of G in the counterclockwise direction. This method allows us to continue from the boundary into the whole region, and it is easy to see that it does not suffer from instabilities such as those observed in the previous paragraph: if the function values on the boundary are shifted from $f(z)$ to $f(z) + \Delta(z)$, the effect this has on the value of $f(z_0)$ given by (1.28) is bounded from above by $\max_{z\in\partial G} |\Delta(z)|$. [When the boundary values change from $f(z)$ to $f(z) + \Delta(z)$, the left hand side of (1.28) changes from $f(z_0)$ to $f(z_0) + \Delta(z_0)$, where

$$\Delta(z_0) = \frac{1}{2\pi i}\int_{\partial G} dz \frac{\Delta(z)}{z - z_0}, \tag{1.29}$$

because the right hand side of (1.28) is linear in f. A simple manipulation

of the function to some larger set $G' \supset G$, but in that case it is not guaranteed that the continuation is unique. If there are two such continuations, one onto a set $G_1 \supset G$ and the other one onto a set $G_2 \supset G$, then there might be points in $G_1 \cap G_2$ where the two continuations differ. In other words, the analytic continuation is generally path dependent. We will return to this issue later in the chapter, but for now we can ignore this possibility.

then yields for its modulus $|\Delta(z)|$ the inequality

$$|\Delta(z_0)| = \left|\frac{1}{2\pi i}\int_{\partial G} dz \frac{\Delta(z)}{z - z_0}\right| \leq \frac{1}{2\pi}\int_{\partial G} |dz| \frac{|\Delta(z)|}{|z - z_0|}. \tag{1.30}$$

We can parameterize the integration over the boundary by the angle ϕ formed by the intersection of the line from z_0 to z and some fixed reference line passing through z_0. Denoting $|z - z_0| = r(\phi)$ we obtain

$$|\Delta(z_0)| \leq \frac{1}{2\pi}\int_{\partial G} |dz| \frac{|\Delta(z)|}{|z - z_0|} = \frac{1}{2\pi}\int_0^{2\pi} r(\phi) d\phi \frac{|\Delta\left(z(\phi)\right)|}{r(\phi)} \leq \max_{z \in \partial G} |\Delta(z)|, \tag{1.31}$$

which demonstrates the stability of this method of continuation.]

The function continued by the formula (1.28) from the boundary values is sometimes called the Cauchy transform. The stability of this method of continuation is closely linked to another result concerning analytic functions, the maximum modulus theorem. This theorem states[3] that the modulus of an analytic function defined on a region G assumes its maximum value on the boundary ∂G.

When one attempts to make use of formulas such as (1.28) in physical applications, the following inconvenience may easily occur: one wishes to consider a form of (1.28) where the region G is very large or unbounded. (Unbounded regions are dealt with in the usual manner of considering a suitable limit in which a bounded region is stretched to infinity.) In order to apply (1.28) one then needs to know the values of f on distant parts of the boundary, which can be a problem. However, there is an important class of cases for which this issue disappears. These are when the modulus of f decreases sufficiently fast so that the integral over the distant part of the boundary is zero. We next limit our attention to these cases and derive the so-called dispersion relations.

[3]The proof of this theorem is very simple if we assume validity of another key result of complex analysis: the open mapping theorem, which states that non-constant analytic functions map open sets onto open sets. Hence, if f is an analytic function defined on a region G and f is not constant, then the set $f(G) = \{f(z) : z \in G\}$ is open, and this means that for any point $\xi \in f(G)$ there is an open ball $B(\xi; r) = \{o : |o - \xi| < r\}$ (for some $r > 0$) contained in $f(G)$: $B(\xi; r) \subset f(G)$. But $B(\xi; r)$ contains points of larger modulus than ξ, and since $B(\xi; r) \subset f(G)$, so does $f(G)$. This shows that there is no point ξ in $f(G)$ such that $|f(z)| \leq |\xi|$ for all $z \in G$. On the other hand, if f is continuous on the closure of G then its modulus must assume its maximal value somewhere on it. If f is not constant, the previous argument then implies that this maximal value of the modulus must be assumed on the boundary ∂G. When f is constant, the maximum modulus is assumed everywhere, including the boundary, which concludes the proof. Observe that in a similar way we can prove analogous statements for the real component of f instead of the modulus, or for the imaginary component, or for the projection onto any fixed line.

1.4 Causal transforms and dispersion relations

There are various ways to obtain dispersion relations. For now, we will stick to the approach that, as we will see, corresponds to the condition that $g(t) = 0$ for $t < 0$. There is also the question of what class of functions or distributions should we consider. We will need to require certain asymptotic properties. Our main goal is to explore the elementary concepts and build some intuition, so for now we will limit our attention to the relatively simple case of functions that are square integrable.

A square integrable function $g(t)$ satisfies

$$\int_{-\infty}^{+\infty} dt\ |g(t)|^2 = C, \tag{1.32}$$

for some finite constant C. These functions can be Fourier transformed and the Parseval's theorem then states that the Fourier transform $\mathcal{F}[g](\omega) = \tilde{g}(\omega)$ is also square integrable. Although in this chapter I try to state some of the results precisely, I do not attempt to be mathematically rigorous in their derivation (if derivations are offered at all). Under these relaxed standards we can obtain the integral of $|\tilde{g}|^2$ as follows

$$\begin{aligned}\int d\omega\ |\tilde{g}(\omega)|^2 &= \int d\omega\ dt\ dt'\ \left[g(t)\mathrm{e}^{i\omega t}\right]\left[g(t')^*\mathrm{e}^{-i\omega t'}\right] \\ &= \int dt\ dt'\ 2\pi\delta(t-t')g(t)g(t')^* = 2\pi C. \end{aligned}\tag{1.33}$$

If $g(t)$ also satisfies the primitive causality condition (1.27), then if we consider a complex $\omega = \alpha + i\beta$, and integrate over the line $\beta = \text{const} > 0$ in the upper half-plane, we obtain

$$\int_{-\infty}^{\infty} d\alpha\ |\tilde{g}(\alpha+i\beta)|^2 = \int_0^{\infty} dt \int_0^{\infty} dt'\ 2\pi\delta(t-t')g(t)g(t')^*\mathrm{e}^{-\beta(t+t')} < 2\pi C, \tag{1.34}$$

because for $\beta > 0$ the factor $\exp(-\beta(t+t'))$ provides an additional suppression of the integrand. Hence, in this case the function $\tilde{g}(\alpha + i\beta)$ is square integrable when considered as a function of α with β fixed at any positive value — in other words, it is square integrable along any line in the upper half-plane and parallel to the real axis — and all such integrals satisfy a common bound. Furthermore, it can be shown rigorously that $\tilde{g}(\omega)$ is analytic in the upper half-plane. Functions with these properties are called *causal transforms*, and we will focus on them for the rest of this section.

To summarize, we consider a function $f(z)$ that has the following properties:

- $f(z)$ is analytic in the upper half-plane;
- $f(x+iy)$, when considered as a function of x, is square integrable for any $y \geq 0$;

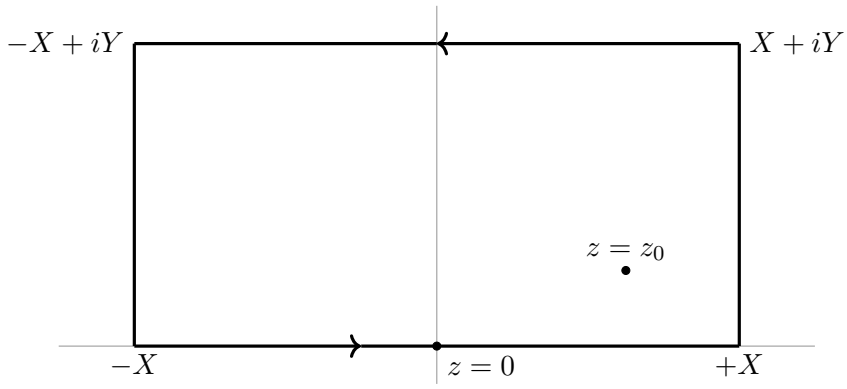


Figure 1.3: An integration contour suitable for the application of the Cauchy's integral formula to square integrable functions analytic in the upper half-plane. Both X and Y are eventually send to $+\infty$.

- there exists a common bound $2\pi C$ such that $\int dx|f(x+iy)|^2 \leq 2\pi C$ for any $y \geq 0$.

Note that the square integrability property implies

$$f(x+iy) \to 0 \text{ when } |x| \to \infty, \tag{1.35}$$

for any $y \geq 0$.

We will now apply the Cauchy's integral formula (1.28) with a suitably chosen integration contour that will make it convenient to exploit these properties. Fix a point $z_0 = x_0 + iy_0$ with $y_0 > 0$. Consider the following rectangular contour of integration (depicted in Fig. 1.3): for some values of X, Y both real and positive, start at the point $-X$, go along the real axis to the point $+X$, then parallel to the imaginary axis to $X+iY$, then along the horizontal line to $-X+iY$, and then vertically back to $-X$. The values of X and Y must be chosen such that $-X < x_0 < X$ and $0 < y_0 < Y$, i.e., that the point z_0 is in the region enclosed by the contour of integration.

We now proceed in steps to consider the integral (1.28) consecutively along each side of this rectangular contour. The integral from X to $X+iY$ can be expressed as

$$\frac{1}{2\pi i}\int_0^Y idy \frac{f(X+iy)}{\sqrt{(X-x_0)^2+(y-y_0)^2}},$$

and its modulus can be bounded by the following expression

$$\frac{1}{2\pi}\max_{0\leq y\leq Y}\{|f(X+iy)|\}\int_0^Y dy \frac{1}{\sqrt{(X-x_0)^2+(y-y_0)^2}}. \tag{1.36}$$

We now consider the behavior of this integral as we send X to the infinity. For a fixed value Y the maximum $\max_{0\leq y\leq Y}\{|f(X+iy)|\}$ of $|f(z)|$ over the

right side of the rectangle goes to zero as $X \to \infty$, because of the property (1.35). Furthermore, for large X the integral in (1.36) is approximately equal[4] to Y/X, and this factor also vanishes with $X \to \infty$. Hence, in the limit of $X \to \infty$ the integral over the right side of the rectangle vanishes. By an entirely similar argument, the integral over the left side of the rectangle vanishes when we send $-X \to -\infty$.

Putting these two results together we see that

$$f(x_0 + iy_0) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} dx \frac{f(x)}{x - (x_0 + iy_0)} + \frac{1}{2\pi i} \int_{+\infty}^{-\infty} dx \frac{f(x + iY)}{(x + iY) - (x_0 + iy_0)}, \quad (1.37)$$

for any $Y > y_0$. The second integral on the right-hand side corresponds to the integration over the upper side of the rectangle whose left and right sides were moved to the minus infinity and the plus infinity, respectively. We can show that this integral vanishes using the Schwarz inequality, which yields

$$\left| \int_{-\infty}^{+\infty} dx \frac{f(x + iY)}{(x - x_0) + i(Y - y_0)} \right| \leq \left[\int_{-\infty}^{+\infty} dx |f(x + iY)|^2 \right]^{1/2} \times \left[\int_{-\infty}^{+\infty} dx \frac{1}{(x - x_0)^2 + (Y - y_0)^2} \right]^{1/2}. \quad (1.38)$$

Now, the first integral on the right hand side is bounded by $2\pi C$ and the second one equals $\pi/(Y - y_0)$. [We can calculate this integral for instance by closing the contour of integration using a large semicircle in one of the half-planes and picking up the contribution of the pole that lies in that half-plane.] This provides us with the bound

$$\left| \int_{-\infty}^{+\infty} dx \frac{f(x + iY)}{(x - x_0) + i(Y - y_0)} \right| \leq \sqrt{\frac{2\pi^2 C}{Y - y_0}}, \quad (1.39)$$

which vanishes with $Y \to +\infty$. In fact, except for this integral no other term in (1.37) depends on the value of Y, which means that the integral

[4] Alternatively, we could also use the substitution $\sqrt{a^2 + y^2} = t - y$, which leads to

$$\frac{dy}{\sqrt{a^2 + y^2}} = \frac{dt}{t},$$

and evaluate the integral explicitly:

$$\int_0^Y dy \frac{1}{\sqrt{(X - x_0)^2 + (y - y_0)^2}} = \left[\ln \left(y - y_0 + \sqrt{(X - x_0)^2 + (y - y_0)^2} \right) \right]_{y=0}^{y=Y} = \ln \frac{Y - y_0 + \sqrt{(X - x_0)^2 + (Y - y_0)^2}}{-y_0 + \sqrt{(X - x_0)^2 + y_0^2}}.$$

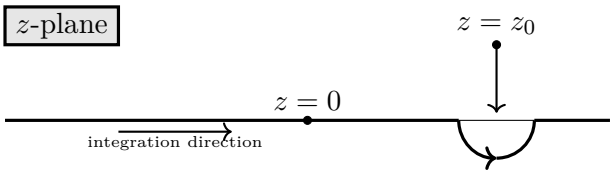


Figure 1.4: A natural way to deform the integration contour to avoid the singularity at $z = z_0$ as z_0 approaches the real axis. For this to be valid the integrand must be analytic not only in the upper half-plane but also in some neighborhood below the real axis. While z_0 is still in the upper half-plane the integrals over the original contour (that is, the real axis) and the deformed contour are the same, because during the deformation the contour did not cross any singularity. Hence, the integral over the deformed contour defines an analytic continuation of the original integral (considered as a function of z_0) to z_0 on the real axis.

itself is also Y-independent. Thus it must be actually equal to zero for all $Y > y_0$.

Altogether, we have found that for functions f that satisfy the three properties enumerated right above the equation (1.35) we can use the integral formula (1.28) to obtain

$$f(z_0) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} dx \frac{f(x)}{x - z_0}, \tag{1.40}$$

for any z_0 with positive imaginary part. This kind of equation is sometimes also called a dispersion relation, but strictly speaking there is one aspect still missing. To obtain full dispersion relations we need to move the point z_0 from the upper half-plane to the real axis. This is not trivial because in the original formula (1.28) the point z_0 must not lie on the boundary of G. However, after we resolve the ensuing issues we obtain a formula of great value. First of all, in physical applications we are often interested primarily in real values of z_0. But there is also another, less obvious utility of that formula. What is now the denominator in (1.40) will become real, and this, combined with the fact that the integration contour is the real axis, will allow us to nicely separate the formula into its real and imaginary components and it turns out that either one of those equations contains the full information on the other one.

In the equation (1.40) the point z_0 lies in the upper half-plane. If we kept the contour as it is and just moved z_0 onto the real axis, the integral on the right-hand side would become ill-defined because the denominator would vanish for $x = z_0$. A natural and very useful approach in situations such as this one is to try to deform the contour of integration away from the singularity of the integrand, as is depicted in Fig. 1.4. Unfortunately, in this

specific case we cannot do so, because we do not assume that the function $f(z)$ is analytic below the real axis.

Nevertheless, since this kind of approach is of tremendous utility in high energy physics let us now take a short detour and describe this technique a little bit better. Consider a situation such as the one depicted in Fig. 1.4 but assume for now that the function f is analytic everywhere. That means that there is only one singularity of the integrand $f(z)/(z - z_0)$ in (1.40): the simple pole at $z = z_0$. We start with z_0 in the upper half-plane, which means that the integral is regular everywhere on the real axis, the original contour of integration. Denote the value of the integral by $I(z_0)$. It is an analytic function of z_0 that is well-defined for all z_0 with positive imaginary part. For z_0 on the real axis, however, the integral (1.40) becomes ill-defined. But just because the function $I(z_0)$, as given by (1.40), is defined only for z_0 from the upper half-plane, we cannot exclude the option that $I(z_0)$ can be analytically continued into a larger region. In fact, from the figure 1.4 it is easy to see how to think about such a continuation. With z_0 still in the upper half-plane, the integral over the original contour — the real axis — and the integral over the contour depicted by the thick line have the same value. This is a consequence of the Cauchy's theorem: the difference between the two contours is the small half-disk protruding into the lower half-plane and since the integrand is analytic on that half-disk, the integral over its boundary vanishes. But the integral over this boundary equals the difference between the integral over the original contour and the integral over the deformed contour. So both the integrals give the same result while z_0 is in the upper half-plane. However, when z_0 moves onto the real line, as indicated in the figure by an arrow, the original integral becomes ill-defined, but the second integral avoids the singularity and remains perfectly finite. In this way we can explicitly define continuations of (1.40) as integrals over deformed contours. As long as we can deform the contour away from all the singularities, the integral remains finite.

There are a few ways in which the above procedure may fail. We will now illustrate one such a case. Suppose that in addition to z_0, which we now consider as a parameter, there is also another singularity at the point z_1. For simplicity, let us assume that z_1 is a function of z_0 and that $z_1 = z_0^*$. (The star indicates complex conjugation.) Hence, the singularities at z_0 and z_1 approach the real axis together, one from above the integration contour and the other one from below. This is illustrated in Fig. 1.5. As the singularities reach the real axis the integral becomes ill-defined and we would like to deform the contour in a way that avoids both of them. However, this not possible. Any deformation that avoids one of the singularities necessarily crosses the other one. (For instance, if we tried to deform the contour to avoid z_0 as we did in Fig. 1.4, we would cross the singularity at z_1.) If this happens we say that the integration contour is *pinched* by those singularities.

Pinches of integration contours are very important in the analysis of

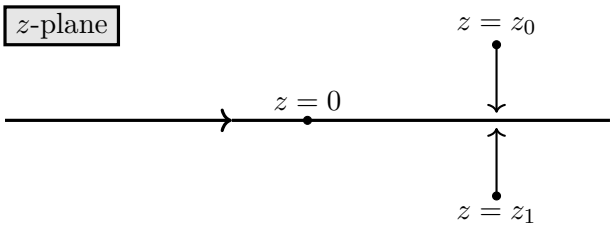


Figure 1.5: An illustration of a pinch. The singularities at z_0 and z_1 move towards each other and they meet near the integration contour, as indicated by the arrows. We would like to deform the contour away from the singularities, but there is no deformation that avoids both of them. For instance, we could try to avoid the singularity at z_0 deforming the contour as was done in Fig. 1.4, but in doing so we would cross the singularity at z_1.

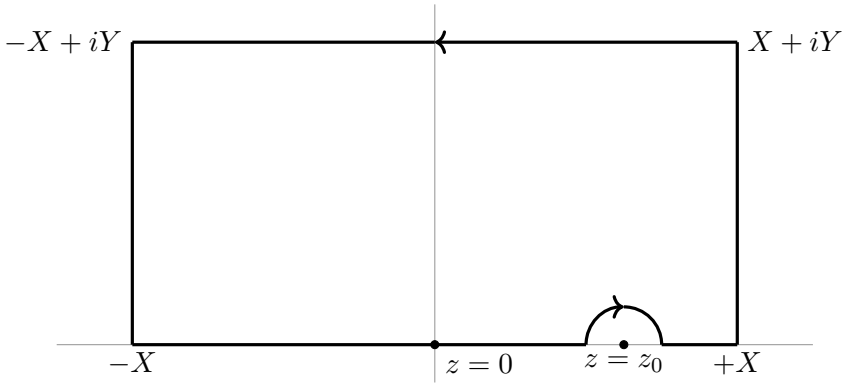


Figure 1.6: An integration contour that is suitable for the derivation of dispersion relations for square integrable functions. The values of X and Y are taken to ∞.

analytic properties of Feynman graphs and scattering amplitudes. In this context they usually give rise to branch points. One can determine possible locations of such pinches using the so-called Landau conditions. More information about these methods can be found for instance in [1], and we will briefly return to this topic later in the section 2.5. (For some issues with the treatment presented in [1] as well as in other standard references, see [6].) Another important application of these techniques is in the Libby-Sterman analysis of zero-mass singularities in Feynman graphs. This analysis identifies regions with dominant contributions in the high-energy limit, which in turn leads to factorization theorems. (For a textbook treatment of this and other aspects of perturbative QCD, see [7].)

Let us now return to our main task, which is to derive a version of (1.40) for real values of z_0. If the function f was analytic in some neighborhood below the real axis, we could just deform the contour to avoid the singular-

ity[5], but since we only assume that f is analytic in the upper half-plane we need to devise a different approach. We proceed similarly as we did in the derivation of (1.40) but instead of the contour depicted in Fig. 1.3, which is suitable only when $\Im\{z_0\} > 0$, we modify the contour to accommodate for the case of real z_0. This modified contour is depicted in Fig. 1.6. Note that this time the integral over the whole contour vanishes because the region enclosed by it contains no singularities. Now, by the same argument as before, the integrals over the two vertical sides of the rectangle vanish in the limit $\pm X \to \pm\infty$. Our previous consideration of the integral over the upper side of the rectangle also remains valid: in the limit $|X| \to \infty$ this integral vanishes (for any $Y > 0$).

Hence, altogether we find that the integral over the bottom part of the contour of integration, depicted in Fig. 1.7, is zero. We will now divide this contour into two parts and consider the corresponding integrals separately. The first part is the integral over the semicircle $S(\epsilon)$ that lies in the upper half-plane and goes from $x_0 - \epsilon$ to $x_0 + \epsilon$. When we express this integral in terms of the angular variable ϕ defined by the substitution $z = x_0 + \epsilon \exp(i\phi)$, we obtain

$$\frac{1}{2\pi i}\int_{S(\epsilon)} dz \frac{f(z_0)}{z - x_0} = \frac{1}{2\pi i}\int_{\pi}^{0} d\left(x_0 + \epsilon \mathrm{e}^{i\phi}\right) \frac{f(x_0 + \epsilon \mathrm{e}^{i\phi})}{\epsilon \mathrm{e}^{i\phi}}$$
$$= -\frac{1}{2\pi}\int_{0}^{\pi} d\phi f(x_0 + \epsilon \mathrm{e}^{i\phi}), \quad (1.41)$$

for any $\epsilon > 0$. We now take the limit $\epsilon \to 0$, which allows us to use the continuity of $f(z)$ in the upper half-plane[6] to get

$$\lim_{\epsilon\to 0^+} \frac{1}{2\pi i}\int_{S(\epsilon)} dz \frac{f(z_0)}{z - x_0} = -\frac{1}{2\pi}\lim_{\epsilon\to 0^+}\int_{0}^{\pi} d\phi f(x_0 + \epsilon \mathrm{e}^{i\phi}) = -\frac{1}{2} f(x_0). \quad (1.42)$$

[5] We could just rewrite (1.40) as

$$f(x_0) = \frac{1}{2\pi i}\lim_{\epsilon\to 0^+}\int_{-\infty}^{\infty} dx \frac{f(x)}{x - x_0 - i\epsilon},$$

and using the contour deformation as in Fig. 1.4 we would in effect obtain the distribution-theoretic identity

$$\lim_{\epsilon\to 0^+} \frac{1}{x - i\epsilon} = \mathcal{P}\left(\frac{1}{x}\right) + i\pi\delta(x),$$

which then leads to the final result

$$f(x_0) = \frac{1}{2} f(x_0) + \frac{1}{2\pi i}\mathcal{P}\int_{-\infty}^{\infty} dx \frac{f(x)}{x - x_0}.$$

We will discuss this later in this and the next section.

[6] Actually, we need f to be continuous on the closure of the upper half-plane, that is, on the union of the upper half-plane and the real axis. We did not specify this explicitly, but have assumed it throughout this section. Note that the functions specified around the equations (1.32)-(1.34) satisfy this property.

The other part of the integral is

$$\frac{1}{2\pi i}\left(\int_{-\infty}^{x_0-\epsilon} dx + \int_{x_0+\epsilon}^{+\infty} dx\right)\frac{f(x)}{x-x_0}. \tag{1.43}$$

Taking the limit $\epsilon \to 0$ then gives

$$\frac{1}{2\pi i}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{f(x)}{x-x_0}, \tag{1.44}$$

where the symbol $\mathcal{P}$ in front of the integral denotes the so-called Cauchy's principal value. This can be nicely defined in distribution-theoretic terms, but it is also appropriate to think of it simply as the limit $\epsilon \to 0^+$ of the previous equation: we integrate over the whole integration region except that we skip a small interval around the pole, and consider the limit when this interval contracts to length zero.

Combining our results we obtain

$$0 = -\frac{1}{2}f(x_0) + \frac{1}{2\pi i}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{f(x)}{x-x_0}, \tag{1.45}$$

or, in a nicer form,

$$f(x_0) = \frac{1}{i\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{f(x)}{x-x_0}. \tag{1.46}$$

We can see that both the integral measure and the denominator of the integrand are real. As promised earlier, this allows us to conveniently separate the real and imaginary part of this equation. Taking the real part yields

$$\Re\{f(x_0)\} = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{\Im\{f(x)\}}{x-x_0}. \tag{1.47}$$

Taking instead the imaginary part leads to

$$\Im\{f(x_0)\} = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{\Re\{f(x)\}}{x-x_0}. \tag{1.48}$$

The equations (1.47) and (1.48) are the dispersion relations that we were looking for. As a consequence of the factor of i in the equation (1.46) they relate the real part of f to its imaginary part.

1.5 Sokhotski-Plemelj identities

Before we continue our discussion of dispersion relations and causal transforms let us take a closer look at a useful mathematical result that we encountered.

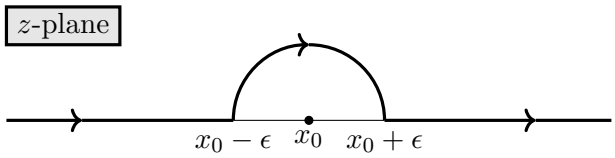


Figure 1.7: Taking the limit $|X| \to \infty$ with the contour depicted in Fig. 1.6 we find that the integral over the contour depicted here is zero. The contour starts on the real axis at $-\infty$ and goes along the axis to the point $x_0 - \epsilon$. Then it describes a semicircle of radius ϵ in the upper half-plane and returns to the real axis at the point $x_0 + \epsilon$. It continues along the real axis all the way to $+\infty$. The value of ϵ is arbitrary positive.

In the previous section we considered the integral of $f(z)/(z-z_0)$, where $f(z)$ was a function analytic in the upper half-plane (and continuous on its closure), over the contour depicted in Fig. 1.7. We showed how to decompose this integral into two parts: one which corresponds to the principal value of that integral and the other one — stemming from the integral over the small semicircle — that contributes the value $-f(x_0)/2$. We can depict this schematically as in Fig. 1.8. That picture attempts to convey the following reasoning. The leftmost diagram corresponds to the integration over the real axis, with the black dot depicting the position of the pole of the integrand just below the real axis. That is, the denominator is of the form $z-(x_0-i\epsilon)$, where ϵ is a small positive quantity. What we are really interested in is the limit $\epsilon \to 0$. Since the integrand is assumed to be analytic in the upper half-plane, we are allowed to deform the contour so that it avoids the pole at $z_0 = x_0 - i\epsilon$ as ϵ is taken to zero. This is depicted in the middle diagram. The rightmost diagram then shows the decomposition of this deformed contour into two parts, exactly as in the previous section. The first term corresponds to the principal value, while the second term contributes with half the residue multiplied by $-2\pi i$. [It is half the residue because we integrate over a semicircle rather than the full circle. We multiply by $-2\pi i$ because we integrate in the clockwise direction. In the previous section we demonstrated that this is correct by an explicit calculation.] If we interpret this result as the distribution $1/(x-(x_0-i\epsilon))$ acting on a test function $f(x)$ we obtain

$$\lim_{\epsilon\to 0^+} \frac{1}{x - x_0 + i\epsilon} = \frac{\mathcal{P}}{x - x_0} - i\pi\delta(x - x_0), \tag{1.49}$$

where we have assumed that $f(x)$ is the boundary value of a function analytic in the upper half-plane.

If, on the other hand, we assume that $f(x)$ is the boundary value of a function analytic in the lower half-plane, we may consider a schematic

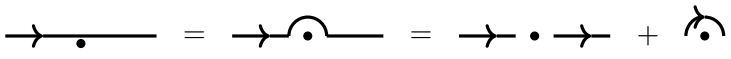


Figure 1.8: A schematic corresponding to the Sokhotski-Plemelj identity for a function analytic in the upper half-plane.

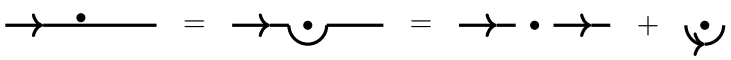


Figure 1.9: A schematic corresponding to the Sokhotski-Plemelj identity for a function analytic in the lower half-plane.

depicted in Fig. 1.9, and this corresponds to the equation

$$\lim_{\epsilon\to 0^+} \frac{1}{x - x_0 - i\epsilon} = \frac{\mathcal{P}}{x - x_0} + i\pi\delta(x - x_0). \tag{1.50}$$

The sign in front of the delta function has been reversed because now the semicircle around the pole is traversed in the counterclockwise direction. These two results are called Sokhotski-Plemelj identities and can be derived by distribution theoretic-methods without explicit restrictions on the analyticity of test functions. (See, e.g., [3].)

We close this section by providing a simple, less rigorous derivation. Assuming that both terms on the right hand side are well-defined, we can write

$$\lim_{\epsilon\to 0^+} \frac{1}{x \pm i\epsilon} = \lim_{\epsilon\to 0^+} \frac{x}{x^2 + \epsilon^2} \mp \lim_{\epsilon\to 0^+} \frac{i\epsilon}{x^2 + \epsilon^2}. \tag{1.51}$$

When considered as a function of x, the first term on the right hand side equals $1/x$ for $x \neq 0$ and 0 for $x = 0$. When considered as a distribution, it can be shown to correspond to $\mathcal{P}/x$. The second term, when considered simply as a function, is zero everywhere except for $x = 0$ where it diverges. However, for any $\epsilon > 0$ we have

$$\int dx \frac{\epsilon}{x^2 + \epsilon^2} = \int dy \frac{1}{y^2 + 1} = [\arctan(y)]_{-\infty}^{+\infty} = \pi, \tag{1.52}$$

so as a distribution the second term corresponds to $\mp i\pi\delta(x)$. Altogether, we have obtained

$$\lim_{\epsilon\to 0^+} \frac{1}{x \pm i\epsilon} = \frac{\mathcal{P}}{x} \mp i\pi\delta(x). \tag{1.53}$$

1.6 Hilbert transforms

We now return to our discussion of dispersion relations (1.47) and (1.48). In this section we will show that either one of them implies the other.

Before we proceed, however, let us first mention several other results, which we will prove useful in our investigation of dispersion relations.

Recall, first, the very useful expression for the Heaviside step function $\theta(t)$, which is defined in (1.20), in terms of its Fourier transform:

$$\theta(t) = \lim_{\epsilon \to 0^+} \frac{1}{-2\pi i} \int_{-\infty}^{+\infty} d\omega \frac{\mathrm{e}^{-i\omega t}}{\omega + i\epsilon}. \tag{1.54}$$

To prove the validity of this formula, consider separately the cases of the positive t and of the negative t. First, if $t < 0$ we can close the contour of integration using a large semicircle in the upper half-plane. In the limit of the infinitely large semicircle the integral over this semicircle vanishes. Then, as a consequence of the fact that the integrand is analytic in the upper half-plane, it follows that for $t < 0$ the integral is zero. If, on the other hand, $t > 0$, we are forced to close the contour in the lower half-plane and pick up the contribution of the residue at $\omega = -i\epsilon$. The value of the integral is therefore $-2\pi i$, which shows that the right hand side equals 1 for $t > 0$, completing our proof of (1.54). Using our earlier notation, we may write (1.54) also as

$$\theta(t) = i\mathcal{F}^{-1}\left[\frac{1}{\omega + i\epsilon}\right](t), \tag{1.55}$$

where on the right hand side it is now implicitly assumed that we are considering the limit $\epsilon \to 0^+$. (Note also that $\theta(-t) = -i\mathcal{F}^{-1}\left[\frac{1}{\omega - i\epsilon}\right](t)$.)

It will be convenient to explicitly consider also the function $\epsilon(t)$, defined by

$$\epsilon(t) = \begin{cases} -1 & \text{if } t < 0, \\ +1 & \text{if } t > 0. \end{cases} \tag{1.56}$$

Using (1.55) and the Sokhotski-Plemelj identities (1.53) we can write

$$\begin{aligned} \epsilon(t) = \theta(t) - \theta(-t) &= i\mathcal{F}^{-1}\left[\frac{1}{\omega + i\epsilon}\right](t) \\ &\quad - \left(-i\mathcal{F}^{-1}\left[\frac{1}{\omega - i\epsilon}\right](t)\right) = 2i\mathcal{F}^{-1}\left[\frac{\mathcal{P}}{\omega}\right](t). \end{aligned} \tag{1.57}$$

More explicitly, this reads

$$\epsilon(t) = \frac{i}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} d\omega \frac{\mathrm{e}^{-i\omega t}}{\omega}. \tag{1.58}$$

We can verify this equation with the help of contour integration. We consider the original contour as going from $-\infty$ to $-\epsilon$ and then from $+\epsilon$ to $+\infty$ (and eventually we take the limit $\epsilon \to 0^+$). If $t > 0$, we close the contour by two circular paths: a large semicircle in the lower half-plane, and a small semicircle from $-\epsilon$ to $+\epsilon$, also in the lower half-plane (see Fig. 1.10). The overall integral is zero, as well as the integral over the large semicircle. The integral of $\exp(-i\omega t)/\omega$ over the small semicircle, in the limit $\epsilon \to 0$, is equal

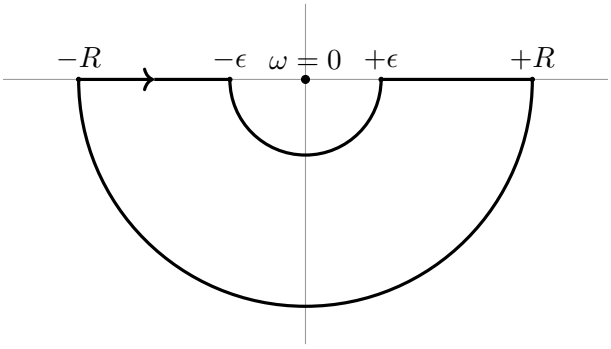


Figure 1.10: An integration contour that can be used to verify the validity of (1.58) for $t > 0$. The integral over the whole contour is zero for all $0 < \epsilon < R$. The correspondence with (1.58) is obtained in the limit $R \to \infty$ and $\epsilon \to 0^+$. In this limit the integral over the horizontal segments is equal to the integral (1.58), the integral over the large semicircle vanishes, and the integral over the small semicircle contributes $2\pi i$ times half the residue.

to πi. This means that the integral over the original contour equals $-\pi i$, and the right hand side of (1.58) is indeed $+1$ for $t > 0$. For $t < 0$ we use a similar argument, but with semicircles in the upper half-plane.

We will also need to use the well-known results regarding Fourier transforms of products and convolutions. In their core, these results state that products are transformed into convolutions and convolutions into products, but there are additional factors of 2π, which complicate this issue. In particular, they depend on the definition of the Fourier transform. We used (1.3), but some definitions contain an extra factor of $(2\pi)^{-1/2}$ or $(2\pi)^{-1}$. Let us therefore state these results explicitly. Earlier, we have already derived the result (1.25) for the Fourier transform of a convolution

$$\mathcal{F}\left[a * b\right] = \mathcal{F}\left[a\right]\ \mathcal{F}\left[b\right]. \tag{1.59}$$

This implies

$$\mathcal{F}^{-1}\left[\tilde{a}\tilde{b}\right] = \mathcal{F}^{-1}\left[\tilde{a}\right] * \mathcal{F}^{-1}\left[\tilde{b}\right], \tag{1.60}$$

where we employed our usual notation $\mathcal{F}[f(t)](\omega) = \tilde{f}(\omega)$. For the inverse Fourier transform of a convolution we find

$$\begin{aligned}\mathcal{F}^{-1}\left[\tilde{a} * \tilde{b}\right](t) &= \frac{1}{2\pi}\int d\omega\ (\tilde{a} * \tilde{b})(\omega)\mathrm{e}^{-i\omega t} \\ &= \frac{1}{2\pi}\int d\omega \int d\omega' \left[\tilde{a}(\omega-\omega')\mathrm{e}^{-i(\omega-\omega')t}\right]\left[\tilde{b}(\omega')\mathrm{e}^{-i\omega' t}\right] \\ = 2\pi\int\frac{d\omega}{2\pi}\int\frac{d\omega'}{2\pi}\left[\tilde{a}(\omega)\mathrm{e}^{-i\omega t}\right]\left[\tilde{b}(\omega')\mathrm{e}^{-i\omega' t}\right] &= 2\pi\mathcal{F}^{-1}\left[\tilde{a}\right](t)\ \mathcal{F}^{-1}\left[\tilde{b}\right](t).\end{aligned} \tag{1.61}$$

And finally, applying $\mathcal{F}$ to both sides of (1.61) we get

$$\mathcal{F}[ab] = \frac{1}{2\pi}\mathcal{F}[a] * \mathcal{F}[b]. \tag{1.62}$$

Now we are ready to turn our attention to the dispersion relations (1.47)

$$\Re\{f(x_0)\} = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{\Im\{f(x)\}}{x - x_0},$$

and (1.48)

$$\Im\{f(x_0)\} = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{\Re\{f(x)\}}{x - x_0}.$$

Notice that these equations have the form of a convolution of $\mathcal{P}/x$ with $\Re\{f(x)\}$ or $\Im\{f(x)\}$. For instance, starting from (1.47) we can write

$$\Re\{f(x_0)\} = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} dx \frac{\Im\{f(x)\}}{x_0 - x} = -\frac{1}{\pi}\left[\frac{\mathcal{P}}{x_0}\right] * [\Im\{f(x_0)\}], \tag{1.63}$$

where we have written the denominator $x - x_0$ as $-(x_0 - x)$ to obtain a form directly corresponding to convolution. Applying now the inverse Fourier transform and using (1.61) we find

$$\mathcal{F}^{-1}[\Re\{f(x_0)\}] = -\frac{1}{\pi}2\pi\mathcal{F}^{-1}\left[\frac{\mathcal{P}}{x_0}\right]\mathcal{F}^{-1}[\Im\{f(x_0)\}], \tag{1.64}$$

and, by the virtue of (1.57), this reads

$$\mathcal{F}^{-1}[\Re\{f\}](t) = i\epsilon(t)\mathcal{F}^{-1}[\Im\{f\}](t). \tag{1.65}$$

This equation is the inverse Fourier transform of the dispersion relation (1.47). Note, however, that $\epsilon(t)^2 = 1$. [We are interested only in the properties of $\epsilon(t)^2$ as a distribution, which means that we do not care about its value at $t = 0$, which is a set of measure zero.] Hence, multiplying by $\epsilon(t)$ both sides of the equation (1.65), we find that

$$\mathcal{F}^{-1}[\Im\{f\}](t) = -i\epsilon(t)\mathcal{F}^{-1}[\Re\{f\}](t). \tag{1.66}$$

This equation, in turn, is the inverse Fourier transform of the dispersion relation (1.48), as we will now show. We apply $\mathcal{F}$ to both sides of (1.66) and use (1.62) to obtain

$$\Im\{f\} = \frac{-i}{2\pi}\mathcal{F}[\epsilon] * \Re\{f\}. \tag{1.67}$$

But (1.57) implies that $\mathcal{F}[\epsilon](\omega) = 2i\mathcal{P}/\omega$, and we recover (1.48).

In a similar way, one can start from the dispersion relation (1.48) and using the same methods derive the dispersion relation (1.47). Hence, we

have shown that either one of those dispersion relations implies the other one. The proof relied on the application of the inverse Fourier transformation to those equations, leading to their counterparts (1.65) and (1.66), for which the equivalence is apparent. [One could say that we have transformed the problem from a momentum representation to a coordinate representation. This kind of a change of representation is often helpful. A result that seems puzzling in one representation may become more transparent in another.]

Notice that nowhere in our derivation of the equivalence of the dispersion relations (1.47) and (1.48) have we relied on the fact that $\Re\{f\}$ and $\Im\{f\}$ are the real and imaginary parts of the same function. The result is in fact quite general, and the equations (1.47) and (1.48) are just an instance of so-called *Hilbert transforms*. The Hilbert transform of a function $a(\omega)$ is defined as its convolution with $\mathcal{P}/(\pi\omega)$

$$\mathcal{H}\left[a\right](\omega) = \frac{1}{\pi}\mathcal{P}\int d\omega' \frac{a(\omega')}{\omega - \omega'}, \tag{1.68}$$

and, as we have shown in this section, its inverse $\mathcal{H}^{-1}$ is provided by the convolution with $-\mathcal{P}/(\pi\omega)$

$$\mathcal{H}^{-1}\left[b\right](\omega) = -\frac{1}{\pi}\mathcal{P}\int d\omega' \frac{b(\omega')}{\omega - \omega'}. \tag{1.69}$$

The dispersion relations (1.47) and (1.48) therefore state that $\Im\{f\}$ is the Hilbert transform of $\Re\{f\}$.

1.7 Titchmarsh's theorem

In the previous section we applied the inverse Fourier transform to the dispersion relations (1.47) and (1.48) to obtain the equations (1.65)

$$\mathcal{F}^{-1}\left[\Re\{f\}\right](t) = i\epsilon(t)\mathcal{F}^{-1}\left[\Im\{f\}\right](t),$$

and (1.66)

$$\mathcal{F}^{-1}\left[\Im\{f\}\right](t) = -i\epsilon(t)\mathcal{F}^{-1}\left[\Re\{f\}\right](t),$$

respectively. This allowed us to see that the two equations are in fact equivalent.

If we also use the fact that $\Re\{f\}$ and $\Im\{f\}$ are the real and imaginary parts of the same function f, we can use (1.65) and (1.66) to show that $\mathcal{F}^{-1}[f]$ satisfies the primitive causality condition (1.27). For instance, using (1.65) we may write

$$\begin{aligned}\mathcal{F}^{-1}\left[f\right](t) &= \mathcal{F}^{-1}\left[\Re\{f\}\right](t) + i\mathcal{F}^{-1}\left[\Im\{f\}\right](t) \\ &= i(\epsilon(t)+1)\,\mathcal{F}^{-1}\left[\Im\{f\}\right](t) = 2i\theta(t)\,\mathcal{F}^{-1}\left[\Im\{f\}\right](t).\end{aligned} \tag{1.70}$$

Similarly, we can use (1.66) and find

$$\mathcal{F}^{-1}\left[f\right](t) = 2\theta(t)\,\mathcal{F}^{-1}\left[\Re\{f\}\right](t). \tag{1.71}$$

In either case, we see that $\mathcal{F}^{-1}[f](t) = 0$ for $t < 0$.

Let us summarize what we have found so far in this and previous sections. We started from a square integrable function $g(t)$ that satisfied the primitive causality condition (1.27). We argued that this implied that its Fourier transform $\tilde{g}(\omega)$ was a boundary value of a function analytic in the upper half-plane and square integrable along any line parallel to the real axis lying in the upper half-plane. This, in turn, implied that $\tilde{g}(\omega)$ satisfied dispersion relations (1.47) and (1.48). And now we have shown that these dispersion relations imply that $g = \mathcal{F}^{-1}[\tilde{g}]$ satisfies the primitive causality condition. This means that we have returned full circle to the original starting point of the argument.[7]

If stated and proven more rigorously this collection of results corresponds to what is known among physicists as the Titchmarsh's theorem (see [8], pages 125–129):

Theorem 1 (Titchmarsh's theorem) *Let $\tilde{g} : \mathbb{R} \to \mathbb{R}$ be a square integrable function. The following statements are equivalent:*

1. *The inverse Fourier transform $g(t) = \mathcal{F}^{-1}[\tilde{g}](t)$ satisfies the condition*

$$g(t) = 0 \quad \textit{if} \quad t < 0.$$

2. *There exists a constant $C > 0$ and a function $G : \mathbb{C} \to \mathbb{C}$ analytic in the upper half-plane, such that*

$$\int_{-\infty}^{+\infty} d\omega\, |G(\omega + i\eta)|^2 < C, \quad \textit{for any} \quad \eta > 0,$$

and G satisfies

$$\lim_{\eta \to 0^+} G(\omega + i\eta) = \tilde{g}(\omega)$$

for almost all ω.

3. *The function $\tilde{g}$ satisfies*

$$\Re\{\tilde{g}(\omega)\} = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\Im\{\tilde{g}(\omega)\}}{\omega' - \omega}.$$

[7] To connect this to our brief discussion of Hilbert transforms, these results also show that for any square integrable real function $a(x)$, the complex function $f(x) = a(x) + i\mathcal{H}[a](x)$ is the boundary value of a function $f(z)$ analytic in the upper half-plane and square integrable along any line parallel to the real axis (in the upper half-plane).

4. The function $\tilde{g}$ satisfies

$$\Im\{\tilde{g}(\omega)\} = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\Re\{\tilde{g}(\omega)\}}{\omega' - \omega}.$$

In other words, during our investigation of square integrable functions of a single real variable we learned that for such functions the primitive causality condition (1.27)

$$g(t) = 0 \text{ if } t < 0,$$

is equivalent to the property that the Fourier transform $\tilde{g}(\omega) = \mathcal{F}[g](\omega)$ is the boundary value of a function analytic in the upper half-plane that satisfies certain boundedness properties. Furthermore, it is also equivalent to $\tilde{g}(\omega)$ satisfying either one of the dispersion relations (1.47) and (1.48).

This concludes our introduction of the link between the causality and analyticity. The restriction to square integrable functions of a single variable is, however, quite significant and it is not at all straightforward to connect the results that have been presented so far to applications in non-relativistic quantum mechanics and quantum field theory. We will concern ourselves with those issues in the next chapter. The rest of the present chapter will instead be devoted to a brief discussion of several disconnected topics in the elementary dispersion theory.

1.8 Propagation of light in a medium

In this section we shall remain in the context of classical physics and briefly discuss some aspects of the propagation of light in a dielectric medium. It was in this context that the first known dispersion relation was recognized and it would be shame if we did not at least mention it here. There are also other reasons to consider this case. First of all, some of the terminology traces its origin to the propagation of light in a medium. Second, we will encounter a simple generalization of the primitive causality condition (1.27) and a common technique that will allow us to use Titchmarsh's theorem to obtain analyticity of $\tilde{g}$ even if that function may not be square integrable. And lastly, we will be able to present a very intuitive argument for why the condition of causality demands that the real and imaginary parts of $\tilde{g}$ be not independent.

A full treatment of this topic, although entirely within our reach, would take us too far from our main line of interest. For this reason I will just state the parts relevant to our exposition and skip most of the details and derivations. A proper textbook treatment can be found in sections 1.4, 1.5, 1.9 and 1.10 of [3], the notation of which we are going to more or less follow. (A very condensed review of this topic can be also found in [4].)

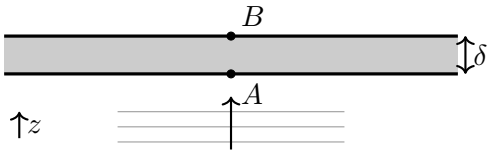


Figure 1.11: A thick layer of a dielectric material is positioned horizontally. The thickness of the slab is δ. A monochromatic plane wave travels along the z-axis vertically and approaches the material form below. We measure the incident electric field E_{in} at the point A and the outgoing electric field E_{out} at the point B.

Consider a monochromatic plane wave propagating in the z-direction. If the light propagates in the vacuum the electric field satisfies

$$E(z,t) = E(0,0)\mathrm{e}^{i(kz-\omega t)} = E(0,0)\mathrm{e}^{i\omega(z/c-t)}.$$

In this section we can interpret E as either corresponding to the vector of the electric field or to any of its components. As usually, the symbols k and ω represent the angular wave number and the angular frequency, respectively. In this formalism, in order to ensure that the field is real, we would need to actually consider frequencies ω and $-\omega$ always together in a superposition $E(0,0)\mathrm{e}^{i\omega(z/c-t)} + E(0,0)^*\mathrm{e}^{i(-\omega)(z/c-t)}$, but to simplify notation we will usually consider complex fields here.

To describe the propagation in a dielectric medium we need to introduce the *complex refractive index* n' and write

$$E(z,t) = E(0,0)\mathrm{e}^{i\omega\left(z\frac{n'}{c}-t\right)}. \tag{1.72}$$

The real part of n' is the *real refractive index* n that determines the phase velocity of the wave. The imaginary part corresponds to the absorption and scattering of the wave. If we define

$$n' = n + i\frac{c\beta}{2\omega}, \tag{1.73}$$

the equation (1.72) becomes

$$E(z,t) = E(0,0)\mathrm{e}^{i\omega\left(z\frac{n}{c}-t\right)}\mathrm{e}^{-\frac{\beta z}{2}}. \tag{1.74}$$

The quantity β is therefore called the *extinction coefficient*. (The intensity decreases in proportion to $\exp(-\beta z)$.)

Consider now the following arrangement (see Fig. 1.11). A thin layer of a dielectric medium is placed perpendicularly to the z-axis. Let us denote its width by δ. The layer extends from $z = 0$ to $z = \delta$ and the plane wave travels in the empty space in the z-direction from the region of negative z

until it reaches the front face of the slab. Denote the value of the electric field at $z = 0$ and time t by $E_{\rm in}(t)$. The plane wave then propagates through the medium and exits the slab at $z = \delta$. Let us denote the value of the electric field at $z = \delta$ and time t by $E_{\rm out}(t)$. In the terminology of the section 1.2 $E_{\rm in}$ is the input and $E_{\rm out}(t)$ the output of the system. The primitive causality condition (1.27) would then correspond to us demanding that if $E_{\rm in}(t) = 0$ for $t < 0$ then also $E_{\rm out}(t) = 0$ for $t < 0$. (Note that the system is time translation invariant, so we are free to single out the time $t = 0$ in the statement of the causality condition.) However, the present system is an example of when the primitive causality condition is not fully appropriate. No signal can travel from one side of the layer to the other with velocity greater than the speed of light c. A stronger condition therefore applies: $E_{\rm in}(t) = 0$ for $t < 0$ implies that $E_{\rm out}(t) = 0$ for $t < \delta/c$. If we write

$$E_{\rm out}(t) = \int_{-\infty}^{+\infty} dt' g(t - t') E_{\rm in}(t'), \tag{1.75}$$

this condition reads

$$g(t) = 0 \quad \text{for} \quad t < \delta/c. \tag{1.76}$$

A condition of this kind, that no signal can propagate faster than the speed of light, is sometimes called the *relativistic causality condition*.

Let us now introduce a simple method that is often very helpful. When we stated the Titchmarsh's theorem the notation indicated that the function whose properties the theorem determined was a Green's function g. That is certainly how we often want to use it, but the theorem is applicable to any square integrable function that satisfies the causality condition. In the present case of the propagation of light we have not provided any argument for why the function g of (1.75) should be square integrable. We only know that it satisfies (1.76) and therefore also the (weaker) primitive causality condition (1.27). This is, however, already enough for us to be able to employ the Titchmarsh's theorem and say something useful about the analyticity properties of g.

The argument goes as follows. Although we do not know if g is square integrable, we certainly can arrange for the incoming wave to be such that $E_{\rm in}(t)$ is square integrable, and that $E_{\rm in}(t) = 0$ for $t < 0$.[8] We can also use the conservation of energy to argue that

$$\int_{-\infty}^{+\infty} \mathbf{E}_{\rm out}(t)^2 \leq \int_{-\infty}^{+\infty} \mathbf{E}_{\rm in}(t)^2,$$

[8] Strictly speaking, for a monochromatic plane wave we cannot arrange either of those things. A plane wave is not localizable in space or time, and $|E_{\rm in}(t)| = \text{const}$ for such a wave. Hence, neither the condition of causality nor the condition of square integrability can be satisfied by such $E_{\rm in}(t)$. What we really have in mind, however, is a suitable superposition of plane waves: a wave packet that occupies a restricted region in space and time, and which can satisfy both the conditions. We rely on the linearity of the system when passing back and forth between the consideration of monochromatic plane waves and that of localizable wave packets.

which means that if all the components of $\mathbf{E}_{\rm in}$ are square integrable so are those of $\mathbf{E}_{\rm out}$. Furthermore, the causality condition implies that also $E_{\rm out}(t) = 0$ for $t < 0$. Hence, both $E_{\rm in}$ and $E_{\rm out}$ satisfy the conditions of Titchmarsh's theorem and both are therefore boundary values of functions analytic in the upper half-plane. Now, the equation (1.75) implies

$$\tilde{E}_{\rm out}(\omega) = \tilde{g}(\omega)\tilde{E}_{\rm in}(\omega), \tag{1.77}$$

for ω real. This allows us to define $\tilde{g}$ in the upper half-plane as $\tilde{E}_{\rm out}/\tilde{E}_{\rm in}$, where $\tilde{E}_{\rm out}$, $\tilde{E}_{\rm in}$ now stand for the continuations of those functions into the upper half-plane. One can see that $\tilde{g}$ so defined is analytic in the upper half-plane except possibly for the points where $\tilde{E}_{\rm in}$ has zeros. [Note, however, that we may consider various different functions $\tilde{E}_{\rm in}$ so in general the possibility of such poles can be excluded.]

Let us now describe the relation between $E_{\rm in}(t)$ and $E_{\rm out}(t)$ in terms of $n'(\omega)$. The propagation of light equation (1.72) implies that for a monochromatic wave we have

$$E_{\rm out}(t) = E_{\rm in}(t)\mathrm{e}^{i\omega n'\frac{\delta}{c}}, \tag{1.78}$$

which means that $\tilde{g}$ of (1.77) is related to n' by[9]

$$\tilde{g}(\omega) = \mathrm{e}^{i\omega n'\frac{\delta}{c}}, \tag{1.79}$$

where n' is a function of ω. The relativistic causality condition (1.76) then implies that

$$\tilde{g}(\omega) = \mathrm{e}^{i\omega n'\frac{\delta}{c}} = \int_{\delta/c}^{\infty} dt\; g(t)\mathrm{e}^{i\omega t}, \tag{1.80}$$

or, in other words,

$$\mathrm{e}^{i\omega(n'(\omega)-1)\frac{\delta}{c}} = \int_{0}^{\infty} dt\; g(t+\delta/c)\mathrm{e}^{i\omega t}. \tag{1.81}$$

We already know that the left-hand side is analytic in the upper half-plane of ω, but this equation allows us to proceed even further. If the medium is an insulator, it is possible to show (see [3]) that $n'(\omega) - 1$ is a causal transform. We can therefore formulate dispersion relations (1.47) and (1.48) for $n' - 1$. In particular, using $\Im\{n'\} = c\beta/2\omega$, we may write (1.47) as

$$n - 1 = \frac{c}{2\pi}\mathcal{P}\int_{-\infty}^{+\infty} \frac{d\omega'}{\omega'}\frac{\beta(\omega')}{\omega' - \omega}. \tag{1.82}$$

[9]This may seem slightly confusing. We formulated the equation of the propagation of light (1.72) for monochromatic waves only, which means that $E_{\rm in}(t)$ and $E_{\rm out}(t)$ of that equation are inverse Fourier transforms of functions $\tilde{E}_{\rm in}(\omega')$ and $\tilde{E}_{\rm out}(\omega')$ which are proportional to the delta function $\delta(\omega' - \omega)$. That explains why $\tilde{g}(\omega)$ can be extracted from (1.78) directly as the ratio $E_{\rm out}(t)/E_{\rm in}(t)$.

If we further use the reality condition $n'(-\omega) = n'(\omega)^*$, which implies that $\beta(-\omega) = \beta(\omega)$, we can write this dispersion relation in the form

$$n - 1 = \frac{c}{\pi}\mathcal{P}\int_0^\infty d\omega' \, \frac{\beta(\omega')}{\omega'^2 - \omega^2}. \tag{1.83}$$

This equation is known as the *Kramers–Kronig* relation, after Ralph Kronig and Hendrik A. Kramers, who derived this equation independently in 1926 and 1927. Historically, it was the first known dispersion relation and the name 'dispersion relations' originates from this context. Furthermore, the real part of $\tilde{g}$ is often called the *refractive* part, and the imaginary part is called the *absorptive* part, in reference to the physical interpretation of the real and imaginary parts of n'.[10]

We close this section by introducing an intuitive argument for why causality implies that there must be some relation between the imaginary and real parts of $n'(\omega)$. For the argument's sake, suppose that the refractive and absorptive part of $n' - 1$ are completely independent and imagine that we could design a material that has the real refractive index n equal to 1 for all ω, and the extinction coefficient $\beta(\omega)$ such that the material is fully transparent for almost all ω except for a very small interval of frequencies near some ω_0, where β sharply becomes very large. That is, suppose that the extinction coefficient corresponds roughly to a delta function, $\beta(\omega) \approx \delta(\omega - \omega_0)$. Such a material would let nearly all frequencies pass untouched, except for essentially an infinitesimal interval of frequencies near ω_0, which would instead be completely removed from the incoming wave.

The causality requires that if a localized wave packet of light travels towards the material, the outgoing wave does not appear until the incident wave packet reaches the material: $E_{\text{out}}(t) = 0$ at least as long as $E_{\text{in}}(t) = 0$. As we will now explain, the material described in the previous paragraph would violate this condition. A monochromatic plane wave of angular frequency ω moving in the z-direction has electric field $E(z,t) = E_\omega \exp\{i\omega(z/c - t)\}$ where E_ω is a complex constant. At this point it will be better to consider real plane waves, in which case we consider a suitable superposition of plane waves of angular frequencies ω and $-\omega$ and obtain

$$E(z,t) = A \, \sin\left(\omega(z/c - t) + \alpha\right), \tag{1.84}$$

where A and α are now some real constants. Therefore, we cannot localize such a wave: it is almost everywhere almost all the time. To obtain a wave packet that is localized in some restricted region we must consider a suitable superposition of such plane waves that cancel each other everywhere

[10] The term absorptive part is nevertheless very fitting in many applications in quantum physics, where $\tilde{g}$ often corresponds to the scattering amplitude. The imaginary part of the scattering amplitude is related to the total cross section by the optical theorem — both in quantum scattering theory and in optics.

except for that restricted region. Now, if the incident wave is such a localized wave packet, the field $E_{\rm in}(t)$, which is measured at the front face of the slab, remains zero until the wave packet arrives. The outgoing wave, measured at the other side of the slab is determined by applying (1.78) to each monochromatic component of the incoming wave separately. But in the case of our imagined material this would correspond to leaving the incoming wave unchanged except for removing frequencies that are infinitesimally near to ω_0. This roughly corresponds to removing a plane wave of the form (1.84) at the frequency ω_0 from that wave packet, and that is the same as adding minus that monochromatic wave to the incoming wave packet. But since the incoming wave packet is localized, once we add a monochromatic wave to it the resulting superposition cannot be localized: everywhere except for the region to which the original wave packet was restricted the resulting superposition is of the form (1.84). The outgoing wave would manifestly violate causality.

We can therefore see why dispersion relations, such as the Kramers–Kronig relation (1.83), are necessary. If we remove or suppress some components in a localized wave packet — which is governed by the function $\beta(\omega)$ — we must adjust the phases of the remaining components – this is governed by $n(\omega)$ — to ensure a suitable cancellation in distant regions and make the resulting wave packet localized.

1.9 Subtractions

There is an important technical issue that needs to be discussed, and that is the topic of subtracted dispersion relations. As we have already mentioned on several occasions, not every analytic function satisfies dispersion relations. In order for us to be able to write down dispersion relations for a function, in addition to being analytic that function must also satisfy certain asymptotic properties. This was suggested already by the formulation of the Titchmarsh's theorem where the requirement for $\tilde{g}$ to be a causal transform was not only that $\tilde{g}(\omega + i\eta)$ be an analytic function of $\omega + i\eta$, but also that when considered as a function of ω, the function $\tilde{g}(\omega + i\eta)$ must be square integrable for any fixed value of $\eta > 0$. Furthermore, we also saw that there must exist a common constant C that provides an upper bound for integrals $\int d\omega |\tilde{g}(\omega + i\eta)|^2$ for all values of $\eta > 0$.

Although the Titchmarsh's theorem does not correspond to the most general situation that we need to consider, asymptotic properties of $\tilde{g}$ continue to play an important role also in those more general circumstances. One often encounters the situation in which $\tilde{g}$ is a function of several complex variables, is analytic in a certain region, and can be bounded by a polynomial. To formulate for it some kind of a dispersion relation we need to make sure that the part of the integral that goes over the distant boundary van-

ishes. Because of that, we cannot directly formulate dispersion relation for $\tilde{g}$ of polynomial growth, and must proceed in a more subtle way by considering a suitable function that is related to $\tilde{g}$ but has appropriate asymptotic properties. In fact, we have already encountered an example of this when we stated the Kramers–Kronig relation (1.83), which was formulated for $n' - 1$ instead of n'. This is because for large values $|\omega|$ the complex refractive index n' approaches the value 1, which means that n' is not square integrable. To obtain a causal transform we needed to subtract from it the constant 1. This illustrates the general case: to obtain dispersion relations we might need to carry out steps that result in so-called *subtractions*. As far as I know, this is an entirely technical issue, without deeper physical or mathematical consequences, so I am not going to discuss it in any great detail. Instead, I will simply present two different methods of how can one think about subtractions. In both cases I will limit my consideration to a function f of a single complex variable z that is $\mathcal{O}(|z|^n)$ for large $|z|$ and some integer $n \geq 0$.[11]

Let us now describe the first approach. Suppose that $f(z) = \mathcal{O}(|z|^n)$ for some $n \geq 0$. Then, for any polynomial $P(z)$ of order $n + 1$ or higher the function $f(z)/P(z)$ is $\mathcal{O}(|z|^{-1})$. This introduces new poles so we usually do not want to apply the Titchmarsh's theorem for square integrable functions. Instead, we can obtain dispersion relations by simply closing the integration contour using a large semicircle in the upper half-plane or by some other suitable path. The integrand $f(z)/[(z - z_0)P(z)]$ is now $\mathcal{O}(|z|^{-2})$, so the integral over the distant boundary vanishes. For our applications we usually need $P(z)$ to be real on the real axis and to not have any zeros there. This implies that zeros of P are placed symmetrically with respect to complex conjugation, both above and below the real axis. The integral over the whole closed contour does in general not vanish, because the integrand is not analytic in the enclosed region, which normally contains both the pole at $z = z_0$ and some of the poles that are located at zeros of $P(z)$. For each zero of $P(z)$ that is inside the integration contour we pick up a contribution from its residue. Note that if a_i is a simple zero of $P(z)$ then we can write

$$\frac{P(z)}{z - a_i} = \left.\frac{dP(z)}{dz}\right|_{z=a_i}.$$

Hence, if all the zeros of $P(z)$ are simple then the residue theorem says that the integral of $f(z)/[(z - z_0)P(z)]$ over the contour closed by an infinitely

[11] Recall that in this context $f(z) = \mathcal{O}(g(z))$ means that the modulus of f grows at most as fast as that of g for large $|z|$. More precisely, it means that there exists a constant C such that

$$\frac{|f(z)|}{|g(z)|} < C,$$

for all large enough $|z|$.

large semicircle in the upper half-plane equals

$$2\pi i \left(\frac{f(z_0)}{P(z_0)} + \sum_{a_i=1}^{k} \frac{f(a_i)}{(a_i - z_0)\frac{dP(z)}{dz}\big|_{z=a_i}} \right), \tag{1.85}$$

where we assumed that z_0 is in the upper half-plane (otherwise the first term would be absent) and where we denoted the zeros of $P(z)$ in the upper half-plane by a_1, ..., a_k. The sum over the zeros of $P(z)$ corresponds to the subtractions. Note that each subtraction depends on the value of $f(z)$ at some zero a_i of $P(z)$. We will see a simple case of this method later in the section 5.5, where we will discuss the spectral representation of the photon propagator.

The other method that we are going to consider in this section is much less general but avoids the explicit introduction of new poles and can be used in conjunction with the Titchmarsh's theorem. It is, however, applicable only if the boundary value $f(x)$ of $f(z)$ on the real axis is sufficiently differentiable.

Let us consider the case when $f(z) = \mathcal{O}(1)$ for large $|z|$. Then we cannot assume that the function is square integrable. We can, however, work with the function $f(z)/(z - x_0)$, for some real constant x_0. This function has already suitable asymptotic properties, but now the problem is its pole at $z = x_0$. To deal with this issue, we can consider yet another function

$$f_1(z) = \frac{f(z) - f(x_0)}{z - x_0}, \tag{1.86}$$

which is regular at $z = x_0$ where it assumes the value $f_1(x_0) = df(x_0)/dz$. The function $f_1(x)$ is therefore square integrable. Unfortunately, that does not automatically mean that $f(z)$ is a causal transform, but putting in some more work[12] one can show that when the function f corresponds to a Green's function satisfying the primitive causality condition then f_1 in fact really is a causal transform. We can then formulate dispersion relations for f_1, and from those one can obtain the dispersion relation for f of the form

$$f(x) = f(x_0) + \frac{x - x_0}{\pi i} \mathcal{P} \int_{-\infty}^{+\infty} \frac{dx'}{x' - x} \frac{f(x') - f(x_0)}{x' - x_0}. \tag{1.87}$$

The first term on the right hand side is the subtraction term. In this case it is the value of $f(x)$ at $x = x_0$. Accordingly, this kind of dispersion relation is called *once subtracted* dispersion relation. If the function f had worse asymptotic properties than $\mathcal{O}(1)$ one could proceed similarly to regularize the function $f(z)/(z - x_0)^n$ for some suitably large n and obtain dispersion

[12] It is not completely straightforward to obtain these results. One can find the detailed argument in the section 1.7 of [3].

relations with n subtractions. Details of this procedure can be found for instance in the section 1.7 of [3].

In general, each subtraction adds one free parameter to the dispersion relation: we need to provide the value of the function f or its derivatives at some fixed point. For this reason it is often best to use dispersion relations with the lowest possible number of subtractions. Nevertheless, there are some cases when one might prefer to proceed differently. As we add subtractions, we trade the number of free parameters for better asymptotic properties. Sometimes this is an advantageous trade-off. For instance, in phenomenological applications when high energy data are imprecise or missing it may be helpful to use the extra powers of $1/(\omega - \omega_0)$ in the dispersion integral to suppress contributions from higher energies — which correspond to high values of ω — even if that means that one must supply values of f at some additional low energy subtraction points.

Chapter 2

Analyticity in quantum theory

We will now proceed to discuss the relation between causality and analyticity in quantum mechanics and quantum field theory. This is a large subject, which is not even completely understood, and we will not attempt to address it fully. Instead, our aim will only be to present a rough overview of the topic, and provide some basic understanding and hints of how the material that has been presented so far applies to the cases of non-relativistic quantum mechanics and relativistic quantum field theory. We will also introduce some of the foundational results that will be needed later.

2.1 Quantum scattering and causality

The paradigmatic setup in which the relation between causality and analyticity is studied in quantum theory is that of a scattering system. Let us illustrate this on a simple example provided by the scattering of a classical scalar field $\phi(\mathbf{x}, t)$ by a spherically symmetric potential $V(r)$. We consider the case of a *finite-range* potential. This means that we assume that there is some finite radius $a > 0$ beyond which the potential vanishes,

$$V(r) = 0 \quad \text{for} \quad r > a. \tag{2.1}$$

This is sometimes described by saying that the interaction is *local*, because the field 'feels' the scatterer only when it is less than distance a away from its

center.[1] We suppose that for $r > 0$ the field satisfies the free wave equation

$$\ddot{\phi} = v^2 \nabla^2 \phi, \tag{2.2}$$

with v standing for the velocity of the field propagation. To make use of the spherical symmetry of the potential we decompose the solution into spherical harmonics. The simplest such a component is of course the s-wave component, which is proportional to the spherical harmonic of angular momentum 0. This solution is therefore spherically symmetric, $\phi(\mathbf{x}, t) = \phi(r, t)$. Recalling that in spherical coordinates the Laplacian is

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2} r + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \tag{2.3}$$

we see that the free wave equation (2.2) applied to an s-wave reads

$$\left[\frac{\partial}{v\partial t}\right]^2 [r\phi(r,t)] = \left[\frac{\partial}{\partial r}\right]^2 [r\phi(r,t)]\,. \tag{2.4}$$

Outside the scatterer, for $r > a$, the general s-wave solution is therefore

$$\phi(r,t) = \frac{1}{r}\int_{-\infty}^{+\infty} dk \left[A(k)\mathrm{e}^{-ik(r+vt)} + B(k)\mathrm{e}^{ik(r-vt)}\right]. \tag{2.5}$$

Here the first term corresponds to the incoming wave and the second term to the outgoing wave.[2] When focusing our attention only the region $r > a$ the coefficient functions $A(k)$ and $B(k)$ might seem to be independent. In general, however, they are not, because the solution (2.5) of the equation (2.2) applies only in the region $r > a$ and must be connected to an appropriate solution for $r < a$. The solution for $r < a$ satisfies the analog of the free wave equation (2.2) for non-zero potential. Furthermore, it must be finite at $r = 0$, and this constraint is inherited by the continuation of the solution into the region $r > a$ in the form of a relation between the coefficient functions of the incoming and outgoing waves in (2.5).

[1] Treatment of non-local interactions is far more complicated. If the potential is not restricted to a finite radius but decreases sufficiently fast, as do for example Yukawa-like potentials with $V(r) \propto r^{-1}\exp(-mr)$, the usual machinery of scattering theory still applies, even though some qualitatively new features may appear in the solutions. Such potentials are sometimes also referred to as being local. On the other hand, once we include genuine long-range forces, for which $V(r) \propto r^{-1}$, some parts of the usual scattering theory begin to break down. A common symptom of these issues is the appearance of infrared divergences. In fact, if long-range forces are present it is not even clear how to define asymptotic states and the S-matrix.

[2] Some comments for the reader's convenience: The signs of $\mp ik$ in the exponents are just a convention, and were chosen such that the time evolution factor is in both terms $\exp(-i\omega t)$, where $\omega = kv$. As for which wave is incoming and which outgoing, we can see directly from the factors $r \pm vt$ in the exponents that the phase velocities are $\mp v$. Furthermore, the group velocity $d\omega/dk$ is in this case the same as the phase velocity ω/k.

Since we are dealing with the second order differential equation, to sew the solution for $r < a$ to the solution (2.5) for $r > a$ we must make sure that the function ϕ and its first derivatives are continuous everywhere. Note that we can do this for each Fourier component (each value of k) separately. This should therefore result in a relation between $A(k)$ and $B(k)$, inherited from the region $r < a$.

Let us provide a simple example. If there is no scattering at all then $V(r) = 0$ for all r and the free wave equation (2.2) holds in the whole space. The solution (2.5) therefore applies all the way to $r = 0$, where it must be finite. Because of the factor $1/r$ in front of the integral this translates to the condition that the expression inside the square brackets in (2.5) vanishes at $r = 0$. Thus, we find that $B(k) = -A(k)$. If V was not identically zero the relation between $A(k)$ and $B(k)$ would become more complicated, but because of the linearity we would still expect it to be of the form

$$B(k) = -S(k)A(k). \tag{2.6}$$

The function $S(k)$ defined by the equation (2.6) corresponds to the *scattering matrix*, or the *S-matrix*. (In our simple case of an elastic s-wave scattering it is just a function.) Note the explicit minus sign on the right hand side of the definition (2.6). The reason that we put it there is that we want the S-matrix to be identically 1 if there is no scattering.

We can now turn our attention to the question of how to connect all of this to our earlier findings regarding the link between causality and analyticity. Recall our earlier considerations of linear causal systems, for which we identified the input $f(t)$, the Green's function $g(t)$, and the output $x(t)$, and how those functions were related by a convolution (see (1.24)). The Fourier transform of that convolution then led to the equation (1.26)

$$\tilde{x}(\omega) = \tilde{g}(\omega)\tilde{f}(\omega).$$

In our classical s-wave scattering problem we instead had the incoming and the outgoing waves and their Fourier transforms with respect to time were basically[3] the coefficient functions $A(k)$ and $B(k)$. Hence, there is an analogy between the coefficient functions $A(k)$ and $B(k)$ of the incoming and outgoing waves, and the Fourier transforms $\tilde{f}(\omega)$ and $\tilde{x}(\omega)$ of the input and the output. Furthermore, our present definition of the S-matrix, the equation (2.6), is of the same form as the equation (1.26) that we quoted above.

[3]More precisely, if we denote by ϕ_{in} the incoming wave $r^{-1}\int_{-\infty}^{+\infty} dk\ A(k)\mathrm{e}^{-ik(r+vt)}$ and by ϕ_{out} the outgoing wave $r^{-1}\int_{-\infty}^{+\infty} dk\ B(k)\mathrm{e}^{ik(r-vt)}$, we obtain

$$\mathcal{F}_t\left[\phi_{\text{in}}\right](\omega) = \frac{2\pi}{r}A(\omega/v)\mathrm{e}^{-i\frac{\omega}{v}r},$$
$$\mathcal{F}_t\left[\phi_{\text{out}}\right](\omega) = \frac{2\pi}{r}B(\omega/v)\mathrm{e}^{+i\frac{\omega}{v}r}.$$

This implies that the S-matrix corresponds roughly to the Fourier transform of the Green's function. Of course, the correspondence just outlined is not exact and one needs to be careful when deriving precise results, but the analogy is correct. In particular, we can formulate a causality condition in a similar fashion as we did before, although we need to make a minor modification. Suppose that we are measuring the wave at some fixed radius $r_0 > a$. A direct application of the primitive causality condition (1.27) would be to demand that the outgoing field at r_0 is zero as long as the incoming field is zero. In fact, a stronger condition applies. After the incoming field reaches the radius r_0, it must still travel (with velocity v) at least to the radius $r = a$ until it can start being scattered by the potential, and the outgoing wave must then travel from $r = a$ back to $r = r_0$ in order for us to measure it. Hence, the appropriate condition is to demand that if the incoming field appeared only at $t = t_0$ then the outgoing field is zero at least until the time $t_0 + 2(r_0 - a)/v$. It is possible to show that this condition implies that the function $S(k)\exp(2ika)$ must be analytic and bounded in the upper half-plane. (For details, see [3, 4].)

Now that we have seen an indication of how to relate a classical scattering system to our earlier findings, let us turn our attention to a simple example in non-relativistic quantum mechanics. We will consider the scattering of a single spinless particle in a local potential $V(\mathbf{x})$. Instead of the classical wave equation, the wave function $\psi(\mathbf{x}, t)$ now solves the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{x}, t) = \left(-\frac{\hbar^2}{2\mu}\nabla^2 + V(\mathbf{x})\right)\psi(\mathbf{x}, t). \tag{2.7}$$

Here μ is either the mass of the particle if we consider a scattering in an external potential $V(\mathbf{x})$, or the reduced mass $\mu = m_1 m_2/(m_1 + m_2)$ if we consider a system of two particles of masses m_1 and m_2. In the latter case the variable $\mathbf{x}$ corresponds to the relative separation of the two particles.

As is well known, one of the important differences between the classical wave equation (2.2) and the Schrödinger equation (2.7) is that the classical wave equation is a differential equation of second order in time, while the Schrödinger equation involves only the first time derivative of the wave function. Consequences of this difference turn out to be important also for our purposes.

Let us illustrate this again on the case of the s-wave scattering in a finite-range spherically symmetric potential. The full solution is again a function of only the radius r and time t, but as a consequence of there being only the first time derivative in the equation (2.7) the vacuum solution is now

$$\psi(r, t) = \frac{1}{r}\int_{-\infty}^{+\infty} dp \left[A(p)\mathrm{e}^{\frac{i}{\hbar}(-pr-Et)} + B(p)\mathrm{e}^{\frac{i}{\hbar}(pr-Et)}\right], \tag{2.8}$$

where p is the momentum, related to the wave number k by $p = \hbar k$, and

where the energy E equals

$$E = \frac{p^2}{2\mu}. \tag{2.9}$$

As a consequence, the energy is non-negative, $E \geq 0$. This is very different from the classical solution (2.5) which contained components of both positive and negative angular frequency ω.

The fact that the solution (2.8) contains only positive frequencies makes it much more difficult to formulate a causality condition. The primitive causality condition (1.27), or its straightforward modifications such as the relativistic causality condition (1.76) or the condition that we formulated earlier in this section for the scattering of a classical field, are all very intuitive but they are meaningful only if it is possible to consider situations in which the input $f(t)$ is identically zero until some time t_0. But in quantum mechanics that is impossible to achieve. If the incoming wave vanishes during any interval of time then it is zero always. This is a consequence of there being only positive frequencies in the solutions of Schrödinger equation. At a fixed radius r the incoming wave of the solution (2.8) is of the form

$$f(t;r) = \psi_{\rm in}(r,t) = \int_0^\infty dE\; K(E)\, {\rm e}^{-iEt/\hbar}, \tag{2.10}$$

where

$$K(E) = \frac{A{\rm e}^{-ipr}}{r}\frac{dp}{dE} = \frac{\mu}{p(E)} A\left(p(E)\right) {\rm e}^{-ip(E)r/\hbar}, \tag{2.11}$$

and $p(E) = \sqrt{2\mu E}$. The fact that the integral in (2.10) is only over positive values of E means that $f(t;r)$ is a boundary value of a function $F(t)$ analytic in the lower half-plane.[4] But an analytic function that vanishes over any interval is zero everywhere, and this translates to the same property for the boundary value $f(t;r)$.

To summarize, we found that in the context of quantum mechanics it is impossible to use any straightforward variation of the primitive causality condition. The incoming wave is almost all the time non-zero near the scatterer. (However, this does not mean that the wave packet cannot be localized. The incoming wave far away from the center of the wave packet must only be non-zero, but can be very small, as for instance being suppressed exponentially with the distance from the wave packet center.) This is an important observation about quantum theory, but we will not explore it here any further. Let us just mention that in order to obtain analyticity properties people have suggested and used various formulations of causality. Some of these are described in detail in [3]. One of them is the *Schützer–Tiomno causality condition*, which requires that the outgoing wave at a radius r_0 and time t_0 cannot depend on the incoming wave at the same radius r_0 and

[4] Mathematical results concerning properties such as this one will be stated more carefully in the section 2.3.

later times $t > t_0$. Perhaps better known is the *Van Kampen's causality condition* which states that if the incoming wave packet in the limit of $t \to -\infty$ represents a single particle, then the total probability of finding the particle outside the scatterer (that is, for $r > a$) can never exceed 1. Another important approach is the usage of the *Wigner's causal inequality* regarding *time delays*, which are related to *phase shifts*. (We won't discuss these concepts here. In the context of causality they are treated, for instance, in [4, 3].)

There is one last issue that I would like to mention before we close this brief discussion of scattering in non-relativistic quantum mechanics. In this section we have been considering only s-wave scattering. A general scattering process, however, is not spherically symmetric. Even if we limit our attention only to spherical potentials $V(\mathbf{x}) = V(r)$, we still need to be able to treat incoming waves that contain components of higher angular momentum. Once we include higher angular momenta, we are in effect considering the angular dependence of the scattering. In particular, if the incoming wave is a plane wave it contains all integer values $l \geq 0$ of angular momentum. For an incoming plane wave traveling in, say, the z-direction the scattering amplitude[5] $f(p, \theta)$ is a function of both the magnitude of the three-momentum p and the scattering angle θ. We have seen that causality implies certain analytic properties with respect to the variable p, but causality on its own does not allow us to say anything about the analyticity in the variable θ. Nevertheless, it can often be shown that the scattering amplitude $f(p, \theta)$ indeed is analytic in a certain region, but this kind of analyticity follows from the locality of interactions rather than from causality. In particular, for finite-range potentials it is possible to show that partial wave amplitudes decrease asymptotically for large values of the angular momentum l as $\exp(-2l \log l)$, and this implies the analyticity of $f(p, \theta)$ as a function of $\cos\theta$ in the whole complex plane.[6] For Yukawa-like potentials the region of analyticity is restricted to an ellipse. For further details the reader is again referred to [4] or [3]. Note that in quantum field theory this distinction between the analyticity in the energy being related to causality and the analyticity in the scattering angle to locality becomes somewhat blurred as

[5] The scattering amplitude is closely related to the S-matrix, and is usually defined by the equation

$$\psi^p_{\text{out}}(\mathbf{x}) \to f(p, \theta) \frac{\mathrm{e}^{ipr}}{r}$$

for $|x| \to \infty$, possibly up to some constant factors. Here ψ^p_{out} is the outgoing wave corresponding to the incoming plane wave of momentum p in the z-direction, with the time dependence stripped away (or just evaluated at $t = 0$).

[6] This follows from the Neumann's theorem, which states that if $f(z) = \sum_{l=0}^{\infty} c_l P_l(z)$, where P_l are the Legendre polynomials, and if asymptotically the coefficients c_l decay as a^{-l} or faster (for some $a > 1$) then the function $f(z)$ is analytic within the ellipse centered at the origin and with the major semi-axis oriented along the real axis and of the length $(a^2 + 1)/2a$ and the minor semi-axis of the length $(a^2 - 1)/2a$. For finite-range potentials, $c_l \propto l^{-2l}$, so $f(z)$ is analytic everywhere.

a result of crossing symmetry.

In quantum field theory it becomes even less clear how to formulate a condition that would correctly capture the physical property of causality. In fact, people encountered difficulties trying to formulate and use generalizations of the primitive causality condition (1.27) that would be applicable in relativistic settings. Instead, a somewhat different approach turned out to be useful. We postulate that

$$[\phi_l(x), \phi_m(y)]_\pm = 0 \quad \text{for } x - y \text{ spacelike}, \tag{2.12}$$

where ϕ_l and ϕ_m are any local fields,[7] and $[A, B]_\pm$ corresponds to the commutator $AB - BA$ if either one of A or B is bosonic and the anti-commutator $AB + BA$ if they are both fermionic. The condition (2.12) is called the condition of *microscopic causality* or *microcausality*. Sometimes it is also referred to as *local commutativity*, especially in the context of axiomatic field theory. Roughly speaking, this condition expresses the idea that no signal propagates faster than with the speed of light. Of course, this condition is just an assumption, or a postulate, and it is not clear that it must hold, especially for very small separations of x and y.

To get some idea of how to obtain analytic properties from the condition of microscopic causality (2.12), consider matrix elements of time-ordered products of field operators, such as

$$(\Psi_B, \mathrm{T}\{\phi_1(x_1) \dots \phi_N(x_N)\} \Psi_A). \tag{2.13}$$

Matrix elements of this kind figure prominently in quantum field theory, and we will discuss two important contexts in which they appear in the next section.

The operation $\mathrm{T}\{\cdots\}$ that appears in (2.13) denotes the time ordering of operators, which is defined by

$$\begin{aligned} &\mathrm{T}\{O_1(x_1) O_2(x_2) \dots O_N(x_N)\} \\ &\qquad = \sum_\pi \theta\left(x^0_{\pi(1)} - x^0_{\pi(2)}\right) \dots \theta\left(x^0_{\pi(N-1)} - x^0_{\pi(N)}\right) \\ &\qquad\qquad \times O_{\pi(1)}\left(x_{\pi(1)}\right) O_{\pi(2)}\left(x_{\pi(2)}\right) \dots O_{\pi(N)}\left(x_{\pi(N)}\right), \end{aligned} \tag{2.14}$$

where the sum is over all permutations π of the indices $1, \dots, N$, and x^0_j is the time component of the four-vector x_j. For instance, if $N = 2$ then (2.14) reads

$$\begin{aligned} \mathrm{T}\{O_1(x_1) O_2(x_2)\} &= \theta\left(x^0_1 - x^0_2\right) O_1(x_1) O_2(x_2) \\ &\quad + \theta\left(x^0_2 - x^0_1\right) O_2(x_2) O_1(x_1). \end{aligned} \tag{2.15}$$

[7] These are operator fields, or, more precisely, operator-valued distributions.

In the definition (2.14) and the example (2.15) we assumed that operators $O_j(x)$ were all bosonic. Fermionic operators satisfy anti-commutation rather than commutation relations, and if they are present, we include a factor of -1 for each interchange of two fermionic operators. In particular, if all the operators O_1, ..., O_N are fermionic, then instead of (2.14) we obtain

$$\begin{aligned}\mathrm{T}\left\{O_1\left(x_1\right) O_2\left(x_2\right) \ldots O_N\left(x_N\right)\right\} \\ = \sum_\pi \operatorname{sgn}(\pi)\Big[\theta\left(x^0_{\pi(1)} - x^0_{\pi(2)}\right) \ldots \theta\left(x^0_{\pi(N-1)} - x^0_{\pi(N)}\right) \\ \times O_{\pi(1)}\left(x_{\pi(1)}\right)\ O_{\pi(2)}\left(x_{\pi(2)}\right) \ldots O_{\pi(N)}\left(x_{\pi(N)}\right)\Big], \quad (2.16)\end{aligned}$$

where $\operatorname{sgn}(\pi) = \pm 1$ is the sign of the permutation π. For instance, when O_1 and O_2 are both fermionic we get

$$\begin{aligned}\mathrm{T}\left\{O_1\left(x_1\right) O_2\left(x_2\right)\right\} = \theta\left(x^0_1 - x^0_2\right) O_1\left(x_1\right) O_2\left(x_2\right) \\ - \theta\left(x^0_2 - x^0_1\right) O_2\left(x_2\right) O_1\left(x_1\right). \quad (2.17)\end{aligned}$$

Now we are ready to indicate how to obtain analytic properties from the microscopic causality condition (2.12). We will limit our attention to the simple case of (2.15), but with both the operators O_1 and O_2 standing for a bosonic field ϕ. To proceed, one needs to re-write the time ordered product $\mathrm{T}\{\phi(x)\phi(y)\}$ in terms of the commutator $[\phi(x), \phi(y)]$ and explicit time-ordering step functions in either one of the following two forms

$$\begin{aligned}\mathrm{T}\{\phi(x)\phi(y)\} = \phi(x)\phi(y) - \theta(y^0 - x^0)[\phi(x), \phi(y)]_- \\ = \phi(y)\phi(x) + \theta(x^0 - y^0)[\phi(x), \phi(y)]_-. \quad (2.18)\end{aligned}$$

We will now focus on the terms containing the commutator.[8] As a consequence of the condition of microscopic causality (2.12), these commutator terms are identically zero outside of the null cone of $x - y$. They also contain explicit time-ordering step functions, which restrict their support even further. The end result is that the term $\theta(y^0 - x^0)[\phi(x), \phi(y)]_-$ must be zero unless $x - y$ is past-pointing and either null or time-like. Similarly, $\theta(x^0 - y^0)[\phi(x), \phi(y)]_-$ must be zero unless $x - y$ is future-pointing and either null or time-like. This directly translates into certain analytic properties of the momentum-space representation of the corresponding matrix elements. We will discuss the mathematics of this in Section 2.3.

[8] The fixed-order terms must be dealt with in some other way. We will see an example of this much later, in the section 5.5, where we discuss the spectral representation of the photon propagator. Note also that as we will see in Section 2.3, vacuum expectation values of fixed-order terms possess certain analytic properties already as functions of x and y.

2.2 Perturbation theory for the S-matrix and time-ordered Green's functions

In the previous section we suggested a strategy for proving analytic properties of momentum-space Green's functions from the condition of microscopic causality (2.12). This strategy can be applied to matrix elements of time-ordered products of coordinate-space fields, that is, to matrix elements of the form (2.13). Such matrix elements play a prominent role in quantum field theory.

This is actually a good place to take a short detour and introduce two important contexts in which matrix elements of this kind appear. One of those is the interaction picture Dyson series for S-matrix elements. This series consists of terms whose form is similar to (2.13), with Ψ_A and Ψ_B representing the appropriate free-particle states, but where the operators $\phi_j(x)$ all stand for the interaction part of the Hamiltonian expressed in the interaction picture. They depend only on the time component of x. This is a well-known topic and our presentation will be very concise. We will introduce some of the concepts and notation. We will also mention some problems of that approach. This will lead us to the second topic that we wish to discuss, which is the topic of time-ordered Green's functions, and of their perturbation series as given by the Gell-Mann–Low formula. I will discuss this second topic in a greater detail because most textbooks do not cover it appropriately.

2.2.1 Dyson series for the S-matrix

The Dyson series for the S-matrix is formulated in the interaction picture. This picture is based on the decomposition of the Hamiltonian H into the free Hamiltonian H_0 and the interaction V,

$$H = H_0 + V. \tag{2.19}$$

Interaction-picture operators, including free fields, evolve according to the free Hamiltonian,

$$A_I(t) = \mathrm{e}^{iH_0 t} A(0) \mathrm{e}^{-iH_0 t}, \tag{2.20}$$

where we used the subscript I to indicate that $A_I(t)$ is the interaction-picture operator. This is different from the time evolution of operators in the Heisenberg picture, which is governed by the equation that is of the same form but with H_0 replaced with the full Hamiltonian H,

$$A(t) = \mathrm{e}^{iHt} A(0) \mathrm{e}^{-iHt}. \tag{2.21}$$

The goal is to calculate (at least approximately) the S-matrix, whose element $S_{\beta\alpha}$ is defined as the scalar product

$$S_{\beta\alpha} = \left(\Psi_\beta^-, \Psi_\alpha^+\right) \tag{2.22}$$

of the asymptotic "out" state Ψ^-_β containing particles denoted by the multi-index β and the asymptotic "in" state Ψ^+_α containing particles denoted by the multi-index α. These asymptotic states are eigenstates of the full Hamiltonian H and are supposed to appear as states containing separate non-interacting particles when observed at a distant past ($t \to -\infty$, this is for the "in" states) or at a distant future ($t \to +\infty$, this is for the "out" states). We label these states according to their particle content, which is discernible either at $t \to -\infty$ or at $t \to +\infty$. For each such a state, the multi-indices therefore determine the number of particles, and their types, spin, and three-momenta — all those quantities as they would be observed at the appropriate asymptotic time. For instance, if the state vector Ψ^+_α corresponds to a physical state that for $t \to -\infty$ appears as containing two particles, then $\alpha = \mathbf{p}_1, \sigma_1, n_1; \mathbf{p}_2, \sigma_2, n_2$, where $\mathbf{p}_1$ is the three-momentum of the first particle, σ_1 stands for its spin z-component or its helicity, and n_1 denotes the type of that particle, all as observed at $t \to -\infty$. Notice that n_1 determines also the total spin and the mass of the particle. The labels $\mathbf{p}_2$, σ_2, and n_2 carry the same kind of information but with respect to the particle 2.

In order to obtain the Dyson series for the S-matrix it is convenient to first define the S-matrix operator S by

$$S_{\beta\alpha} = (\Phi_\beta, S\Phi_\alpha)\,, \tag{2.23}$$

where Φ_β and Φ_α are the corresponding free-particle states. Free-particle states are eigenstates of the free Hamiltonian H_0, and our approach assumes that in the limit $t \to -\infty$ the free-particle state Φ_α appears to be the same as the "in" state Ψ^+_α, and that in the limit $t \to +\infty$ the free-particle state Φ_β appears to be the same as the "out" state Ψ^-_β. In this statement the free-particle states are meant to be evolved using the free-particle Hamiltonian H_0 and asymptotic "in" and "out" states are meant to be evolved using the full Hamiltonian H. Note also that none of these states, when considered on their own, are localized in space or time. So for the above asymptotic conditions to make any sense one needs to consider suitable wave packets.

One then defines the evolution operator $U(t_2, t_1)$ as

$$U(t_2, t_1) = \mathrm{e}^{iH_0 t_2}\mathrm{e}^{-iH(t_2-t_1)}\mathrm{e}^{-iH_0 t_1}, \tag{2.24}$$

and observes — as we will see later, somewhat incorrectly — that the S-operator is the limit of $U(t_2, t_1)$ for $t_1 \to -\infty$ and $t_2 \to +\infty$. The evolution operator U satisfies the differential equation

$$\frac{d}{dt}U(t, t_0) = -iV_I(t)U(t, t_0), \tag{2.25}$$

and the initial condition $U(t_0, t_0) = 1$. From this one can obtain the integral equation

$$U(t_2, t_1) = U(t_1, t_1) + \int_{t_1}^{t_2} dt\, \frac{dU(t, t_1)}{dt} = 1 - i\int_{t_1}^{t_2} dt\, V_I(t)U(t, t_1). \tag{2.26}$$

This gives an implicit solution for $U(t_2, t_1)$ — a solution that is expressed in terms of an integral of $U(t, t_1)$ at times $t < t_2$. One can express the operator $U(t, t_1)$ on the right-hand side in terms of this solution, which leads to the equation

$$U(t_2, t_1) = 1 - i \int_{t_1}^{t_2} dt\; V_I(t) + (-i)^2 \int_{t_1}^{t_2} dt\; V_I(t) \int_{t_1}^{t} dt'\; V_I(t') U(t', t_1). \tag{2.27}$$

By iterating this procedure one arrives at the infinite series

$$\begin{aligned} U(t_2, t_1) = 1 + (-i) \int_{t_1}^{t_2} dt\; V_I(t) + (-i)^2 \int_{t_1}^{t_2} dt\; \int_{t_1}^{t} dt'\; V_I(t) V_I(t') \\ + (-i)^3 \int_{t_1}^{t_2} dt\; \int_{t_1}^{t} dt'\; \int_{t_1}^{t'} dt''\; V_I(t) V_I(t') V_I(t'') + \dots . \end{aligned} \tag{2.28}$$

Notice that in this series the interaction operators are always time-ordered. For instance, in the term of the third order in V_I the time values t, t', and t'' satisfy $t_2 \geq t \geq t' \geq t'' \geq t_1$. Recalling that the interaction V is a bosonic operator, we can rewrite this time-ordered series using the time-ordered product operator in the following manner

$$\begin{aligned} U(t_2, t_1) = 1 + (-i) \int_{t_1}^{t_2} dt\; \mathrm{T}\left\{V_I(t)\right\} + \frac{(-i)^2}{2!} \int_{t_1}^{t_2} dt\; \int_{t_1}^{t_2} dt'\; \mathrm{T}\left\{V_I(t) V_I(t')\right\} \\ + \frac{(-i)^3}{3!} \int_{t_1}^{t_2} dt\; \int_{t_1}^{t_2} dt'\; \int_{t_1}^{t_2} dt''\; \mathrm{T}\left\{V_I(t) V_I(t') V_I(t'')\right\} + \dots , \end{aligned} \tag{2.29}$$

where all the time variables are now integrated from t_1 to t_2. This extension of the integration region is where the combinatorial factors $1/N!$ come from. To see this, let us for the sake of simplicity focus only on the term of the second order in V_I and consider two values t_a, t_b of time, satisfying $t_2 > t_b > t_a > t_1$. Then, in the time-ordered integral $\int_{t_1}^{t_2} dt\; \int_{t_1}^{t} dt'\; V_I(t) V_I(t')$ these values appear only once, when the integrand is $V_I(t_b) V_I(t_a)$. In the integral $\int_{t_1}^{t_2} dt\; \int_{t_1}^{t_2} dt'\; \mathrm{T}\left\{V_I(t) V_I(t')\right\}$, however, they appear twice: when the integrand is $\mathrm{T}\left\{V_I(t_b) V_I(t_a)\right\}$ and when it is $\mathrm{T}\left\{V_I(t_a) V_I(t_b)\right\}$.

We can now rewrite the series (2.29) as

$$U(t_2, t_1) = \mathrm{T}\left\{\exp\left(\int_{t_1}^{t_2} dt\; V_I(t)\right)\right\}, \tag{2.30}$$

and if we also take the limit $t_1 \to -\infty$ and $t_2 \to +\infty$ (assuming that it exists) we obtain the Dyson series for the S-matrix operator

$$S = \mathrm{T}\left\{\exp\left(\int_{-\infty}^{+\infty} dt\; V_I(t)\right)\right\}. \tag{2.31}$$

In order to obtain manifestly Lorentz-covariant S-matrix or Green's functions, we usually work with Hamiltonians of the form

$$H(t) = \int d^3\mathbf{x}\; \mathcal{H}(\mathbf{x}, t), \tag{2.32}$$

where $\mathcal{H}(\mathbf{x}, t)$ is the Hamiltonian density. The interaction density $\mathcal{V}(x)$ is defined similarly, and we can write

$$\mathcal{V}(x) = \mathcal{H}(x) - \mathcal{H}_0(x), \tag{2.33}$$

where $\mathcal{H}_0(x)$ is the density of the free Hamiltonian. The Dyson series for the S-matrix operator then reads

$$S = \mathrm{T}\left\{\exp\left(\int d^4x\ \mathcal{V}_I(x)\right)\right\}, \tag{2.34}$$

or, more explicitly,

$$\begin{aligned} S = 1 + (-i)\int d^4x_1\ \mathcal{V}_I(x_1) + \frac{(-i)^2}{2!}\int d^4x_1\ \int d^4x_2\ \mathrm{T}\left\{\mathcal{V}_I(x_1)\mathcal{V}_I(x_2)\right\} \\ + \frac{(-i)^3}{3!}\int d^4x_1\ \int d^4x_2\ \int d^4x_3\ \mathrm{T}\left\{\mathcal{V}_I(x_1)\mathcal{V}_I(x_2)\mathcal{V}_I(x_3)\right\} + \ldots. \end{aligned} \tag{2.35}$$

Instead of integrating only over the time variable, which of course depends on the frame of reference, as we did in (2.31), we have now obtained a series in which we integrate with the Lorentz-invariant measure d^4x over the whole space-time. We can see that this expression is manifestly Lorentz covariant, provided that:

- The interaction density is a Lorentz scalar.
- For $x_i - x_j$ space-like, $[\mathcal{V}_I(x_i), \mathcal{V}_I(x_j)] = 0$. (This is because for space-like separations the time-ordering of events is frame-specific. But if the interaction density operators commute for such separations, this frame-specificity of the time-ordered product has no consequences and we obtain a Lorentz invariant result. Please also note that while this commutativity condition is sufficient for the S-operator to be Lorentz invariant, it is not necessary.)

Here we have encountered another context in which one wishes to impose the condition of microscopic causality (2.12). Let us also note that the first condition, that the interaction density is a Lorentz scalar, is quite problematic and does not always apply.[9] This is related to the fact that the Hamiltonian formalism singles out a specific time coordinate and the canonical quantization often leads to non-covariant terms in the Hamiltonian, even if one starts with a Lorentz invariant Lagrangian density. Nevertheless, in the context of the canonical formalism it is possible to show that if the Lagrangian density is Lorentz invariant, then the S-matrix must be Lorentz-covariant. Effects

[9] In fact, in some circumstances the interaction density is not even a local function of the fields. A well-known example of this is the Coulomb term that appears in the canonical quantization of classical electrodynamics in the Coulomb gauge.

of those terms in the interaction density that are not Lorentz scalars are exactly canceled by non-covariance of propagators. It is therefore usually possible to drop all these non-covariant terms, both in the interaction and in propagators, and formulate the corresponding Feynman rules in terms of covariant quantities. (Another — and usually much more convenient — way to show the Lorentz-covariance of the S-matrix and formulate covariant Feynman rules is to use path-integral methods.) A further discussion of both the derivation of the Dyson series for the S-matrix and of the issues of Lorentz covariance in the canonical formalism can be found, for instance, in chapters 3, 6 and 7 of [9].[10]

To calculate S-matrix elements one needs to sandwich the S-matrix operator in between two free-particle states. The series (2.35) then yields

$$S_{\beta\alpha} = \delta(\beta-\alpha) + (-i)\int d^4x_1 \ (\Phi_\beta, \mathcal{V}_I(x_1)\Phi_\alpha) + \frac{(-i)^2}{2!}\int d^4x_1 \int d^4x_2 \ (\Phi_\beta, \mathrm{T}\left\{\mathcal{V}_I(x_1)\mathcal{V}_I(x_2)\right\}\Phi_\alpha) + \dots . \quad (2.36)$$

Recall that the multi-indices α and β contain information on numbers and types of the particles, and their spin and three-momenta components. The delta function $\delta(\beta-\alpha)$ is accordingly an abbreviation for a product of three-momentum-conservation delta functions and Kronecker delta symbols corresponding to discrete labels in the multi-indices. (In fact, it represents a particular sum of such products for all possible permutations of indistinguishable particles.) As we can see, each term of (2.36) does indeed contain a factor of the form (2.13).

2.2.2 Gell-Mann–Low formula

Although the derivation of the perturbation series for the S-matrix presented above may look convincing, and although it is satisfactory in non-relativistic quantum mechanics, when one attempts to apply it in the context of relativistic quantum field theory serious problems appear.[11] One issue is presented by the Haag's theorem [12], which for our purposes can be roughly stated as saying that if a relativistic quantum field theory has a well-defined interaction picture, then it must be a theory of non-interacting free fields. In other words, as Streater and Wightman [5] put it, the interaction picture

[10] I should mention that the treatment of the perturbation theory derived in the operator formalism for the S-matrix that was just presented here, as well as the treatment of the same topics in [9], are flawed, and although the discussion in [9] is very readable and informative, it should be not considered as entirely correct. The book [9] does not discuss those problems at all. Its treatment of the topics related to the canonical formalism, on the other hand, is in my opinion excellent.

[11] I base these remarks mainly on my reading of an unfinished version of [10], which provides a discussion of many of these issues, but has not been published yet. Another place where discussion of some of these issues can be found is [11].

exists only if there is no interaction. Another issue — which appears to be quite independent of the complications associated with the Haag's theorem — is that the asymptotic states of a full theory never really cease to "feel" the interaction. Particles in those states, even when being far apart from each other, always continue to self-interact. It is therefore not possible to say that such particles, in some asymptotic regime, appear as if they were free, at least not in the straightforward way that works in non-relativistic quantum mechanics. In fact, as a consequence of this self-interaction, each such a particle represents a superposition of free-particle states of various numbers of particles. Hence, to define a valid correspondence between the asymptotic states and the free-particle states is not nearly as simple as presented in many introductory treatments of the perturbation theory for S-matrix, such as the one that we sketched above.[12]

It seems that one way to avoid these problems is to focus not directly on the S-matrix but to instead consider time-ordered Green's functions

$$(\Psi_0, \mathrm{T}\{\phi_1(x_1)\dots\phi_N(x_N)\}\Psi_0)\,, \tag{2.37}$$

where the fields $\phi_j(x)$ are Heisenberg-picture operators in the full theory. The state vector Ψ_0 corresponds to the true vacuum and is normalized to unity

$$(\Psi_0, \Psi_0) = 1. \tag{2.38}$$

If we knew how to evaluate or approximate the vacuum expectation values (2.37), we could then obtain matrix elements between any asymptotic states — and, in particular, also the S-matrix — by the application of the *Lehmann–Symanzik–Zimmermann reduction formula* (LSZ formula). (The original reference for the LSZ formula is [13], and it is presented in many textbooks of quantum field theory, including [14, 9]. A modern treatment of this topic, including a discussion of some limitations of the usual approaches, can be found in [11].)

It turns out that the vacuum expectation values (2.37) can be calculated

[12] There is one more issue that probably should be mentioned at this point, and that is that it is not even clear how to *define* an S-matrix if there are massless particles present in the theory.

perturbatively according to the *Gell-Mann–Low formula*[13] (GML formula)

$$(\Psi_0, \mathrm{T}\,\{\phi_1(x_1)\dots\phi_N(x_N)\}\,\Psi_0) = \lim \frac{\left(\Psi_0^{\mathrm{free}}, \mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int d^4x\,\mathcal{V}_I(x)}\right\}\Psi_0^{\mathrm{free}}\right)}{\left(\Psi_0^{\mathrm{free}}, \mathrm{T}\left\{\mathrm{e}^{-i\int d^4x\,\mathcal{V}_I(x)}\right\}\Psi_0^{\mathrm{free}}\right)}. \quad (2.39)$$

This remarkable formula deserves some explanation. First, notice that while on the left-hand side there are Heisenberg-picture fields ϕ_j, which evolve according to the full Hamiltonian as in (2.21), and the vacuum Ψ_0 is the true vacuum state, on the right-hand side only corresponding free-field or interaction-picture quantities appear. The fields ϕ_I are free fields that evolve according to the interaction-picture evolution equation (2.20), and the state Ψ_0^{free} is the free-particle vacuum. In the exponential there is the integral over the whole space-time of the interaction Hamiltonian density expressed in the interaction picture. Neither the numerator, not the denominator of the expression on the right-hand side are well-defined on their own. Instead, we need to consider their ratio in a certain limit, in which their respective singularities cancel out. (This limit will be specified below.)

Since this topic is not adequately treated in many of the standard textbooks, let us now present a derivation of the GML formula (2.39). To make sense of the factors on its right-hand side we first put the system in a box of a finite volume and restrict its time evolution to a finite interval of time that begins at some initial time T_+ and ends at some final final time T_-. The size of the box and the time interval must be chosen such that all the coordinate variables x_1, ..., x_n fall inside that region of space-time. The numerator factor is now

$$\left(\Psi_0^{\mathrm{free,box}}, \mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\int_{\mathrm{box}}d^3\mathbf{x}\,\mathcal{V}_I(x)}\right\}\Psi_0^{\mathrm{free,box}}\right). \quad (2.40)$$

Our next goal will be to express this amplitude in terms of Heisenberg-picture quantities. To make the notation more compact, let us write

$$\int_{\mathrm{box}} d^3\mathbf{x}\,\mathcal{V}_I(\mathbf{x},t) = V_I(t).$$

[13] I do not know the original reference. If one is looking for a textbook presentation, the formula is discussed, for instance, in [14]. That book also presents a derivation of that formula, but with some important issues ignored. Some aspects of the GML formula and their derivation are also explained in [9], where the formula itself is used quite extensively (especially in Chapter 9 in the context of path-integral methods), but where it is never properly introduced. The presentation given here is based mainly on the unpublished text [10] by J. Collins.

The time-ordered product in (2.40) then reads

$$\begin{aligned}
&\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\}\\
&=\sum_{K=0}^{\infty}\frac{(-i)^K}{K!}\int_{T_+}^{T_-}dt_1\cdots\int_{T_+}^{T_-}dt_K\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)V_I(t_1)\dots V_I(t_K)\right\}
\end{aligned}\tag{2.41}$$

Let us denote by u_1, ..., u_N the time values x_1^0,..., x_N^0 but ordered so as to satisfy $u_1 \geq \cdots \geq u_N$. Recall also that the values T_- and T_+ were chosen such that $T_- > u_1$ and $u_N > T_+$. Consider now a single term in the series (2.41); in other words, consider a fixed value of K. Each of the variables of integration t_1, ..., t_K, as we integrate it between T_+ and T_-, passes through each of the intervals (T_+, u_N), (u_N, u_{N-1}), ..., (u_1, T_-). At each point of the integration over these variables, there will be some number $K_0 \geq 0$ of them that fall into the interval (u_1, T_-), some number $K_1 \geq 0$ of them that fall into the interval (u_2, u_1), etc., and these non-negative integers K_0, ..., K_N satisfy the condition $K_0 + \cdots + K_N = K$. (We ignore the set of measure zero when some of the variables t_1, ..., t_K appear in the boundary of any of the intervals.) Furthermore, as the variables t_1, ..., t_K are integrated from T_+ to T_-, all possible values of K_0, ..., K_N (that is, all non-negative values that add up to K) will be covered. Hence, for a given value of K, we can separate the integral into several terms, one for each permissible list of values K_0, ..., K_N. In each such a term we integrate only over that subset of the total integration volume in which K_0 of the integration variables fall into the interval (u_1, T_-) , K_1 of them fall into (u_2, u_1), etc. For each such a term we rename the variables as $t_l^{(m)}$ where m denotes the interval to which those variables belong and l distinguishes between the variables inside a single interval. In the interval m there are K_m variables, which means that l can assume values $l = 1, \dots, K_m$. Changing to these new variables results in a new combinatorial factor, because there are $K!/(K_0!K_1!\dots K_N!)$ ways in which we can assign K variables into $N+1$ groups, such that there are K_0 of them in the first group, K_1 of them in the second group, etc. Hence, we can write

$$\begin{aligned}
&\int_{T_+}^{T_-}dt_1\cdots\int_{T_+}^{T_-}dt_K\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)V_I(t_1)\dots V_I(t_K)\right\}\\
&=\sum_{K_0+\cdots+K_N=K}\frac{K!}{K_0!\dots K_N!}\Big[\int_{u_1}^{T_-}dt_1^{(0)}\cdots\int_{u_1}^{T_-}dt_{K_0}^{(0)}\mathrm{T}\left\{V_I(t_1^{(0)})\dots V_I(t_{K_0}^{(0)})\right\}\\
&\quad\times\phi_I(u_1)\int_{u_2}^{u_1}dt_1^{(1)}\cdots\int_{u_2}^{u_1}dt_{K_1}^{(1)}\mathrm{T}\left\{V_I(t_1^{(1)})\dots V_I(t_{K_1}^{(1)})\right\}\phi_I(u_2)\dots\\
&\quad\times\phi_I(u_N)\int_{T_+}^{u_N}dt_1^{(N)}\cdots\int_{T_+}^{u_N}dt_{K_N}^{(N)}\mathrm{T}\left\{V_I(t_1^{(N)})\dots V_I(t_{K_N}^{(N)})\right\}\Big],
\end{aligned}\tag{2.42}$$

where the sum is over all permissible lists of K_0, ..., K_N. (For simplicity, we also dropped the index label from the field operators.) $K!$ from the combinatorial factor in (2.42) cancels with the $1/K!$ factor in (2.41), and we can therefore write

$$\begin{aligned}
\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\} = \sum_{K=0}^{\infty}\sum_{K_0+\dots+K_N=K}\Bigg[\\
\frac{(-i)^{K_0}}{K_0!}\int_{u_1}^{T_-}dt_1^{(0)}\cdots\int_{u_1}^{T_-}dt_{K_0}^{(0)}\mathrm{T}\left\{V_I(t_1^{(0)})\dots V_I(t_{K_0}^{(0)})\right\}\phi_I(u_1) \\
\times\frac{(-i)^{K_1}}{K_1!}\int_{u_2}^{u_1}dt_1^{(1)}\cdots\int_{u_2}^{u_1}dt_{K_1}^{(1)}\mathrm{T}\left\{V_I(t_1^{(1)})\dots V_I(t_{K_1}^{(1)})\right\}\phi_I(u_2) \\
\times\dots\phi_I(u_N)\frac{(-i)^{K_N}}{K_N!}\int_{T_+}^{u_N}dt_1^{(N)}\cdots\int_{T_+}^{u_N}dt_{K_N}^{(N)}\mathrm{T}\left\{V_I(t_1^{(N)})\dots V_I(t_{K_N}^{(N)})\right\}\Bigg].
\end{aligned} \tag{2.43}$$

Notice now that instead of summing over all values of K from 0 to ∞ and then for each such K summing over all non-negative values of K_0, ..., K_N satisfying the constraint $K_0+\dots+K_N=K$, we can just as well sum over each K_0, ..., K_N independently from 0 to ∞. This results in

$$\begin{aligned}
\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\} = \\
\Bigg[\sum_{K_0=0}^{\infty}\frac{(-i)^{K_0}}{K_0!}\int_{u_1}^{T_-}dt_1^{(0)}\cdots\int_{u_1}^{T_-}dt_{K_0}^{(0)}\mathrm{T}\left\{V_I(t_1^{(0)})\dots V_I(t_{K_0}^{(0)})\right\}\Bigg]\phi_I(u_1) \\
\times\Bigg[\sum_{K_1=0}^{\infty}\frac{(-i)^{K_1}}{K_1!}\int_{u_2}^{u_1}dt_1^{(1)}\cdots\int_{u_2}^{u_1}dt_{K_1}^{(1)}\mathrm{T}\left\{V_I(t_1^{(1)})\dots V_I(t_{K_1}^{(1)})\right\}\Bigg]\phi_I(u_2) \\
\times\dots\phi_I(u_N)\Bigg[\sum_{K_N=0}^{\infty}\frac{(-i)^{K_N}}{K_N!}\int_{T_+}^{u_N}dt_1^{(N)}\cdots\int_{T_+}^{u_N}dt_{K_N}^{(N)}\mathrm{T}\left\{V_I(t_1^{(N)})\dots V_I(t_{K_N}^{(N)})\right\}\Bigg].
\end{aligned} \tag{2.44}$$

The sums in the square brackets of (2.44) correspond to the series (2.30) for the evolution operators between the endpoints of their respective intervals. We can therefore write

$$\begin{aligned}
\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\} \\
= U(T_-,u_1)\phi_I(u_1)U(u_1,u_2)\phi_I(u_2)\dots\phi_I(u_N)U(u_N,T_+).
\end{aligned} \tag{2.45}$$

We will now show that the right-hand side of (2.45) is nearly the same as the product of the corresponding Heisenberg-picture operators. To see this, notice first that using the equation (2.20) for the time evolution of operators

in the interaction picture, and the equation (2.21) for the time evolution in the Heisenberg-picture, one can find the following relation between those two pictures

$$A(t) = \mathrm{e}^{iHt}\mathrm{e}^{-iH_0t}A_I(t)\mathrm{e}^{iH_0t}\mathrm{e}^{-iHt} = \Omega(t)A_I(t)\Omega(t)^\dagger, \tag{2.46}$$

where $\Omega(t)^\dagger$ stands for the Hermitian conjugate of $\Omega(t)$. The operator $\Omega(t_1)$ is a unitary operator, and when applied to some Heisenberg-picture state vector (that is, the Schrödinger-picture state vector evaluated at $t = 0$), it evolves the state according to the free Hamiltonian from time $t = 0$ to time $t = t_1$ and then according to the full Hamiltonian in the opposite time direction from t_1 to 0. If the limits $t_1 \to \pm\infty$ existed, the operators $\Omega(-\infty)$ and $\Omega(+\infty)$ would transform free-particle states into their asymptotic "in" and "out" counterparts. (That is, we would have $\Psi_\alpha^{\mathrm{in}} = \Omega(-\infty)\Psi_\alpha^{\mathrm{free}}$ and $\Psi_\alpha^{\mathrm{out}} = \Omega(+\infty)\Psi_\alpha^{\mathrm{free}}$.) Our system is, however, placed inside a finite box, which makes the time evolution to a distant past or to a distant future complicated. If we removed the box the Haag's theorem would apply, and we would encounter problems already in the stage of interpreting the transformation (2.46). For that reason, we remain in the box and consider only finite time arguments, $T_+ \leq t \leq T_-$, in the equation (2.46).

In terms of the operator Ω we see from its definition (2.24) that the evolution operator $U(t_2, t_1)$ can be expressed as

$$U(t_2, t_1) = \Omega(t_2)^\dagger\Omega(t_1). \tag{2.47}$$

But this means that we can write (2.45) as

$$\begin{aligned}
\mathrm{T}\left\{\phi_{I1}(x_1)\ldots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\} = \\
\Omega(T_-)^\dagger\left[\Omega(u_1)\phi_I(u_1)\Omega(u_1)^\dagger\right]\left[\Omega(u_2)\phi_I(u_2)\Omega(u_2)^\dagger\right]\ldots \\
\times\left[\Omega(u_N)\phi_I(u_N)\Omega(u_N)^\dagger\right]\Omega(T_+) \\
= \Omega(T_-)^\dagger\phi(u_1)\phi(u_2)\ldots\phi(u_N)\Omega(T_+),
\end{aligned} \tag{2.48}$$

where in the third line we used (2.46) to express the product in terms of the corresponding Heisenberg-picture operators. If we reinstate the time-ordering operator (and return field labels), this reads

$$\begin{aligned}
\mathrm{T}\left\{\phi_{I1}(x_1)\ldots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\} \\
= \Omega(T_-)^\dagger\mathrm{T}\left\{\phi_1(x_1)\phi_2(x_2)\ldots\phi_N(x_N)\right\}\Omega(T_+),
\end{aligned} \tag{2.49}$$

which is a very useful result on its own.

Similarly, from (2.30) and (2.47) (or directly from (2.49)) we see that the time-ordered operator in the denominator on the right-hand side of the GML formula is simply

$$\mathrm{T}\left\{\mathrm{e}^{-i\int_{T_+}^{T_-}dt\,V_I(t)}\right\}=\Omega(T_-)^\dagger\Omega(T_+). \tag{2.50}$$

The whole ratio (with the system still in the box and limited to a finite time interval)

$$\frac{\left(\Psi_0^{\mathrm{free,box}},\mathrm{T}\left\{\phi_{I1}(x_1)\dots\phi_{IN}(x_N)\mathrm{e}^{-i\int_{T_+}^{T_-}dt\int_{\mathrm{box}}d^3\mathbf{x}\,\mathcal{V}_I(x)}\right\}\Psi_0^{\mathrm{free,box}}\right)}{\left(\Psi_0^{\mathrm{free,box}},\mathrm{T}\left\{\mathrm{e}^{-i\int_{T_+}^{T_-}dt\int_{\mathrm{box}}d^3\mathbf{x}\,\mathcal{V}_I(x)}\right\}\Psi_0^{\mathrm{free,box}}\right)} \tag{2.51}$$

therefore reads

$$\frac{\left(\Psi_0^{\mathrm{free,box}},\Omega(T_-)^\dagger\mathrm{T}\left\{\phi_1(x_1)\dots\phi_N(x_N)\right\}\Omega(T_+)\Psi_0^{\mathrm{free,box}}\right)}{\left(\Psi_0^{\mathrm{free,box}},\Omega(T_-)^\dagger\Omega(T_+)\Psi_0^{\mathrm{free,box}}\right)}. \tag{2.52}$$

If we were working in the context of non-relativistic quantum mechanics with the interaction becoming effectively irrelevant when the particles are far apart, we could make sense of the limits of $\Omega(t)$ for $t\to\pm\infty$. We could also remove the box, and the operator $\Omega(-\infty)$ would transform free-particle states into their corresponding "in" states and the operator $\Omega(+\infty)$ would transform free-particle states into their corresponding "out" states. In that case we could directly take the limit $T_+\to-\infty$ and $T_-\to+\infty$ in (the "box-less" version of) (2.52) and obtain

$$\frac{\left(\Psi_0^-,\mathrm{T}\left\{\phi_1(x_1)\dots\phi_N(x_N)\right\}\Psi_0^+\right)}{\left(\Psi_0^-,\Psi_0^+\right)}. \tag{2.53}$$

Here we have denoted the "in" vacuum by Ψ_0^+ and the "out" vacuum by Ψ_0^-. Both these vacua, as well as (what we called) the true vacuum Ψ_0, should all correspond to the same physical state. They might, however, differ by phase factors. Let us therefore write $\Psi_0^-=\mathrm{e}^{i\alpha_-}\Psi_0$, and $\Psi_0^+=\mathrm{e}^{i\alpha_+}\Psi_0$, for some real numbers α_-, α_+. Substituting these expressions into (2.53) we can see that the phase factors cancel between the numerator and the denominator, and after we also use the normalization condition (2.38)

$$(\Psi_0,\Psi_0)=1,$$

we recover the left-hand side of the Gell-Mann–Low formula (2.39).

In quantum field theory, however, particles never stop interacting, and we cannot easily use the above argument. Instead, we proceed in the following steps. First, we use the fact that $H_0\Psi_0^{\rm free} = 0$ to remove the appearance of the free Hamiltonian in the Ω-operators in (2.52). This leads to

$$\frac{\left(\Psi_0^{\rm free,box}, {\rm e}^{-iHT_-}{\rm T}\left\{\phi_1(x_1)\dots\phi_N(x_N)\right\}{\rm e}^{iHT_+}\Psi_0^{\rm free,box}\right)}{\left(\Psi_0^{\rm free,box}, {\rm e}^{-iH(T_- - T_+)}\Psi_0^{\rm free,box}\right)}. \tag{2.54}$$

Next, we continue (2.54), as a function of T_- and T_+, to complex values of T_- and T_+ of the form

$$T_- = (1 - i\epsilon)T, \tag{2.55}$$
$$T_+ = (-1 + i\epsilon)T, \tag{2.56}$$

where $0 < \epsilon$ is a small positive number and T is real. We are interested in the limit of $T \to +\infty$. The imaginary components of T_- and T_+, defined in (2.55) and (2.56), will be instrumental in isolating the contribution of matrix elements between the true vacuum states. To see this, consider any complete basis $\left\{\Psi_a^{\rm box}\right\}$ composed of eigenstates of the full Hamiltonian, which contains the vacuum state $\Psi_0^{\rm box}$,

$$H\Psi_a^{\rm box} = E_a\Psi_a^{\rm box}, \tag{2.57}$$
$$\Psi_0^{\rm box} \in \left\{\Psi_a^{\rm box}\right\}. \tag{2.58}$$

Consider now, for instance, the factor ${\rm e}^{iHT_+}\Psi_0^{\rm free,box}$ from the numerator of (2.54). Expressing this factor in the basis $\left\{\Psi_a^{\rm box}\right\}$ it reads

$${\rm e}^{iHT_+}\Psi_0^{\rm free,box} = \sum_a \Psi_a^{\rm box}{\rm e}^{iE_aT_+}\left(\Psi_a^{\rm box}, \Psi_0^{\rm free,box}\right), \tag{2.59}$$

where the sum includes integration over the continuous parts of the label a. Among the states in $\left\{\Psi_a^{\rm box}\right\}$ the vacuum $\Psi_0^{\rm box}$ has the lowest value of energy and no other basis state has the same energy. (That is, $E_a > E_0$ for $a \neq 0$.) Hence, if we now continue this expression to complex values of T_+ according to (2.56), then relative to the vacuum component the contribution of the components with $a \neq 0$ will be exponentially suppressed in the limit of $T \to \infty$,

$$\begin{aligned}{\rm e}^{iHT_+}\Psi_0^{\rm free,box} \to {\rm e}^{-\epsilon E_0T}\Big[&\Psi_0^{\rm box}{\rm e}^{iE_0T}\left(\Psi_0^{\rm box}, \Psi_0^{\rm free,box}\right)\\ &+ \text{ terms suppressed as } {\rm e}^{-\epsilon(E_a - E_0)T}\Big].\end{aligned} \tag{2.60}$$

We can apply the same reasoning also to the part $\left(\Psi_0^{\text{free,box}}, \mathrm{e}^{-iHT_-} \dots\right)$ of the numerator of the fraction (2.54) and to the corresponding portions of the denominator. This shows that if we make the continuations (2.55) and (2.56), then in the limit of $T \to +\infty$ the fraction (2.54) becomes

$$\frac{\left|\left(\Psi_0^{\text{box}}, \Psi_0^{\text{free,box}}\right)\right|^2 \mathrm{e}^{-2iE_0T}\mathrm{e}^{-2\epsilon E_0T}\left(\Psi_0^{\text{box}}, \mathrm{T}\left\{\phi_1(x_1)\dots\phi_N(x_N)\right\}\Psi_0^{\text{box}}\right)}{\left|\left(\Psi_0^{\text{box}}, \Psi_0^{\text{free,box}}\right)\right|^2 \mathrm{e}^{-2iE_0T}\mathrm{e}^{-2\epsilon E_0T}\left(\Psi_0^{\text{box}}, \Psi_0^{\text{box}}\right)} + \text{ suppressed terms.} \quad (2.61)$$

The suppressed terms contain factors of the form $\exp\left(-\epsilon T(E_a - E_0)\right)$ for some $E_a > E_0$. Notice that in the leading term in (2.61) there are several factors that cancel between the numerator and the denominator. The factors e^{-2iE_0T} correspond to the time evolution of an energy eigenstate with eigenvalue E_0 over the period of $2T$, and the factors $\mathrm{e}^{-2\epsilon E_0T}$ appear as a consequence of the analytic continuation. If we renormalized the full Hamiltonian — e.g., by normal-ordering it — such that $E_0 = 0$, then these factors would be all equal to one. However, we do not need to do that: these factors cancel between the numerator and the denominator anyway, as they should, since physics should depend only on the differences in energy levels of various energy eigenstates and not on their absolute values. Another factor that appears both in the numerator and in the denominator is $\left|\left(\Psi_0^{\text{box}}, \Psi_0^{\text{free,box}}\right)\right|^2$, and it is crucial that it cancels out as well. This is because as we remove the box, that is, as we pass to the limit of infinite volume, these factors each approach zero. In other words, if the normalization condition $\left(\Psi_0^{\text{box}}, \Psi_0^{\text{box}}\right) = 1$ is enforced throughout, then as the box increases,[14]

$$\left(\Psi_0^{\text{box}}, \Psi_0^{\text{free,box}}\right) \to 0. \quad (2.62)$$

Nevertheless, since these factors cancel out, the limit of infinite volume on the right-hand side of the Gell-Mann–Low formula (2.39) is well-defined. Thus, in these limits (first of $T \to +\infty$ and then of the infinite volume of the box) we recover from (2.61) the quantity $(\Psi_0, \mathrm{T}\left\{\phi_1(x_1)\dots\phi_N(x_N)\right\}\Psi_0)$. But this is just the left-hand side of the GML formula, as we needed to prove.

For practical calculations with Feynman diagrams the implications of the Gell-Mann–Low formula are as follows. The denominator on the right-hand side of the formula (2.39) corresponds to the contribution of all Feynman graphs that contain an arbitrary number of bubble diagrams — connected subdiagrams without any external legs. The numerator factor, on the other hand, represents the contribution of all Feynman diagrams, connected or

[14] I will not show this here. One can an get an indication of this behavior if one normalizes to unity the free-particle vacuum and calculates the full vacuum in the first-order of the time-independent perturbation theory.

disconnected, that have in total N external legs — each corresponding to one of the fields appearing in the Green's function — and may contain any number of bubble diagrams. When we divide this value by the denominator we remove the contribution of the bubble diagrams. The formula therefore states that in order to calculate a Green's function we can use the standard Feynman rules but must exclude any bubble diagrams.

2.3 Analytic properties of Laplace transforms

In many applications we need to consider more general quantities than square-integrable functions of a single variable satisfying the primitive causality condition or its simple generalizations. In this section we will briefly state some very general results that are relevant in quantum field theory. We will follow the treatment of [5].

In quantum field theory the fields are usually postulated to be operator valued tempered distributions,[15] and we therefore need to investigate analytic properties of Fourier transforms of such distributions. The Fourier

[15] Distributions are continuous linear functionals defined on some specific set of test functions. The notion of their continuity is usually inherited directly from the corresponding notions (of a metric, or convergence, etc.) for the space of test functions. We will consider test functions that are functions of either a single or several real variables. The most general set of distributions that we will consider is the set $\mathscr{D}'$ of what we will simply call *distributions*. The space of test functions for distributions in $\mathscr{D}'$ is denoted by $\mathscr{D}$. This is the space of all infinitely differentiable complex-valued functions of compact support. Another important space of test functions is $\mathscr{S}$, defined to contain all infinitely differentiable complex-valued functions which together with all their derivatives approach zero (for large values of their arguments) faster than polynomially. Every function from $\mathscr{D}$ is also a member of $\mathscr{S}$, that is, $\mathscr{D} \subset \mathscr{S}$. Distributions defined on the set of test functions $\mathscr{S}$ are called *tempered distributions* and denoted by $\mathscr{S}'$. Tempered distributions can be Fourier transformed and the Fourier transform of a tempered distribution is a tempered distribution. Note that any tempered distribution is also a distribution, $\mathscr{S}' \subset \mathscr{D}'$. [This is a consequence of the fact that $\mathscr{D} \subset \mathscr{S}$, which implies that any $T \in \mathscr{S}'$ is defined and linear on functions from $\mathscr{D}$, and the fact that continuity in $\mathscr{S}$ implies continuity in $\mathscr{D}$.] Coming back to quantum field theory, what we mean by the statement that fields are operator-valued tempered distributions is that for any two vectors ψ_1 and ψ_2 from the domain of a field $\Phi(x)$ the matrix element

$$(\psi_1, \Phi(f)\ \psi_2)\,,$$

when considered as a functional of f, is a tempered distribution. We are using a somewhat ambiguous notation here: $\Phi(x)$ denotes an operator-valued distribution in the variable x — and in this section we will prefer the notation Φ_x for that same object — but $\Phi(f)$ denotes the result of the application of the functional $\Phi(x)$ on the test function $f(x)$. If the operator-valued distribution $\Phi(x)$ is also a function, then we can write

$$\Phi(f) = \int dx\ f(x)\Phi(x).$$

In this thesis I will not provide an introduction to the theory of distributions. The reader can find good introductions for instance in [3, 5] and references cited therein.

transform $\mathcal{F}[T]$ of a tempered distribution T is defined by

$$\mathcal{F}[T](f) = T(\mathcal{F}f). \tag{2.63}$$

For tempered distributions that are also functions this corresponds to

$$\int d\omega \left[\int dx\, T(x)\mathrm{e}^{i\omega x}\right] f(\omega) = \int dx \left[\int d\omega\, f(\omega)\mathrm{e}^{ix\omega}\right] T(x), \tag{2.64}$$

where we could change the order of integration because of the integrability properties of the product $T(x)f(\omega)$ of functions $T \in \mathscr{S}'_x$, $f \in \mathscr{S}_\omega$. (A note regarding the notation: Throughout this section I will sometimes employ various subscripts to make the notation more transparent. For example, the subscript $\mathbf{v}$ in $\mathscr{S}_\mathbf{v}$ means that these test functions are functions of the variable $\mathbf{v}$. This subscript therefore determines the dimension of the domain of test functions. For example, if $\mathbf{v}$ in this example was a three-dimensional vector, $\mathscr{S}_\mathbf{v}$ would be the space of those functions from $\mathbb{R}^3$ to $\mathbb{C}$ which are test functions for tempered distributions. If $\mathbf{v}$ and $\mathbf{u}$ belong to spaces of the same dimension then $\mathscr{S}_\mathbf{v}$ and $\mathscr{S}_\mathbf{u}$ denotes the same space of test functions, and for $f \in \mathscr{S}_\mathbf{v}$ the expression $f(\mathbf{u})$ is perfectly valid. Subscripts for the spaces of distributions carry analogous meanings.)

Recall that our earlier results concerned Fourier transforms of certain functions and some kind of their continuation from the real line to some region of the complex plane. This actually corresponds to the *Laplace* transform, which for functions of a single real variable we will define by the equation

$$\mathcal{L}[f](\xi + i\eta) = \int_{-\infty}^{+\infty} dt\, f(t)\mathrm{e}^{i(\xi+i\eta)t}, \tag{2.65}$$

if the right hand side converges.[16] If the function f is defined on $\mathbb{R}^n$ and $\mathbf{k}\cdot\mathbf{p}$ denotes some scalar product of vectors $\mathbf{k}$ and $\mathbf{p}$ from $\mathbb{R}^n$, we can generalize (2.65) to

$$\mathcal{L}[f](\boldsymbol{\xi} + i\boldsymbol{\eta}) = \int d^n\mathbf{t}\, f(\mathbf{t})\mathrm{e}^{i(\boldsymbol{\xi}+i\boldsymbol{\eta})\cdot\mathbf{t}}. \tag{2.66}$$

Notice that for $\eta = 0$ we recover our definition of the Fourier transform, or its straightforward generalization to $\mathbb{R}^n$. And, as we will discuss below, if

[16] A more common definition is

$$\mathcal{L}[f](p) = \int_0^\infty dt\, f(t)\mathrm{e}^{-pt}.$$

Our definition (2.65) corresponds instead to the *two-sided Laplace* transform, where one integrates over the whole real line. Furthermore, to obtain a better correspondence with applications in quantum theory we substituted $p \to -i(\xi + i\eta)$, which is just a matter of convention. Also, notice that for non-zero η the integral (2.65) exists only for a very restricted class of functions because the exponential factor blows up for either positive or for negative values of t.

this Fourier transform exists, then the Laplace transform for non-zero values of η represents its continuation to the complex plane.

We may use this relation between the Fourier and the Laplace transforms to extend the definition of the Laplace transform to distributions. Consider any distribution $T_{\mathbf{t}} \in \mathscr{D}'_{\mathbf{t}}$. Then

$$\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}} T_{\mathbf{t}} \tag{2.67}$$

is also a distribution in $\mathscr{D}'_{\mathbf{t}}$. (Note that here $\mathbf{t}$ is the variable of the distribution and $\boldsymbol{\eta}$ figures only as a real parameter.) It may happen that for some value of $\boldsymbol{\eta}$ this new distribution is not only in $\mathscr{D}'_{\mathbf{t}}$ but also in $\mathscr{S}'_{\mathbf{t}}$. That is, $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}} T_{\mathbf{t}}$ might be a tempered distribution. In such a case it can be Fourier transformed, and for such values of $\boldsymbol{\eta}$ we define the Laplace transform of $T_{\mathbf{t}}$ as

$$\mathcal{L}[T_{\mathbf{t}}](\boldsymbol{\xi} + i\boldsymbol{\eta}) = \mathcal{F}\left[\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}} T_{\mathbf{t}}\right]_{\boldsymbol{\xi}}. \tag{2.68}$$

On the right-hand side we Fourier transform from the variable $\mathbf{t}$ to the variable $\boldsymbol{\xi}$, and we collect this variable together with the parameter $\boldsymbol{\eta}$ into the complex variable $\boldsymbol{\xi} + i\boldsymbol{\eta}$ of the Laplace transform on the left-hand side. Notice that for distributions $T_{\mathbf{t}}$ that are also functions this definition is consistent with the definition (2.66).

Consider any fixed $T_{\mathbf{t}} \in \mathscr{D}'_{\mathbf{t}}$ and denote the set of all $\boldsymbol{\eta}$ for which $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}} T_{\mathbf{t}}$ is a tempered distribution by Γ. The Laplace transform of $T_{\mathbf{t}}$ is therefore defined on the set $\mathbb{R}^n + i\Gamma$. (By $\mathbb{R}^n + i\Gamma$ we mean the set $\{\boldsymbol{\xi} + i\boldsymbol{\eta} : \boldsymbol{\xi} \in \mathbb{R}^n,\ \boldsymbol{\eta} \in \Gamma\}$.) It is not unusual that Γ is empty, but one can show that it must always be convex.[17] We will call a set of the form $\mathbb{R}^n + i\Gamma$ where Γ is convex a *tube*.

A very important result is that the Laplace transform is always an analytic function on the tube where it exists! Furthermore, it has certain boundedness properties, and any analytic function defined on a tube with those boundedness properties is a Laplace transform of some distribution. Let us quote this remarkable theorem in its entirety.

Theorem 2 *Let $T_{\mathbf{t}}$ be a distribution in $\mathscr{D}'_{\mathbf{t}}$ (with $\mathbf{t} \in \mathbb{R}^n$) and Γ be an open convex set in $\mathbb{R}^n$ such that $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}} T_{\mathbf{t}} \in \mathscr{S}'_{\mathbf{t}}$ for all $\boldsymbol{\eta} \in \Gamma$. Then $\mathcal{L}[T_{\mathbf{t}}](\boldsymbol{\xi} + i\boldsymbol{\eta})$ is an analytic function of $\boldsymbol{\xi} + i\boldsymbol{\eta}$ in the tube $\mathbb{R}^n + i\Gamma$. Furthermore, for any compact subset $K \subset \Gamma$ the Laplace transform $\mathcal{L}[T_{\mathbf{t}}](\boldsymbol{\xi} + i\boldsymbol{\eta})$ satisfies the boundedness condition*

$$\left|\mathcal{L}[T_{\mathbf{t}}](\boldsymbol{\xi} + i\boldsymbol{\eta})\right| \leq \left|P_K(\boldsymbol{\xi})\right|, \tag{2.69}$$

for some polynomial P_K and all $\eta \in K$.

Conversely, every function that is analytic in the tube $\mathbb{R}^n + i\Gamma$ and satisfies the boundedness condition (2.69) for each compact subset K of Γ and

[17] Proofs or partial proofs for all the results quoted in this section can be found in [5].

some polynomial P_K is the Laplace transform of a uniquely determined distribution $T_{\mathbf{t}} \in \mathscr{D}'_{\mathbf{t}}$, such that $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}}T_{\mathbf{t}}$ is in $\mathscr{S}'_{\mathbf{t}}$ for all $\boldsymbol{\eta} \in \Gamma$.

Now we can see clearly that the analytic properties that we have been exploring in this and the previous chapter are not limited to considerations of causality and are instead a general property of Laplace transforms. In fact, causality or locality do not appear in the above theorem at all. As we will see, they are instead related to properties of the set Γ of those values of $\boldsymbol{\eta}$ for which the Laplace transform exists.

The primitive causality condition (1.27) and its direct generalizations all required some function $f(t)$ to vanish for all $t < t_0$, with t_0 a constant. The locality condition of finite-range potentials was also of a similar form. These were special cases of a more general situation in which the support of a distribution is limited to some half-space. Suppose that the support of $T_{\mathbf{t}}$ lies in a half-space $\mathbf{t}\cdot\mathbf{a} > A$, where $\mathbf{a}$ is some fixed vector and A is a real number. If a distribution $T_{\mathbf{t}}$ is a function, then the convergence of the Laplace transform

$$\int d^n\mathbf{t}\; T(\mathbf{t})\mathrm{e}^{i\mathbf{t}\cdot(\boldsymbol{\xi}+i\boldsymbol{\eta})} \tag{2.70}$$

implies that the integral

$$\int d^n\mathbf{t}\; T(\mathbf{t})\mathrm{e}^{i\mathbf{t}\cdot(\boldsymbol{\xi}+i\boldsymbol{\eta})}\mathrm{e}^{-u(\mathbf{t}\cdot\mathbf{a}-A)} = \mathrm{e}^{uA}\int d^n\mathbf{t}\; T(\mathbf{t})\mathrm{e}^{i\mathbf{t}\cdot(\boldsymbol{\xi}+i[\boldsymbol{\eta}+u\mathbf{a}])} \tag{2.71}$$

must also converge for any $u \geq 0$. Hence, if $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{t}}T(t) \in \mathscr{S}'_{\mathbf{t}}$ then also $\mathrm{e}^{-(\boldsymbol{\eta}+u\mathbf{a})\cdot\mathbf{t}}T(t) \in \mathscr{S}'_{\mathbf{t}}$ for all $u \geq 0$. This also holds for distributions that are not functions, as is stated in the following theorem.

Theorem 3 *Let $T_{\mathbf{t}}$ be a distribution in $\mathscr{D}'_{\mathbf{t}}$ and let the convex set Γ to consist of all $\boldsymbol{\eta} \in \mathbb{R}^n$ such that $\mathrm{e}^{-\mathbf{t}\cdot\boldsymbol{\eta}}T_{\mathbf{t}}$ is a tempered distribution. If the support of $T_{\mathbf{t}}$ lies in a half-space $\mathbf{t}\cdot\mathbf{a} > A$ (for some $\mathbf{a} \in \mathbb{R}^n$ and $A \in \mathbb{R}$), then Γ contains all points of the form $\boldsymbol{\eta} + u\mathbf{a}$ with $\boldsymbol{\eta} \in \Gamma$ and $u \geq 0$.*

Note that this theorem is of little use if Γ is empty. However, in cases when $T_{\mathbf{t}}$ is a tempered distribution we know that Γ contains the point $\boldsymbol{\eta} = 0$ and the above theorem then implies that $\{u\mathbf{a} : u \geq 0\} \subset \Gamma$.

In quantum field theory we usually apply these considerations to matrix elements of products of fields. Since each field carries a spacetime variable it will be convenient to use the symbol $\mathbf{x}$ instead of $\mathbf{t}$ to denote the variables of distributions on which we act with the Laplace transform. Then $\mathbf{x} \in \mathbb{R}^{4m}$, where m depends on the number of fields in the product[18] and 4 corresponds

[18] As we will discuss in more detail near the end of the present section, translation symmetry implies that the vacuum expectation value of a product of fields depends only on relative positions of the fields. Hence, if there are N fields in the product at positions y_1, ..., y_N, the vacuum expectation value depends on $N-1$ variables, and we can choose those for instance as $y_2 - y_1$, $y_3 - y_2$, ..., $y_N - y_{N-1}$. In this case the distribution $T_{\mathbf{x}}$

to the usual dimension of the spacetime. The scalar product $\mathbf{a}\cdot\mathbf{b}$ is then defined as

$$\mathbf{a}\cdot\mathbf{b}=\sum_{i=1}^{m}a_{\mu}^{(i)}b^{(i)\mu}, \tag{2.72}$$

where each $a^{(i)}$ or $b^{(i)}$ is a four-vector and $x_\mu y^\mu$ denotes the usual scalar product in Minkowski space.

Let us denote by V_+ the set of all real future time-like four-vectors

$$V_+=\left\{x\in\mathbb{R}^4 : x^2=\left(x^0\right)^2-\sum_{j=1}^{3}\left(x^j\right)^2>0;\ x^0>0\right\}, \tag{2.73}$$

and by $\bar{V}_+$ we denote its closure, the set of all x such that $x^2\geq 0$ and $x^0\geq 0$. (The notation here can be confusing: x^2 in the previous sentence stands for the square of x, $x^2=x_\mu x^\mu$, while x^0 denotes the zeroth (contravariant) component of x.) In quantum field theory an important case is when the support of $T_{\mathbf{x}}$ is restricted to a subset $\bar{\Gamma}_+$ of $\mathbb{R}^{4m}$ for which each of the four-vectors $x^{(1)},...,x^{(m)}$ is within the closure of the future light-cone,

$$\text{supp}\ T_{\mathbf{x}}\subset\bar{\Gamma}_+=\left\{\left(x^{(1)},\ ...,x^{(m)}\right) : x^{(j)}\in\bar{V}_+ \text{ for each } j=1,\ ...,m\right\}. \tag{2.74}$$

We denote the interior of $\bar{\Gamma}_+$ by Γ_+. That is, Γ_+ is the subset of all $\mathbf{x}=\left(x^{(1)},...,x^{(m)}\right)\in\mathbb{R}^{4m}$ for which each $x^{(j)}$ is in V_+.

Now, for $\mathbf{a}\in\Gamma_+$ each $a^{(j)}$ is a future time-like four-vector, and for $\mathbf{x}\in\bar{\Gamma}_+$ each $x^{(j)}$ is either a future time-like or a future null four-vector. The scalar product of any future time-like four-vector with any future null or time-like four-vector is always positive. Hence, if $T_{\mathbf{x}}$ satisfies the support condition (2.74), then for any $\mathbf{a}\in\Gamma_+$ the product $\mathbf{a}\cdot\mathbf{x}>0$ for all $\mathbf{x}$ in the support of $T_{\mathbf{x}}$. Furthermore, if $T_{\mathbf{x}}$ is also a tempered distribution, then its Laplace transform exists for $\boldsymbol{\eta}=0$. Theorem 3 then implies that the Laplace transform of $T_{\mathbf{x}}$ exists in the tube $\mathbb{R}^{4m}+i\Gamma_+$.

For distributions with their support in $\bar{\Gamma}_+$ one can prove stronger boundedness properties than those stated in Theorem 2, as we will now present.

Theorem 4 *Let $T_{\mathbf{x}}\in\mathscr{D}'_{\mathbf{x}}$ be such that $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{x}}T_{\mathbf{x}}\in\mathscr{S}'_{\mathbf{x}}$ for all $\boldsymbol{\eta}$ in Γ_+. [Here the scalar product is the one defined by (2.72) and the set Γ_+ is defined just below the equation (2.74).] Suppose, furthermore, that* $\text{supp}\ T_{\mathbf{x}}\subset\bar{\Gamma}_+$. *Then for each $\boldsymbol{\eta}$ in Γ_+ there is a polynomial $P_{\boldsymbol{\eta}}$ such that*

$$\left|\mathcal{L}\left[T_{\mathbf{x}}\right](\boldsymbol{\xi}+i\left[\boldsymbol{\eta}+\mathbf{a}\right])\right|\leq\left|P_{\boldsymbol{\eta}}\left(\boldsymbol{\xi}+i\mathbf{a}\right)\right|, \tag{2.75}$$

would be defined on $\mathbb{R}^{4(N-1)}$, and we would have $\mathbf{x}=\left(x^{(1)},...,x^{(N-1)}\right)$, with variables $x^{(j)}$ defined as $x^{(1)}=y_2-y_1$, $x^{(2)}=y_3-y_2$, etc. This comment largely applies even if the matrix element is not a vacuum expectation value because the dependence on the absolute position (say, on y_1 or on some other weighted average of y_1, ..., y_N) can be easily factored out.

for all $\boldsymbol{\xi} \in \mathbb{R}^{4m}$ *and all* $\mathbf{a} \in \Gamma_+$.

Conversely, if $f(\boldsymbol{\xi} + i\boldsymbol{\eta})$ *is a function analytic in* $\mathbb{R}^{4m} + i\Gamma_+$ *and for each* $\boldsymbol{\eta} \in \Gamma_+$ *there exists a polynomial* $P_{\boldsymbol{\eta}}$ *such that the condition (2.75) is satisfied, then* f *is the Laplace transform of a distribution with support in* $\bar{\Gamma}_+$.

If also $T_{\mathbf{x}} \in \mathscr{S}'_{\mathbf{x}}$ then one can derive further boundedness properties for the Laplace transform, but we will not quote them here. (The reader can find them in [5].)

Note that in Theorem 4 it is not assumed that the Laplace transform exists for $\boldsymbol{\eta} = 0$. In fact, sometimes such a boundary value does not exist even in the sense of distributions. For $T_{\mathbf{x}} \in \mathscr{S}'_{\mathbf{x}}$, however, the boundary value exists and is just the Fourier transform of $T_{\mathbf{x}}$.

Theorem 5 *Let* $T_{\mathbf{x}} \in \mathscr{S}'_{\mathbf{x}}$ *be such that* $\mathrm{e}^{-\boldsymbol{\eta}\cdot\mathbf{x}} T_{\mathbf{x}} \in \mathscr{S}'_{\mathbf{x}}$ *for all* $\boldsymbol{\eta}$ *in* Γ_+. *Then, for any* $f \in \mathscr{S}_{\boldsymbol{\xi}}$,

$$\lim_{\boldsymbol{\eta}\to 0} \int d^{4m}\boldsymbol{\xi}\, \mathcal{L}\left[T_{\mathbf{x}}\right](\boldsymbol{\xi} + i\boldsymbol{\eta})\, f(\boldsymbol{\xi}) = \left(\mathcal{F}\left[T_{\mathbf{x}}\right]\right)_{\boldsymbol{\xi}}(f)\,, \tag{2.76}$$

where the limit $\boldsymbol{\eta} \to 0$ *is taken with* $\boldsymbol{\eta}$ *restricted to some closed cone within* Γ_+. *In other words,* $\mathcal{L}\left[T_{\mathbf{x}}\right]$ *converges in* $\mathscr{S}'_{\boldsymbol{\xi}}$ *to* $\mathcal{F}\left[T_{\mathbf{x}}\right]$ *for* $\boldsymbol{\eta} \to 0$ *within any closed cone in* Γ_+.

Conversely, if $\mathcal{L}\left[T_{\mathbf{x}}\right]$ *exists for all* $\boldsymbol{\eta}$ *in* Γ_+ *and converges in* $\mathscr{S}'_{\boldsymbol{\xi}}$ *for* $\boldsymbol{\eta} \to 0$ *in any closed cone in* Γ_+, *then* $T_{\mathbf{x}}$ *is a tempered distribution.*

Note that when considered as a function of $\boldsymbol{\xi}$ the Laplace transform $\mathcal{L}\left[T\right](\boldsymbol{\xi} + i\boldsymbol{\eta})$ is a tempered distribution. This is a consequence of Theorem 2, which implies that as a function of $\boldsymbol{\xi}$ the Laplace transform $\mathcal{L}\left[T\right](\boldsymbol{\xi} + i\boldsymbol{\eta})$ is bounded by a polynomial. (The variable $\boldsymbol{\eta}$ is then a parameter. That is, for each admissible value of $\boldsymbol{\eta}$ we get a distribution from $\mathscr{S}'_{\boldsymbol{\xi}}$.) When Theorem 5 states that '$\mathcal{L}\left[T_{\mathbf{x}}\right]$ converges in $\mathscr{S}'_{\boldsymbol{\xi}}$' the Laplace transform is interpreted in this sense: as a function of $\boldsymbol{\xi}$ with $\boldsymbol{\eta}$ being treated as a parameter.

This concludes our presentation of relevant mathematical results. In the remaining part of this section we will introduce what is perhaps the most fundamental application of these results to the theory of relativistic quantum fields. We will consider the analytic properties of vacuum expectation values of fixed-order products of fields, such as

$$(\Psi_0, \phi_1(y_1) \ldots \phi_N(y_N)\Psi_0)\,, \tag{2.77}$$

where, as usual, Ψ_0 stands for the vacuum state, and ϕ_1, ..., ϕ_N are components of any Lorentz-covariant fields. Recall our assumption that for any field ϕ_j and any state vectors Ψ_A and Ψ_B from the domain of that field the matrix element $(\Psi_B, \phi_j(y)\Psi_A)$ is a tempered distribution in $\mathscr{S}'_y$.

The Schwartz nuclear theorem then implies that (2.77) is a tempered distribution in all the variables y_1, ..., y_N together. (See, e.g., [5] and references cited therein.) Furthermore, as we already mentioned, properties of such matrix elements under space-time translations imply that the vacuum expectation value (2.77) depends only on the differences of the coordinates of the fields.[19] It will prove convenient to express it in terms of the variables $x_1 = y_1 - y_2$, $x_2 = y_2 - y_3$, ..., $x_{N-1} = y_{N-1} - y_N$. For any fixed choice of the fields ϕ_1, ..., ϕ_N in the product in (2.77) we can then write

$$(\Psi_0, \phi_1(y_1) \dots \phi_N(y_N)\Psi_0) = W(x_1, ..., x_{N-1}), \tag{2.78}$$

where $W \in \mathscr{S}'_{(x_1,...,x_{N-1})}$. We will now argue that W is analytic in the tube $\mathbb{R}^{4(N-1)} - i\Gamma_+$. (According to our earlier definition, Γ_+ is the subset of $\mathbb{R}^{4(N-1)}$ for which each x_j, $j = 1, ..., N-1$, is inside the future light-cone, $x_j \in V_+$.) Note that this region does not include the real hyperplane, and at this point we do not claim anything about the analyticity of W for real values of x_1, ..., x_{N-1}. In fact, we know from the renormalization theory that W is quite singular when any of the separations x_1, ..., x_{N-1} approach zero.

As we have seen in the first part of this section, we can expect this kind of analytic properties to result from some restrictions on the support of the Fourier transform of $W(x_1, ..., x_{N-1})$, which we denote by $\tilde{W}(q_1, ..., q_{N-1})$,

$$\begin{aligned}\tilde{W}(q_1, ..., q_{N-1}) &= \mathcal{F}[W](q_1, ..., q_{N-1}) \\ &= \int d^4x_1 \dots d^4x_{N-1} \mathrm{e}^{i\sum_{j=1}^{N-1} q_j x_j} W(x_1, ..., x_{N-1}).\end{aligned} \tag{2.79}$$

(Note that if we were to strictly follow our earlier mathematical statements, we would need to consider support restrictions of the *inverse* Fourier transform of W, and not of its Fourier transform. This difference is not very important though: the inverse Fourier transform differs from the Fourier transform only by the inversion of its argument and the multiplication by a suitable constant factor.) The four-vector q_j corresponds to the four-momentum flowing from the vertex[20] $j+1$ corresponding to the field ϕ_{j+1} in (2.78) to the vertex j corresponding to the field ϕ_j. For the point $(q_1, ..., q_{N-1})$ to be in the support of $\tilde{W}$, each q_k ($k = 1$, ..., $N-1$) must lie inside the spectrum of the four-momenta of the physical states. We will provide arguments for both of these claims in a moment, but for now let us just take them for granted. Now, the four-momentum of a physical state must be either time-like or null and must be future-pointing. (Except for the vacuum, the four-momentum of which is zero.) This means that if each of q_1, ..., q_{N-1}

[19] A derivation of this well-known property is presented near the end of this section.

[20] Note that these are *not* vertices in the (sums of the) usual Feynman graphs, because we are considering fixed-order Green's functions.

lies in the four-momentum spectrum of the states, then each q_k is either inside or on the future light-cone. In other words, $q_k \in \bar{V}_+$ for all k. This means that the support of $\tilde{W}$ lies in $\bar{\Gamma}_+$. Theorem 3, and the discussion that followed it, then imply that

$$W(x_1, ..., x_{N-1}) = \frac{1}{(2\pi)^{4(N-1)}} \mathcal{L}\left[\tilde{W}\left(q_1, ..., q_{N-1}\right)\right](-x_1, ..., -x_{N-1}) \quad (2.80)$$

is analytic in $\mathbb{R}^{4(N-1)} - i\Gamma_+$. (Note the minus sign in front of the $i\Gamma_+$, and the related minus signs in the argument of the Laplace transform on the right-hand side of (2.80). This is because to obtain W from $\tilde{W}$ we need to apply the *inverse* Fourier transform, and this results in the sign reversal of the variable of the corresponding Laplace transform.) Theorem 4 gives us some information about the boundedness properties of this analytic function $W(x_1, ..., x_{N-1})$, and Theorem 5 says that in the limit when we approach the real hyperplane we recover the tempered distribution $W(x_1, ..., x_{N-1})$ defined in (2.78).

Let us now explain in what sense can we consider the four-vector q_j in (2.79) as representing the four-momentum flowing from the vertex $j+1$ to the vertex j. It will be good to start by considering the momentum-space representation of (2.77),

$$\int d^4y_1 \dots d^4y_N \mathrm{e}^{i\sum_{j=1}^{N} p_j y_j} \left(\Psi_0, \phi_1(y_1) \dots \phi_N(y_N)\Psi_0\right). \quad (2.81)$$

Here the four-vector p_j represents the four-momentum exiting[21] the amplitude at the vertex j. To relate the Fourier transform (2.81) to $\tilde{W}$, we first change from the variables y_1, ..., y_N to the set consisting of y_N and the variables x_1, ..., x_{N-1} introduced earlier. In terms of these new variables

[21] It may be somewhat confusing to decide which sign corresponds to the four-momentum entering the amplitude and which sign corresponds to the four-momentum exiting the amplitude. Here I provide a mnemonic-like argument that could be helpful. (See also the discussion of the transformation properties of amplitudes under space-time translations at the end of this section.) The matrix element $(\Psi_0, \phi(x)\Psi_1)$ — where Ψ_0 is the vacuum, Ψ_1 is a single-particle state containing a particle n of four-momentum $p = (E, \mathbf{p})$, and where the field ϕ annihilates this particle — represents a particle that travels to the space-time event x where it is annihilated. That is, at x the four-momentum p exits the amplitude. The phase of a particle of four-momentum p depends on the position x as $\exp(-iEx^0 + i\mathbf{p}\cdot\mathbf{x}) = \exp(-ipx)$ — one can recall this by recalling that the time evolution of such a particle is given by the factor $\exp(-iEt)$ — and this is also the x dependence of the matrix element. To transform to the momentum representation at this four-momentum — that is, to isolate this value of the four-momentum — we must multiply the matrix element with $\exp(iEx^0 - i\mathbf{p}\cdot\mathbf{x}) = \exp(ipx)$ and integrate over d^4x. This means that multiplying a vertex at x with $\exp(ipx)$ and integrating over x extracts the four-momentum component in which the four-momentum p *exits* the amplitude at that vertex.

the variables y_1, ..., y_N can be expressed as

$$\begin{aligned} y_N &= y_N, && (2.82)\\ y_{N-1} &= x_{N-1} + y_N, && (2.83)\\ y_{N-2} &= x_{N-2} + x_{N-1} + y_N, && (2.84)\\ &\dots && (2.85)\\ y_1 &= x_1 + x_2 + \cdots + x_{N-1} + y_N. && (2.86) \end{aligned}$$

The momentum-space matrix element (2.81) can then be written as

$$\begin{aligned} &\int d^4y_1 \dots d^4y_N \mathrm{e}^{i\sum_{j=1}^N p_j y_j} \left(\Psi_0, \phi_1(y_1) \dots \phi_N(y_N)\Psi_0\right) \\ &\quad = \int d^4y_N d^4x_{N-1} \dots d^4x_1 \mathrm{e}^{i(p_N y_N + p_{N-1}(x_{N-1}+y_N) + \cdots + p_1(x_1+\cdots+y_N))} \\ &\qquad\qquad \times \left(\Psi_0, \phi_1(x_1 + x_2 + \cdots + y_N) \dots \phi_N(y_N)\Psi_0\right). \quad (2.87) \end{aligned}$$

To proceed further, we use the translation-invariance property of vacuum expectation values, which implies that if we shift all the field-coordinates y_j by the same amount $-y_N$ — that is, if we substitute $y_j \to y_j - y_N$ for each $j = 1$, ..., N — then the amplitude does not change,

$$\begin{aligned} &\left(\Psi_0, \phi_1(x_1 + \cdots + x_{N-1} + y_N)\phi_2(x_2 + \cdots + x_{N-1} + y_N) \dots \phi_N(y_N)\Psi_0\right) \\ &\quad = \left(\Psi_0, \phi_1(x_1 + \cdots + x_{N-1})\phi_2(x_2 + \cdots + x_{N-1}) \dots \phi_N(0)\Psi_0\right). \quad (2.88) \end{aligned}$$

After substituting this into the right-hand side of (2.87) the integral over the variable y_N becomes trivial, and we obtain

$$\begin{aligned} &\int d^4y_1 \dots d^4y_N \mathrm{e}^{i\sum_{j=1}^N p_j y_j} \left(\Psi_0, \phi_1(y_1) \dots \phi_N(y_N)\Psi_0\right) \\ &= (2\pi)^4 \delta^4\left(p_1 + \cdots + p_N\right) \int d^4x_{N-1} \dots d^4x_1 \mathrm{e}^{i(p_{N-1}x_{N-1} + \cdots + p_1(x_1+\cdots+x_{N-1}))} \\ &\qquad\qquad \times \left(\Psi_0, \phi_1(x_1 + \cdots + x_{N-1}) \dots \phi_{N-1}(x_{N-1})\phi_N(0)\Psi_0\right). \quad (2.89) \end{aligned}$$

Comparing the integral on the right-hand side with the definition of $\tilde{W}$ in (2.79), we see that

$$\begin{aligned} &\int d^4y_1 \dots d^4y_N \mathrm{e}^{i\sum_{j=1}^N p_j y_j} \left(\Psi_0, \phi_1(y_1) \dots \phi_N(y_N)\Psi_0\right) \\ &\quad = (2\pi)^4 \delta^4\left(p_1 + \cdots + p_N\right) \tilde{W}\left(p_1, p_1 + p_2, ..., p_1 + \cdots + p_{N-1}\right). \quad (2.90) \end{aligned}$$

Note the presence of the four-momentum-conservation delta function.

The equation (2.90) shows that the four-momentum q_1 of (2.79) is equal to p_1, the four-momentum exiting the amplitude at the vertex 1, the four-momentum q_2 is equal to $p_1 + p_2$, the sum of four-momenta exiting the

amplitude at the vertex 1 and the vertex 2, etc. Now, if we adopt the picture that at each vertex j the four-momentum arrives from the vertex $j+1$, which is just to the right of the vertex j, and then either directly exits the diagram as p_j or flows further left into the vertex $j-1$, and if we also assume[22] that the four-momentum is conserved at each vertex, then we can easily see that the four-momentum q_j corresponds to the flow of four-momentum from $j+1$ to j. For, if the four-momentum is conserved at each vertex and flows from right to left, then the four-momentum that flows into the vertex j from the vertex $j+1$ must be equal to the total four-momentum that leaves the amplitude either at the vertex j or at some other vertex further downstream. That is, it must be equal to the sum $p_j + p_{j-1} + \cdots + p_1$. But according to (2.90) this is just q_j, which therefore does indeed represent the four-momentum flowing into the vertex j from the vertex $j+1$.

Next we turn our attention to the more important part of the preceding argument: we need to show that if any of the q_js lies outside of the spectrum of the four-momenta of physical states then the distribution $\tilde{W}(q_1, ..., q_{N-1})$ vanishes. For this purpose we need to consider a complete basis $\{\Psi_a\}$ of physical states consisting of eigenvectors of the operator P^μ of the total four-momentum,

$$P^\mu \Psi_a = p_a^\mu \Psi_a, \tag{2.91}$$

$$\langle \Psi_b | \Psi_a \rangle = 0 \text{ if } a \neq b, \tag{2.92}$$

$$\int \frac{da}{m(a)} \, |\Psi_a\rangle \, \langle \Psi_a| = \mathbf{1}. \tag{2.93}$$

The first one of these equations expresses the condition that the states are eigenvectors of the operator P^μ, and p_a^μ stands there for the total four-momentum of the state Ψ_a. The second equation indicates that the states in $\{\Psi_a\}$ are orthogonal, and the third equation represents the completeness relation for the basis — it says that if we sum over the intermediate states in $\{\Psi_a\}$ we recover the identity operator $\mathbf{1}$. The symbol $\int da$ is meant to represent the integral over all continuous labels as well as the sum over all discrete labels in the multi-index a. We added an extra factor $1/m(a)$ to account for the possibility that a non-trivial measure needs to be employed in this integration. (This, of course, depends on the normalization of the states in $\{\Psi_a\}$.) Now we can insert this sum over the intermediate states $\{\Psi_a\}$ in

[22] From what has been said so far these are just assumptions. After all, we have not even specified what exactly do we mean by diagrams that would correspond to the fixed-order Green's functions.

between any two neighboring operators in (2.81). We obtain[23]

$$\int \frac{da}{m(a)} \int d^4y_1 \dots d^4y_N \mathrm{e}^{i\sum_{j=1}^{N} p_j y_j} \, (\Psi_0, \phi_1(y_1) \dots \phi_k(y_k)\Psi_a) \\ \times (\Psi_a, \phi_{k+1}(y_{k+1})\phi_N(y_N)\Psi_0) \,, \quad (2.94)$$

where the sum over intermediate states has been inserted between the k-th and the k+1-th member of the product. The conservation of four-momentum requires that

$$\int d^4y_1 \dots d^4y_k \mathrm{e}^{i\sum_{j=1}^{k} p_j y_j} \, (\Psi_0, \phi_1(y_1) \dots \phi_k(y_k)\Psi_a) \quad (2.95)$$

vanishes unless $p_a = p_1 + \dots + p_k$. Similarly,

$$\int d^4y_{k+1} \dots d^4y_N \mathrm{e}^{i\sum_{j=k+1}^{N} p_j y_j} \, (\Psi_a, \phi_{k+1}(y_{k+1})\phi_N(y_N)\Psi_0) \quad (2.96)$$

vanishes unless $p_a = -p_{k+1} - \dots - p_N$. These two conditions are equivalent because of the conservation of the total four-momentum,

$$p_1 + \dots + p_N = 0,$$

which is enforced by the delta function in (2.90). Furthermore, as we can also see from (2.90), both these conditions can be expressed as saying that the matrix elements (2.95) and (2.96) vanish unless $p_a = q_k$. This implies that if there is no state Ψ_a for which $p_a = q_k$ then the whole sum over the intermediate states in (2.94) vanishes. The same argument applies also to every other q_j. Hence, the support of (2.81) is restricted to those lists $(p_1, \dots, p_N)$ for which each $q_j = p_1 + \dots + p_j$ $(1 \leq j \leq N-1)$ lies in $\bar{V}_+$. Finally, the equation (2.90) then implies that the same holds for $\tilde{W}(q_1, ..., q_{N-1})$.

In the preceding argument, as well as several other places in this thesis, we have employed the properties of momentum conservation and/or translation covariance of matrix elements of products of field operators. It may be a good idea to present here a simple derivation of those properties. The four-momentum operator P^μ is the generator of translations in space and time. (The symmetry of a physical theory under such transformations then implies that this operator is conserved.) This means that for any state Ψ and any infinitesimal interval dx^μ the state

$$(\mathbf{1} + iP^\mu dx_\mu)\Psi, \quad (2.97)$$

[23] In the current section we strive for a somewhat higher level of mathematical rigor than in the rest of this work. For this reason we should at least remark that this argument is not completely valid: there are issues related to the fact that some of the intermediate states are not normalizable. For a discussion of this issue see [5] and especially the references cited therein.

corresponds to the same physical system as that described by the original state Ψ but as seen by an observer located at the position $x_0^\mu - dx^\mu$, shifted by $-dx$ relative to the position x_0 of the original observer. (In other words, it corresponds to the coordinate transformation $x^\mu \to x^\mu + dx^\mu$. Here the transformations are considered as passive.[24] If, instead, we interpreted the transformation in the active sense, then the effect of (2.97) would be to shift the physical system by $+dx$.) Accordingly, for any field $\phi(x)$, we have

$$\phi(x+dx)(\mathbf{1} + iP^\mu dx_\mu) = (\mathbf{1} + iP^\mu dx_\mu)\phi(x). \tag{2.98}$$

That is, shifting the reference frame by $-dx$ and then applying the field at $x + dx$ (in the new coordinates) is the same as applying the field at x (in the old coordinates) and shifting the reference frame only afterwards. This equation is equivalent to the commutation relation

$$[P^\mu, \phi(x)]_- = -i\frac{\partial}{\partial x_\mu}\phi(x). \tag{2.99}$$

Consider now any matrix element of a product of field operators

$$(\Psi_b, \phi_1(y_1)\dots\phi_N(y_N)\Psi_a) \tag{2.100}$$

between two eigenstates of P^μ. Since Ψ_a and Ψ_b are four-momentum eigenstates, we can write

$$\begin{aligned}(\Psi_b, [P^\mu, \phi_1(y_1)\dots\phi_N(y_N)]\,\Psi_a)& \\ = (p_b - p_a)^\mu\,(\Psi_b, \phi_1(y_1)\dots\phi_N(y_N)\Psi_a)\,,&\end{aligned} \tag{2.101}$$

where p_a and p_b are four-momenta corresponding to the states Ψ_a and Ψ_b, respectively. However, using the equation (2.99) we also have

$$\begin{aligned}(\Psi_b, [P^\mu, \phi_1(y_1)\dots\phi_N(y_N)]\,\Psi_a)& \\ = -i\left(\frac{\partial}{\partial y_{1\mu}} + \dots + \frac{\partial}{\partial y_{N\mu}}\right)(\Psi_b, \phi_1(y_1)\dots\phi_N(y_N)\Psi_a)\,.&\end{aligned} \tag{2.102}$$

[24] A passive symmetry transformation corresponds to a change in the way in which we describe the system, such as a change of coordinates. These transformations do *not* have any influence on the physical system itself. Active transformation are understood differently: under these transformations the way in which we describe the system remains unchanged and it is the physical system itself what undergoes the transformation. These two concepts are closely related. For example, the passive transformation of rotating our frame of reference clockwise by an angle α is equivalent to the active transformation of rotating instead the physical system in the opposite direction, that is, counterclockwise by the angle α. Also, to see that the sign in (2.97) is correct, consider a translation in the time direction corresponding to $dx^\mu = (dt, \mathbf{0})$. The expression (2.97) then reads $(\mathbf{1} + iP^0 dx_0)\Psi = (\mathbf{1} + iP^0 dx^0)\Psi = (\mathbf{1} + iHdt)\Psi$. Since the time evolution of state is $\exp(-iHt)\Psi$, if Ψ is the state of system at time $t = t_0$ then the state $(\mathbf{1} + iHdt)\Psi$ corresponds to the same system but as observed at the time $t = t_0 - dt$. That is, the observer shifted her location in the space-time by $-dt$ along the time-axis, or, in other words, by $-dx^\mu$; just as we claimed below (2.97).

Comparing these two results we see that the matrix element satisfies the differential equation

$$(p_b - p_a)^\mu \, (\Psi_b, \phi_1(y_1) \dots \phi_N(y_N) \Psi_a) \\ = -i \left(\frac{\partial}{\partial y_{1\mu}} + \dots + \frac{\partial}{\partial y_{N\mu}} \right) (\Psi_b, \phi_1(y_1) \dots \phi_N(y_N) \Psi_a) \,. \quad (2.103)$$

We can write the solution of this equation in the form[25]

$$(\Psi_b, \phi_1(y_1) \dots \phi_N(y_N) \Psi_a) \\ = \mathrm{e}^{i(p_b - p_a)_\mu y_N^\mu} \, (\Psi_b, \phi_1(y_1 - y_N) \dots \phi_{N-1}(y_{N-1} - y_N) \phi_N(0) \Psi_a) \,. \quad (2.104)$$

Note that the same argument also applies to matrix elements of time-ordered products of field operators. To summarize, we found that if a matrix element of any product of field operators — either fixed-ordered or time-ordered — is taken between four-momentum eigenstates Ψ_b and Ψ_a, then under overall translation of all space-time arguments of the fields, such as $y_k \to y_k + \Delta y$ for all k, it transforms by simply acquiring a phase factor of the form $\exp\left(i(p_b - p_a) \cdot \Delta y\right)$. In particular, vacuum expectation values are invariant under such translations: they depend only on the differences between the field coordinates.

From the transformation law (2.104) we can derive the conservation of four-momentum. Multiplying both sides of the equation by $\exp\left(i \sum p_j y_j\right)$ and integrating over all variables y_j we find that the integral over y_N on the right-hand side is trivial and leads to the four-momentum-conservation delta function,

$$\int d^4y_1 \dots d^4y_N \mathrm{e}^{i \sum_{j=1}^N p_j y_j} \, (\Psi_b, \phi_1(y_1) \dots \phi_N(y_N) \Psi_a) \\ = \int d^4y_N d^4 \left(y_1 - y_N\right) \dots d^4 \left(y_{N-1} - y_N\right) \mathrm{e}^{i(p_b - p_a + p_1 + \dots + p_N) y_N} \\ \times \mathrm{e}^{i \sum_{j=1}^{N-1} p_j (y_j - y_N)} \, (\Psi_b, \phi_1(y_1 - y_N) \dots \phi_{N-1}(y_{N-1} - y_N) \phi_N(0) \Psi_a) \\ = (2\pi)^4 \delta^4 \left(p_b + p_1 + \dots + p_N - p_a\right) \int d^4y_1 \dots d^4y_{N-1} \mathrm{e}^{i \sum_{j=1}^{N-1} p_j y_j} \\ \times \left(\Psi_b, \phi_1(y_1) \dots \phi_{N-1}(y_{N-1}) \phi_N(0) \Psi_a\right) . \quad (2.105)$$

[25] Or, more generally, the solution is

$$(\Psi_b, \phi_1(y_1) \dots \phi_N(y_N) \Psi_a) = \mathrm{e}^{i(p_b - p_a)_\mu y^\mu} F(y_1 - y_2, ...),$$

where y is any weighted average of y_1, ..., y_N,

$$y = \sum_{j=1}^N \alpha_j y_j,$$

for some real α_j satisfying $\sum \alpha_j = 1$, and the function F depends only on the differences of the variables y_1, ..., y_N.

The matrix element therefore vanishes unless $p_b + p_1 + \cdots + p_N = p_a$. The total four-momentum exiting the amplitude, either at any of the vertices (as p_1 to p_N) or in the final state (as p_b) must be equal to the incoming four-momentum p_a.

There is one more issue that needs to be mentioned. We have seen that the functions $W(x_1, ..., x_{N-1})$ are analytic on the tube $\mathbb{R}^{4(N-1)} - i\Gamma_+$. However, this domain of analyticity can be extended even further if we make full use of the fact that the fields that appear in the products of the form (2.77) are covariant under the restricted subgroup of Lorentz transformations.[26] It can be shown that the Laplace transforms W are analytic and single-valued on a larger set called the *extended tube*. The extended tube consists of all points that can be reached by any proper *complex* Lorentz transformation from any point in $\mathbb{R}^{4(N-1)} - i\Gamma_+$. Functions W transform covariantly on this extended tube. This uniqueness and analyticity on the extended tube, together with the covariance under complex Lorentz transformations, is used in many applications, although often just implicitly. Let us also note that unlike the original tube $\mathbb{R}^{4(N-1)} - i\Gamma_+$, the extended tube contains some real points, that is, it contains vectors whose all components are real. These real points are called *Jost points*, and the knowledge of their existence can be exploited in various ways. For example, the fact that the functions W can be extended to this real environment can be used to show[27] that Ws that correspond to different permutations (orderings) of fields in (2.77) are analytic continuations of the same analytic function into different regions of $\mathbb{C}^{4(N-1)}$.

[26] Recall that restricted Lorentz transformations are those (real) Lorentz transformations which are both proper — their determinant is $+1$ — and ortochronous — their component ${\Lambda^0}_0 \geq 1$. In some cases the fields are covariant under a larger group of Lorentz transformations, for instance, the group that also contains space inversions.

[27] To provide a rough idea of why this is true consider the simple case of products containing only two fields,

$$\begin{aligned} W_1(y_1 - y_2) &= (\Psi_0, \phi_1(y_1)\phi_2(y_2)\Psi_0)\,, \\ W_2(y_2 - y_1) &= (\Psi_0, \phi_2(y_2)\phi_1(y_1)\Psi_0)\,. \end{aligned}$$

Denote $x = y_1 - y_2$. Now, any Lorentz transformation, either real or complex, must preserve the invariant square x^2. Since points of the extended tube are obtained by Lorentz transforming points from the original tube $\mathbb{R} - iV_+$, for any point on the extended tube with some value of x^2 there must be a point on the original tube with the same value of x^2. Jost points are real, so for them x^2 must be real, and for each Jost point there must exist a point in the original tube with the same value of x^2. But for a point $x = \xi - i\eta$ in the original tube $\mathbb{R} - iV_+$ the condition that $x^2 = \xi^2 - \eta^2 - 2i\xi\cdot\eta$ is real implies that $\xi\cdot\eta = 0$. But $\eta \in V_+$ is time-like, and so ξ must be space-like. As a consequence, any such real $x^2 = \xi^2 - \eta^2$ must negative. At Jost points the separation between y_1 and y_2 is therefore space-like and the condition of local commutativity yields $W_1(y_1 - y_2) = W_2(y_2 - y_1)$.

2.4 Field theoretic argument for the existence of poles

In the previous section we have introduced methods that can be used to show that Fourier transforms of various quantities are boundary values of analytic functions, provided that those quantities vanish on suitable subsets of their domains. We then proceeded to apply those methods to the case of vacuum expectation values of fixed-order products of field operators. Since these functions (or rather distributions) depend on space-time coordinates of the fields, they are sometimes referred to as being expressed in the space-time representation or in the coordinate representation. Their Fourier transforms depend on variables that represent various four-momenta, and, accordingly, they are sometimes referred to as being expressed in the momentum representation.

As we saw, the fact that the coordinate-representation fixed-order products were boundary values of analytic functions depended on particular restrictions satisfied by the support of their Fourier transforms — their momentum space counterparts. The support of those momentum-space matrix elements was restricted to those regions in which the four-momentum flowing between adjacent vertices fell into the energy-momentum spectrum of physical states.

Of course, we can try to argue also in the opposite direction: if we managed to establish suitable restrictions on the support of matrix elements in the space-time representation, then analytic properties of matrix elements in momentum-space representation would follow. In fact, we briefly mentioned how to do this in the case of time-ordered products. At the end of the section 2.1 we considered a simple example of how the support of certain parts of such a product was restricted by the condition of microscopic causality.

In the present section we will continue investigating this kind of matrix elements. That is, we will focus our attention on products that are time-ordered. However, while until now we have been mostly interested in showing for some matrix element that it *is* analytic in a certain region, in the present section we are instead going to start investigating the other side of that same coin: we will be interested in a particular condition when the matrix element is *not* analytic. What we will find is that such matrix elements contain poles, which correspond to single-particle states.

In fact, although we will not discuss it here in greater detail, apart from poles corresponding to single-particle states, several other types of singularities are known to be present in Green's functions. These usually correspond to branch points and are related to processes involving several particles. At the level of Feynman diagrams they can be detected by the application of the Landau conditions, which we briefly mentioned in the section 1.4, and which we will also discuss in the section 2.5. They can be interpreted in

terms of processes involving classical propagation of several on-shell particles,[28] as was realized by Coleman and Norton [15], and this point of view lends itself quite well to attempts at a generalization to non-perturbative settings. In any case, by these or some other kinds of arguments one can attempt to establish a procedure to determine a set of known singularities. It then may be reasonable to simply assume that the actual amplitude is analytic everywhere else except for that set of singularities. This is known as the principle of *maximal analyticity.*

Returning to our topic, we are now going to concentrate on vacuum expectation values of time-ordered products of field operators, such as

$$(\Psi_0, \mathrm{T}\left\{\phi_1(y_1)\dots\phi_N(y_N)\right\}\Psi_0)\,. \tag{2.106}$$

As we already mentioned, by the application of the LSZ reduction formula to vacuum expectation values such as (2.106), one can obtain matrix elements of time-ordered products between any asymptotic states, including the S-matrix. The expectation values (2.106) are often referred to as the (space-time representation/coordinate-space) Green's functions. We are interested in exploring the presence of poles in their Fourier transforms. That is, we are going to investigate momentum-space Green's functions, such as

$$G(q_1,\dots,q_n) = \int d^4y_1\dots d^4y_n\, \mathrm{e}^{iq_1y_1}\dots\mathrm{e}^{iq_ny_n}\,(\Psi_0,\ T\left\{\phi_1(y_1)\dots\phi_n(y_n)\right\}\Psi_0)\,. \tag{2.107}$$

Here, Ψ_0 is the true vacuum state and ϕ_j are fields expressed in the Heisenberg picture. (Our findings, however, apply also to matrix elements containing free-fields and free-particle vacua, since those simply represent the special case of a theory of non-interacting fields.) As was already explained in the previous section, the four-momentum q_j in (2.107) corresponds to the four-momentum that exits the amplitude at the vertex j.

Now suppose that there is an integer a, $1 \le a < n$, and a single-particle state $\Psi_{\mathbf{p},\sigma}$ that has non-vanishing matrix elements both with $\phi_{a+1}\dots\phi_n\Psi_0$ and with $\phi_a^\dagger\dots\phi_1^\dagger\Psi_0$. Our result is then that if we consider G as a function of q^2, where

$$q = q_1 + \dots + q_a = -q_{a+1} - \dots - q_n, \tag{2.108}$$

then G has a pole at $q^2 = m^2$, where m is the mass of the particle corresponding to the state $\Psi_{\mathbf{p},\sigma}$. Furthermore, the residue of this pole can be calculated using the usual Feynman rules as if a field for that particle (represented by $\Psi_{\mathbf{p},\sigma}$) was present in the Lagrangian. (To provide a fully unambiguous statement of this result would require a bit more care than what

[28] In addition to the original work [15], this criterion is explained also in many textbooks or lectures, such as [1, 4]. In the context of the Libby-Sterman analysis, and with an illuminating discussion of the coordinate-space representation, it is explained and used in the textbook [7]. See also [6].

we will offer. Our main purpose is only to show that the aforementioned pole is indeed present in the amplitude.)

To specify the value of the residue we will use matrix elements between asymptotic states. These depend on the normalization of those states. Furthermore, any precise statement of the Feynman rules also depends on various conventions. For the purposes of this section we will therefore need to choose between various popular conventions. Since our treatment will closely follow that of [9], I will adopt the convention from that book, except that for the reader's convenience I will keep using the metric $\mathrm{diag}(1,-1,-1,-1)$ instead of switching to the metric $\mathrm{diag}(-1,1,1,1)$ of [9].

We will normalize single-particle states to the delta function:

$$(\Psi_{\mathbf{p}',\sigma',n'}, \Psi_{\mathbf{p},\sigma,n}) = \delta_{n'n}\, \delta_{\sigma'\sigma}\, \delta^3\left(\mathbf{p}' - \mathbf{p}\right), \tag{2.109}$$

where $\Psi_{\mathbf{p}\sigma n}$ corresponds to a single particle of type n, spin z-component (or helicity) σ and three-momentum $\mathbf{p}$, and similarly for $\Psi_{\mathbf{p}'\sigma'n'}$. This leads to particularly simple commutation or anti-commutation relations for the corresponding creation and annihilation operators, which are just

$$\begin{aligned}
\left[a(\mathbf{p},\sigma,n),\ a(\mathbf{p}',\sigma',n')\right]_{\pm} &= 0, && (2.110)\\
\left[a^\dagger(\mathbf{p},\sigma,n),\ a^\dagger(\mathbf{p}',\sigma',n')\right]_{\pm} &= 0, && (2.111)\\
\left[a(\mathbf{p},\sigma,n),\ a^\dagger(\mathbf{p}',\sigma',n')\right]_{\pm} &= \delta_{n'n}\, \delta_{\sigma'\sigma}\, \delta^3\left(\mathbf{p}' - \mathbf{p}\right). && (2.112)
\end{aligned}$$

Another advantage is that if we sum over intermediate states (such as in (2.93)) from a basis that satisfies this normalization convention then the summation over the single-particle sector does not include any additional measure factors. That is, we have

$$\mathbf{1} = \sum_{n,\sigma} \int d^3\mathbf{p}\ |\Psi_{\mathbf{p},\sigma,n}\rangle \langle \Psi_{\mathbf{p},\sigma,n}| + \text{ a sum over multiparticle states.} \tag{2.113}$$

However, there is a price that we pay for this convenience. The unitary operator $U(\Lambda)$, which corresponds to a restricted Lorentz transformation $x \to \Lambda x$, does not transform $\Psi_{\mathbf{p},\sigma,n}$ directly to a superposition of $\Psi_{\mathbf{p}_\Lambda,\sigma',n}$ that would be given only by the transformation properties within the appropriate little group, but the state also acquires an additional, energy-dependent factor. (Here, $\mathbf{p}_\Lambda$ stands for the space components of the four-vector Λp: $\Lambda p = ((\Lambda p)^0, \mathbf{p}_\Lambda)$.) More specifically, we have

$$U(\Lambda)\Psi_{\mathbf{p},\sigma,n} = \sqrt{\frac{(\Lambda p)^0}{p^0}} \sum_{\sigma'} D^{(j)}_{\sigma'\sigma}\left(W(\Lambda,p)\right) \Psi_{\mathbf{p}_\Lambda,\sigma',n}, \tag{2.114}$$

where j is the spin of the particle n, $W(\Lambda,p)$ is the Wigner little group transformation corresponding to the Lorentz transformation Λ applied to

the four-momentum p, and $D^{(j)}_{\sigma'\sigma}$ is the spin j representation of the little group. (This is for massive particles. The little group then consists only of rotations. Massless particles must be treated somewhat differently. These issues are not directly related to our present topic, so I do not explain them here. They are introduced in detail in Chapter 2 of [9], or also in [5].) While the superposition $\sum_{\sigma'} D^{(j)}_{\sigma'\sigma}\left(W(\Lambda,p)\right)\Psi_{\mathbf{p}_\Lambda,\sigma',n}$ is a direct consequence of Wigner's construction of the (infinite-dimensional) unitary representation of the restricted Lorentz group, the extra factor $(\Lambda p)^0/p^0$ results from our particular normalization convention. From this point of view the normalization (2.109) may appear as perhaps being not particularly natural, even though it is convenient.

Furthermore, in order to speak explicitly about the Feynman rules, we also need to fix the convention for the coefficient functions of the annihilation and creation operators that appear in the free versions of the fields ϕ_j. We will write a general Lorentz-covariant free field ϕ_{Ij} describing a particle of species n in the form

$$\phi_{Ij}(x) = \frac{1}{(2\pi)^{3/2}} \sum_\sigma \int d^3\mathbf{p} \left\{ u_j(\mathbf{p},\sigma) \mathrm{e}^{-ipx} a(\mathbf{p},\sigma,n) + v_j(\mathbf{p},\sigma)\mathrm{e}^{ipx} a^\dagger(\mathbf{p},\sigma,n^{\mathrm{c}}) \right\}. \tag{2.115}$$

Here, n^{c} is the antiparticle for n, and the operators $a(\mathbf{p},\sigma,n)$ and $a^\dagger(\mathbf{p},\sigma,n^{\mathrm{c}})$ annihilate and create the corresponding free particles. The coefficient functions u and v depend on the type of the field. For example, for a scalar field we have $u(\mathbf{p}) = v(\mathbf{p}) = 1/\sqrt{2p^0}$, where $p^0 = \sqrt{\mathbf{p}^2 + m_n^2}$. The field ϕ_{Ij} may transform according to any representation of the Lorentz group. Note that here the coefficients u, v are just complex numbers, and not tensors or spinors. That is, the field ϕ_{Ij} can be also thought of as a component of a field.

The definition (2.115) fixes the convention for the coefficient functions. [This is because up to its overall normalization the field itself is fixed by its properties under Lorentz transformations and translations. Its normalization is in turn determined by the canonical commutation or anticommutation relations. The right-hand side of (2.115) contains only two kinds of not fully specified quantities: the coefficient functions and the creation/annihilation operators. But the creation and annihilation operators become fully specified once a normalization of single-particle states, such as (2.109), is selected.] Now that we are equipped with the conventions (2.109)–(2.115), the Feynman rules follow in principle unambiguously (except for the choice of gauge, etc). In particular, we can now specify the free propagator in terms of the coefficient functions u and v. The propagator between two free fields can be defined as

$$i\Delta_{lm}(x,y) = (\Psi_0^{\mathrm{free}}, T\{\phi_{Il}(x)\phi_{Im}^\dagger(y)\}\Psi_0^{\mathrm{free}}), \tag{2.116}$$

and expressed in the form

$$i\Delta_{lm}(x,y) = \frac{i}{(2\pi)^4}\int d^4q \frac{P_{lm}(q)\mathrm{e}^{-iq(x-y)}}{q^2-m_n^2+i\epsilon}, \tag{2.117}$$

where P_{lm} is a polynomial in the components of q. On the mass-shell, this polynomial is given in terms of the coefficient functions as

$$P_{lm}(\mathbf{q},\sqrt{\mathbf{q}^2+m_n^2}) = 2\sqrt{\mathbf{p}^2+m_n^2}\sum_\sigma u_l(\mathbf{q},\sigma)u_m^*(\mathbf{q},\sigma), \tag{2.118}$$

and can be continued to off-shell momenta covariantly.[29]

Now we are ready to state and interpret the result regarding the value of the residue. The full result is that if there is a single-particle physical state $\Psi_{\mathbf{p},\sigma}$ that has non-vanishing matrix elements both with $\phi_{a+1}\dots\phi_n\Psi_0$ and with $\phi_a^\dagger\dots\phi_1^\dagger\Psi_0$, then the amplitude G, defined in (2.107)

$$\begin{aligned} &G(q_1,\dots,q_n) \\ &\qquad = \int d^4y_1\dots d^4y_n\ \mathrm{e}^{iq_1y_1}\dots\mathrm{e}^{iq_ny_n}\left(\Psi_0,\ T\left\{\phi_1(y_1)\dots\phi_n(y_n)\right\}\Psi_0\right), \end{aligned}$$

when understood as a function of q^2, with the variable q defined in (2.108)

$$q = q_1+\dots+q_a = -q_{a+1}-\dots-q_n,$$

has a pole at $q^2=m^2$, and its residue is given by

$$\begin{aligned} G \xrightarrow[q^2\to m^2]{} i\frac{2\sqrt{q^2+m^2}}{q^2-m^2+i\epsilon}&(2\pi)^7\delta^4(q_1+\dots+q_n) \\ &\times\sum_\sigma M_{0|\mathbf{q},\sigma}(q_2,\dots,q_a)M_{\mathbf{q},\sigma|0}(q_{a+2},\dots,q_n). \end{aligned} \tag{2.119}$$

Here $M_{0|\mathbf{q},\sigma}$ and $M_{\mathbf{q},\sigma|0}$ are defined by

$$\begin{aligned} \int d^4y_1\dots d^4y_a\mathrm{e}^{iq_1y_1}\dots\mathrm{e}^{iq_ay_a}&\left(\Psi_0,T\left\{\phi_1(y_1)\dots\phi_a(y_a)\right\}\Psi_{\mathbf{q},\sigma}\right) \\ &= (2\pi)^4\delta^4(q_1+\dots+q_a-q)M_{0|\mathbf{q},\sigma}(q_2,\dots,q_a), \end{aligned} \tag{2.120}$$

[29] Arriving at these covariant propagators is not always straightforward. If we apply the usual operator techniques to the definition (2.116) we can obtain a non-covariant result. This can be dealt with by adding suitable non-covariant terms to the interaction density that effectively cancel out the non-covariant terms in the propagator and restore the Lorentz-covariance of the resulting Feynman rules. If we start from a Lorentz-invariant Lagrangian density and use the canonical formalism to derive the interaction density, then these non-covariant terms needed to restore the covariance of the propagator appear automatically. However, to obtain covariant Feynman rules it is usually more convenient to simply use the path-integral methods and avoid all the aforementioned difficulties.

and

$$\int d^4y_{a+1}\dots d^4y_n e^{iq_{a+1}y_{a+1}}\dots e^{iq_ny_n}\left(\Psi_{\mathbf{q},\sigma},T\left\{\phi_{a+1}(y_{a+1})\dots\phi_n(y_n)\right\}\Psi_0\right)$$
$$=(2\pi)^4\delta^4(q_{a+1}+\dots+q_n+q)M_{\mathbf{q},\sigma|0}(q_{a+2},\dots,q_n). \quad (2.121)$$

As we partially discussed in the section 2.2, the equation (2.120) means that

$$(2\pi)^4\delta^4(q_1+\dots+q_a-q)M_{0|\mathbf{q},\sigma}(q_2,\dots,q_a)$$

corresponds to the sum of all Feynman diagrams in which a single external line corresponding to the particle $\Psi_{\mathbf{q},\sigma}$ enters the diagram and which have a external vertices corresponding to the operators ϕ_1, ..., ϕ_a through which the four-momenta q_1, ..., q_a leave the diagram, but with all bubble subdiagrams excluded. Similarly, $(2\pi)^4\delta^4(\dots)M_{\mathbf{q},\sigma|0}(q_{a+2},\dots,q_n)$ corresponds to the sum of diagrams with a single external line of the particle $\Psi_{\mathbf{q},\sigma}$ leaving the diagram and $n-a$ extra vertices ϕ_{a+1}, ..., ϕ_n through which the momenta q_{a+1}, ..., q_n exit the diagram, with all bubble subdiagrams excluded.

To better appreciate the result (2.119) let us suppose for a moment that there is a field in the Lagrangian (and the Feynman rules) corresponding to the particle of $\Psi_{\mathbf{q},\sigma}$. With the commutation relations of the form (2.110)–(2.112) and the definition of the coefficient functions implicit in (2.115), the (momentum-space) Feynman rules associate with the external particle line in $(2\pi)^4\delta^4(\dots)M_{0|\mathbf{q},\sigma}$ the kinematic factor

$$\frac{1}{(2\pi)^{3/2}}u(\mathbf{q},\sigma). \quad (2.122)$$

Similarly, the external line of $(2\pi)^4\delta^4(\dots)M_{\mathbf{q},\sigma|0}$ yields the factor $u^*(\mathbf{q},\sigma)/(2\pi)^{3/2}$. Consider now the expression

$$\sum_\sigma\left[(2\pi)^4\delta^4(q_1+\dots+q_a-q)M_{0|\mathbf{q},\sigma}(q_2,\dots,q_a)\right]$$
$$\times\frac{i}{(2\pi)^4}\frac{1}{q^2-m^2+i\epsilon}(2\pi)^3 2q^0 \quad (2.123)$$
$$\times\left[(2\pi)^4\delta^4(q_{a+1}+\dots+q_n+q)M_{\mathbf{q},\sigma|0}(q_{a+2},\dots,q_n)\right].$$

The external particles lines from the two expressions in the square brackets each contribute with the factor of $(2\pi)^{-3/2}$. This cancels the factor of $(2\pi)^3$ in the middle line of (2.123). These two external lines also contribute the factor $u(\mathbf{q},\sigma)u^*(\mathbf{q},\sigma)$. But the middle line of (2.123) contains the factor $2q^0$, and according to (2.118), this factor, together with $u(\mathbf{q},\sigma)u^*(\mathbf{q},\sigma)$ from the external particle lines, yields simply P_{lm} after we sum over the quantum number σ. But (2.117) then implies that the middle line, together with the factors from the external particle lines, corresponds after the summation over σ just to the momentum-space propagator of the particle represented

by $\Psi_{\mathbf{q},\sigma}$! This means that in terms of Feynman diagrams (2.123) corresponds to the sum of all Feynman diagrams (excluding bubble subdiagrams) that consist of two parts, one having a external vertices corresponding to fields ϕ_1, ..., Φ_a and a single external propagator line corresponding to the particle of $\Psi_{\mathbf{q},\sigma}$, which connects it to the second part of the diagram, which except for this external propagator line also contains $n-a$ external vertices, corresponding to fields ϕ_{a+1}, ..., ϕ_n.

Now, if we integrate (2.123) over d^4q we obtain

$$\begin{aligned}
&\int d^4q \sum_\sigma \Big[(2\pi)^4\delta^4(q_1+\cdots+q_a-q)M_{0|\mathbf{q},\sigma}(q_2,\ldots,q_a)\Big] \\
&\times \frac{i}{(2\pi)^4}\frac{(2\pi)^3 2q^0}{q^2-m^2+i\epsilon}\Big[(2\pi)^4\delta^4(q_{a+1}+\cdots+q_n+q)M_{\mathbf{q},\sigma|0}(q_{a+2},\ldots,q_n)\Big] \\
&\qquad = (2\pi)^7\delta^4(q_1+\cdots+q_n)\frac{i2q^0}{q^2-m^2+i\epsilon} \\
&\qquad\quad \times \sum_\sigma M_{0|\mathbf{q},\sigma}(q_2,\ldots,q_a)M_{\mathbf{q},\sigma|0}(q_{a+2},\ldots,q_n), \quad (2.124)
\end{aligned}$$

where $q = q_1+\cdots+q_a$. This is the same as the limit in (2.119). Hence, in the limit of $q^2 \to m^2$ the pole structure of the full unperturbed amplitude G is the same as would arise from the propagator of the particle $\Psi_{\mathbf{q},\sigma}$ connecting the first a and the last $n-a$ vertices. It is important to realize that this is true even if there is no field of the particle of mass m in the Lagrangian. That is, the pole structure is the same even if the particle of the state $\Psi_{\mathbf{q},\sigma}$ is a composite particle, a bound state of particles whose fields do appear in the Lagrangian. The only thing that matters is that there is a physical state $\Psi_{\mathbf{q},\sigma}$ corresponding to that particle. This also implies that if there is a field in the Lagrangian whose associated particle is not observed in the physical spectrum, then the corresponding pole should *not* appear in the full Green's function (2.107). To illustrate this point with a well-known example, consider the theory of quantum chromodynamics. In that theory, the product (2.107) of quark and/or gluon fields should contain poles corresponding to mesons, such as π- or K-mesons, or baryons, such as nucleons or hyperons, but should not contain quark or gluon poles, because due to the color confinement those particles do not appear in asymptotic states.

Let us now proceed with the proof. Our assumption was that both the matrix elements $(\Psi_{\mathbf{p},\sigma}, \phi_{a+1}\ldots\phi_n\Psi_0)$ and $(\Psi_0, \phi_1\ldots\phi_a\Psi_{\mathbf{p},\sigma})$ do not vanish. For this reason we restrict our attention only to those regions in the integral (2.107) where in the time-ordered product all the operators ϕ_1, ..., ϕ_a appear to the left of the operators ϕ_{a+1}, ..., ϕ_n

$$\begin{aligned}
G = \int d^4y_1\ldots d^4y_n\,\theta\left(\min(y_1^0,\ldots,y_a^0)-\max(y_{a+1}^0,\ldots,y_n^0)\right)\mathrm{e}^{iq_1y_1}\ldots\mathrm{e}^{iq_ny_n} \\
\times(\Psi_0,\ T\{\phi_1(y_1)\ldots\phi_n(y_n)\}\Psi_0)\ +\ \text{other contributions.} \quad (2.125)
\end{aligned}$$

Here "other contributions" refers to the contribution of the rest of the integral, when $\min(y_1^0,\ldots,y_a^0) < \max(y_{a+1}^0,\ldots,y_n^0)$. Because of the explicit ordering enforced by the theta function we can split the time-ordered product into two parts

$$\begin{aligned}\theta\left(\min(y_1^0,\ldots,y_a^0)-\max(y_{a+1}^0,\ldots,y_n^0)\right)&\\ \times T\left\{\phi_1(y_1)\ldots\phi_a(y_a)\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\right\}&\\ =\theta\left(\min(y_1^0,\ldots,y_a^0)-\max(y_{a+1}^0,\ldots,y_n^0)\right)&\\ \times T\left\{\phi_1(y_1)\ldots\phi_a(y_a)\right\}\,T\left\{\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\right\}.&\end{aligned}$$

This allows us to insert the summation over a complete set of intermediate physical states $\{\Psi_a\}$ in between $T\{\phi_1(y_1)\ldots\phi_a(y_a)\}$ and $T\{\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\}$,

$$\begin{aligned}&T\left\{\phi_1(y_1)\ldots\phi_a(y_a)\right\}\,T\left\{\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\right\}=\\ &\quad T\left\{\phi_1(y_1)\ldots\phi_a(y_a)\right\}\left(\int da\,|\Psi_a\rangle\,\langle\Psi_a|\right)T\left\{\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\right\}.\end{aligned}$$

Here the integral $\int da$ symbolizes integration over continuous labels and sums over discrete labels in the multi-index a. (Note that unlike in (2.93) we did not explicitly indicate the presence of a non-trivial measure. This is because at present we are following the normalization convention (2.109).) We choose the set $\{\Psi_a\}$ such that it includes the states $\Psi_{\mathbf{q},\sigma}$. Now we can further restrict our attention and among all the states in $\int da\,|\Psi_a\rangle\,\langle\Psi_a|$ focus only on the states $\Psi_{\mathbf{q},\sigma}$. In this manner, we write

$$\begin{aligned}G=\int d^4y_1\ldots d^4y_n\theta\left(\min(y_1^0,\ldots,y_a^0)-\max(y_{a+1}^0,\ldots,y_n^0)\right)\mathrm{e}^{iq_1y_1}\ldots\mathrm{e}^{iq_ny_n}&\\ \times\sum_\sigma\int d^3\mathbf{q}\,(\Psi_0,T\left\{\phi_1(y_1)\ldots\phi_a(y_a)\right\}\Psi_{\mathbf{q},\sigma})&\\ \times(\Psi_{\mathbf{q},\sigma},T\left\{\phi_{a+1}(y_{a+1})\ldots\phi_n(y_n)\right\}\Psi_0)&\\ +\text{ other contributions},&\end{aligned}\tag{2.126}$$

where the "other contributions" now contains not only the integration over the region $\min(y_1^0,\ldots,y_a^0) < \max(y_{a+1}^0,\ldots,y_n^0)$ but also the contribution over the region $\min(y_1^0,\ldots,y_a^0) > \max(y_{a+1}^0,\ldots,y_n^0)$ from the intermediate states other than $\Psi_{\mathbf{q},\sigma}$.

In our next step we need to make use of the transformation properties (2.104) of matrix elements between four-momentum eigenstates under translation. They imply

$$\begin{aligned}&(\Psi_0,T\left\{\phi_1(y_1)\phi_2(y_2)\ldots\phi_a(y_a)\right\}\Psi_{\mathbf{q},\sigma})\\ &\quad=\mathrm{e}^{-iqy_1}\,(\Psi_0,T\left\{\phi_1(0)\phi_2(y_2-y_1)\ldots\phi_a(y_a-y_1)\right\}\Psi_{\mathbf{q},\sigma})\end{aligned}\tag{2.127}$$

and

$$\begin{aligned}&(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(y_{a+1})\phi_{a+2}(y_{a+2})\dots\phi_n(y_n)\right\}\Psi_0)\\&\quad= \mathrm{e}^{iqy_{a+1}}\,(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0)\phi_{a+2}(y_{a+2}-y_{a+1})\dots\phi_n(y_n-y_{a+1})\right\}\Psi_0)\,.\end{aligned} \tag{2.128}$$

We now make the following change of the integration variables[30]

$$\begin{aligned}x_1 &= y_1,\\ x_2 &= y_2 - y_1,\\ &\dots\\ x_a &= y_a - y_1,\\ x_{a+1} &= y_{a+1},\\ x_{a+2} &= y_{a+2} - y_{a+1},\\ &\dots\\ x_n &= y_n - y_{a+1}.\end{aligned} \tag{2.129}$$

In these new variables, and after using (2.127) and (2.128), the equation (2.126) reads

$$\begin{aligned}G = \sum_\sigma \int d^3\mathbf{q} \int d^4x_1\dots d^4x_n \Big[&\\ \theta\left(x_1^0 - x_{a+1}^0 + \min(0, x_2^0, \dots, x_a^0) - \max(0, x_{a+2}^0, \dots, x_n^0)\right)&\\ \times \mathrm{e}^{i(q_1+\dots+q_a-q)x_1}\mathrm{e}^{iq_2x_2}\dots\mathrm{e}^{iq_ax_a}\mathrm{e}^{i(q_{a+1}+\dots q_n+q)x_{a+1}}\mathrm{e}^{iq_{a+2}x_{a+2}}\dots\mathrm{e}^{iq_nx_n}&\\ \times (\Psi_0, T\left\{\phi_1(0)\dots\phi_a(x_a)\right\}\Psi_{\mathbf{q},\sigma})\,(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0)\dots\phi_n(x_n)\right\}\Psi_0)\Big]&\\ +\ \text{other terms.}&\end{aligned} \tag{2.130}$$

The integration over the space components of x_1 and x_{a+1} is now trivial and yields delta function factors $(2\pi)^3\delta^3(\mathbf{q_1}+\dots+\mathbf{q_a}-\mathbf{q})$ and $(2\pi)^3\delta^3(\mathbf{q_{a+1}}+\dots+\mathbf{q_n}+\mathbf{q})$. These factors, in turn, make trivial the integration over $\mathbf{q}$,

[30] These variables are *not* the same as the variables x_j that we used in the previous section.

and we obtain

$$G = (2\pi)^6 \delta^3(\mathbf{q_1} + \cdots + \mathbf{q_n}) \sum_\sigma \int dx_1^0 d^4x_2 \ldots d^4x_a dx_{a+1}^0 d^4x_{a+2} \ldots d^4x_n \Big[\\ \times \theta\left(x_1^0 - x_{a+1}^0 + \min(0, x_2^0, \ldots, x_a^0) - \max(0, x_{a+2}^0, \ldots, x_n^0)\right) \\ \times \mathrm{e}^{i(q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2}) x_1^0} \mathrm{e}^{iq_2 x_2} \ldots \mathrm{e}^{iq_a x_a} \\ \times \mathrm{e}^{i(q_{a+1}^0 + \ldots q_n^0 + \sqrt{\mathbf{q}^2 + m^2}) x_{a+1}^0} \mathrm{e}^{iq_{a+2} x_{a+2}} \ldots \mathrm{e}^{iq_n x_n} \\ \times \left(\Psi_0, T\left\{\phi_1(0) \ldots \phi_a(x_a)\right\} \Psi_{\mathbf{q},\sigma}\right) \left(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0) \ldots \phi_n(x_n)\right\} \Psi_0\right) \Big] \\ + \text{ other terms,} \quad (2.131)$$

where $\mathbf{q}$ is now fixed at the value $\mathbf{q} = \mathbf{q}_1 + \cdots + \mathbf{q}_a$. (Also not that since q stands for the four-momentum of a real particle it must be on-shell.) We still need to evaluate the integrals over x_1^0 and x_{a+1}^0. In order to do so we express the step function θ in terms of its Fourier transform (1.54),

$$\theta(t) = \frac{1}{-2\pi i} \int_{-\infty}^{+\infty} d\omega \frac{\mathrm{e}^{-i\omega t}}{\omega + i\epsilon},$$

where the limit of $\epsilon \to 0^+$ is understood implicitly. Substituting this into (2.131) yields

$$G = (2\pi)^6 \delta^3(\mathbf{q_1} + \cdots + \mathbf{q_n}) \sum_\sigma \int d\omega dx_1^0 d^4x_2 \ldots d^4x_a dx_{a+1}^0 d^4x_{a+2} \ldots d^4x_n \Big[\\ \times \frac{i}{2\pi(\omega + i\epsilon)} \mathrm{e}^{-i\omega\left[\min(0, x_2^0, \ldots, x_a^0) - \max(0, x_{a+2}^0, \ldots, x_n^0)\right]} \\ \times \mathrm{e}^{i(q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2} - \omega) x_1^0} \mathrm{e}^{iq_2 x_2} \ldots \mathrm{e}^{iq_a x_a} \\ \times \mathrm{e}^{i(q_{a+1}^0 + \ldots q_n^0 + \sqrt{\mathbf{q}^2 + m^2} + \omega) x_{a+1}^0} \mathrm{e}^{iq_{a+2} x_{a+2}} \ldots \mathrm{e}^{iq_n x_n} \\ \times \left(\Psi_0, T\left\{\phi_1(0) \ldots \phi_a(x_a)\right\} \Psi_{\mathbf{q},\sigma}\right) \left(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0) \ldots \phi_n(x_n)\right\} \Psi_0\right) \Big] \\ + \text{ other terms.} \quad (2.132)$$

It is now possible to perform the integration over x_1^0 and x_{a+1}^0, which leads

to

$$G = (2\pi)^7 \delta^4(q_1 + \cdots + q_n) \sum_\sigma \int d\omega d^4x_2 \ldots d^4x_a d^4x_{a+2} \ldots d^4x_n \Big[\\ \times 2\pi\delta(q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2} - \omega) \\ \times \frac{i}{2\pi(\omega + i\epsilon)} \mathrm{e}^{-i\omega\left[\min(0, x_2^0, \ldots, x_a^0) - \max(0, x_{a+2}^0, \ldots, x_n^0)\right]} \\ \times \mathrm{e}^{iq_2x_2} \ldots \mathrm{e}^{iq_ax_a} \mathrm{e}^{iq_{a+2}x_{a+2}} \ldots \mathrm{e}^{iq_nx_n} \\ \times \left(\Psi_0, T\left\{\phi_1(0) \ldots \phi_a(x_a)\right\} \Psi_{\mathbf{q},\sigma}\right) \left(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0) \ldots \phi_n(x_n)\right\} \Psi_0\right) \Big] \\ + \text{ other terms.} \quad (2.133)$$

Finally, we can also integrate over the variable ω to obtain

$$G = (2\pi)^7 \delta^4(q_1 + \cdots + q_n) \sum_\sigma \int d^4x_2 \ldots d^4x_a d^4x_{a+2} \ldots d^4x_n \Big[\\ \frac{i}{q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon} \times \\ \times \mathrm{e}^{-i(q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2})\left[\min(0, x_2^0, \ldots, x_a^0) - \max(0, x_{a+2}^0, \ldots, x_n^0)\right]} \\ \times \mathrm{e}^{iq_2x_2} \ldots \mathrm{e}^{iq_ax_a} \mathrm{e}^{iq_{a+2}x_{a+2}} \ldots \mathrm{e}^{iq_nx_n} \\ \times \left(\Psi_0, T\left\{\phi_1(0) \ldots \phi_a(x_a)\right\} \Psi_{\mathbf{q},\sigma}\right) \left(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0) \ldots \phi_n(x_n)\right\} \Psi_0\right) \Big] \\ + \text{ other terms.} \quad (2.134)$$

At this point we are nearly finished. Let us denote the sum $q_1^0 + \cdots + q_a^0$ by q^0. Then we can write[31]

$$\frac{1}{q_1^0 + \cdots + q_a^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon} = \frac{1}{q^0 - \sqrt{\mathbf{q}^2 + m^2} + i\epsilon} = \\ = \frac{q^0 + \sqrt{\mathbf{q}^2 + m^2} - i\epsilon}{(q^0)^2 - \mathbf{q}^2 - m^2 + i\epsilon} = \frac{q^0 + \sqrt{\mathbf{q}^2 + m^2}}{q^2 - m^2 + i\epsilon}.$$

We are investigating the limit of G for q approaching its mass-shell, that is, $q^0 > 0$ and $q^2 \to m^2$. In this limit we can substitute $\sqrt{\mathbf{q}^2 + m^2}$ for q^0 in the numerator of the previous equation. Furthermore, the leading part of the expression $\exp\left(-i(q^0 - \sqrt{\mathbf{q}^2 + m^2})\left[\min(0, y_2^0, \ldots, y_a^0) - \max(0, y_{a+2}^0, \ldots, y_n^0)\right]\right)$

[31] Here we employ "allowed" manipulations with the "$i\epsilon$"-terms. These terms are meant to shift the poles away from the contour of integration so that the overall integral is defined as the limit of regular integrals for $\epsilon \to 0^+$. This has several consequences. In general, we may ignore these terms in numerators. Furthermore, it is only important that ϵ is a small positive number, which means that for any positive real number x, we can substitute $x\epsilon \to \epsilon$.

in (2.134) is just 1. Altogether, we arrive at

$$G \xrightarrow[q^2\to m^2]{} (2\pi)^7\delta^4(q_1+\cdots+q_n)\sum_\sigma \int d^4x_2\ldots d^4x_a d^4x_{a+2}\ldots d^4x_n\Big[\\ \frac{2i\sqrt{q^2+m^2}}{q^2-m^2+i\epsilon}\mathrm{e}^{iq_2x_2}\ldots\mathrm{e}^{iq_ax_a}\mathrm{e}^{iq_{a+2}x_{a+2}}\ldots\mathrm{e}^{iq_nx_n} \\ \times\left(\Psi_0, T\left\{\phi_1(0)\ldots\phi_a(x_a)\right\}\Psi_{\mathbf{q},\sigma}\right)\left(\Psi_{\mathbf{q},\sigma}, T\left\{\phi_{a+1}(0)\ldots\phi_n(x_n)\right\}\Psi_0\right)\Big], \tag{2.135}$$

where we have also dropped the "other terms" because those should not (in the absence of degeneracy) produce any pole at the point $q^2 = m^2$. (The terms corresponding to other single particle states produce poles at different values of q^2, while terms corresponding to multi-particle states lead to branch points in q^2. The terms corresponding to different time orderings produce poles and branch points in other variables.) Thus, we have finally obtained (2.119).

2.5 Normal thresholds

In addition to poles that correspond to single-particle states, another common type of singularity that appears in Green's functions are branch points. As we already mentioned, these can be usually interpreted as related to processes involving two or more particles, and at the level of perturbation theory we can detect them by the application of the Landau conditions. Our aim in this section will be to introduce one particular kind of branch point, called *normal threshold*. Other branch points are often called *anomalous thresholds*. There is, however, nothing really anomalous about them, and the terminology is due to historical reasons. Normal thresholds were the easiest to predict and understand, because they follow in a relatively simple way from the condition of unitarity. In the second part of this section we will describe normal thresholds from exactly that point of view. But first we will consider two very simple examples that are meant to indicate how branch points appear in perturbation theory.

2.5.1 Branch points in Feynman integrals

There are various approaches to perturbation theory, with the most common being the covariant formalism that we discussed briefly in Sec. 2.2. As a rule, beyond the first order they usually lead to terms in which we integrate over various loop variables. Integrals of this type are often called *Feynman integrals*, and they are parametric integrals of functions that are analytic except for a number of singularities, usually poles, the location of which

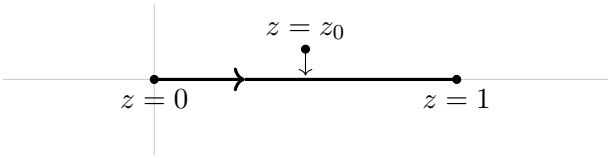


Figure 2.1: The contour of integration as determined by the equation (2.136). We integrate the function $1/(z - z_0)$ along the straight line from $z = 0$ to $z = 1$. In the figure we located the pole at z_0 just above the contour of integration. If we move it downwards as indicated by the arrow, the integral (2.136) becomes undefined when the pole crosses the contour of integration.

may depend on the parameters. Now we will proceed to indicate how such integral may give rise to branch points.

Consider first the following integral

$$I(z_0) = \int_0^1 dx \frac{1}{x - z_0}. \tag{2.136}$$

When z_0 is a real number between 0 and 1 (including the endpoints) then the integral (2.136) is not defined, because at $x = z_0$ the integrand $1/(x-z_0)$ becomes infinite. However, for any other complex value of z_0 the integral (2.136) is well-defined and finite. In fact, in this simple case it is not difficult to obtain the explicit solution,

$$I(z_0) = \log\left(\frac{z_0 - 1}{z_0}\right), \quad z_0 \notin [0, 1]\,. \tag{2.137}$$

This result is simple but we can learn a lot from it if we consider it carefully. Notice, first, the following two things:

- The integral (2.136) is a well-defined, single-valued function of z_0 on $\mathbb{C} - [0, 1]$. The complex logarithm, which figures in its explicit solution (2.137), is however not single-valued and has one branch point. Therefore, in order for (2.137) to really be the solution of the integral (2.136) we must pick the correct branch of the logarithm. The integral (2.136) is not defined for $z_0 \in [0, 1]$, and the argument $(z_0 - 1)/z_0$ of the logarithm in (2.137) for such values of z_0 assumes values from the interval $(-\infty, 0]$. This means that we must choose a branch of the logarithm with the branch cut on the negative real axis. (Because then the branch cut of the logarithm will be restricted to those values of z_0 for which (2.136) is not defined, and the solution (2.137) will be well-defined and single-valued on $\mathbb{C} - [0, 1]$.) Furthermore, considering that for real $z_0 < 0$ the integral (2.136) must also be real, the correct choice turns out to be the principal branch of the logarithm

$$\log z = \log|z| + i\phi, \text{ with } \phi \in (-\pi, \pi]. \tag{2.138}$$

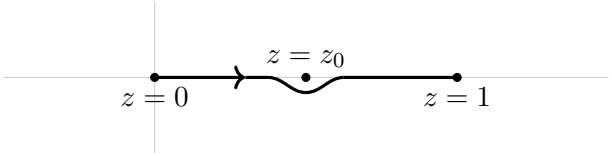


Figure 2.2: An illustration of a continuation of $I(z_0)$, defined by the integral (2.136), to a point on the original contour of integration. For this to be possible, we first need to slightly deform the original contour before we move the pole onto the interval $[0, 1]$. In this way it avoids the pole as the pole is moved to its target position at z_0 as depicted in the figure. (Note that the original, undeformed contour is depicted in Fig. 2.1.)

- When considered as a function of its parameter z_0, the result $I(z_0)$ is analytic everywhere where the integral (2.136) is defined. That is, the solution (2.137) is an analytic function on $\mathbb{C} - [0, 1]$. This is not a surprise, because the integrand $1/(x - z_0)$ is an analytic function of z_0 everywhere except for its one simple pole at $z_0 = x$, and for $z_0 \notin [0, 1]$ it is integrated, as a function of x, over a compact contour that does not cross the singularity.

Using the branch of the logarithm defined in (2.138) the solution (2.137) correctly describes the integral (2.136) everywhere on $\mathbb{C} - [0, 1]$. Notice, however, that (2.137) is singular only at $z_0 = 0$ and $z_0 = 1$. We can therefore analytically continue the solution (2.137) to and beyond all the points on the interval $(0, 1)$. The maximal such a continuation leads to a Riemann surface with an infinite number of sheets, directly inherited from the Riemann surface on which the logarithm is defined. The interval $[0, 1]$ then becomes the branch cut on the first Riemann sheet, which corresponds to the integral (2.136).

Now we are ready to ask the central question of this discussion. Is it possible to interpret this analytic continuation in terms of the integral (2.136)? The answer is yes, and, in fact, we have already discussed this kind of interpretation briefly in Section 1.4. Consider the integral (2.136) with z_0 positioned just above the contour of integration. (See Fig. 2.1.) As we push the singularity at z_0 downwards, the integral becomes undefined as soon as the singularity touches the integration contour. But recall that as a consequence of the Cauchy's theorem, any continuous deformation of the integration contour (with the endpoints fixed) does not change the value of the integral as long as it does not cross any singularity of the integrand. Thus, as z_0 approaches the line $[0, 1]$ we can deform the contour of integration slightly downwards in order to avoid the singularity. (See Fig. 2.2.) This defines an analytic continuation of (2.136). That is, in terms of the integral (2.136), an analytic continuation of $I(z_0)$ corresponds to an appropriate deformation of the integration contour: as the location of the singularity at

z_0 moves over the complex plane it pushes the deformation contour away. Notice that this means that the location of the branch cut of an integral, such as the branch cut of the function $I(z_0)$, is determined by the definition of the original contour of integration, which in the case of $I(z_0)$ is given by (2.136) as the interval $[0, 1]$. It is also important to recognize, however, that it is only branch points that have the real physical significance. Branch cuts must originate in branch points but otherwise they are quite arbitrary. It is up to us to choose how to draw them, and different choices correspond to different descriptions of the same mathematical structure. In this sense the choice of a branch cut is somewhat similar to the choice of coordinates.

This point of view makes it also quite intuitive that the points $z_0 = 0$ and $z_0 = 1$ are real singularities of $I(z_0)$. This is because the endpoints of the integration contour are fixed, and we therefore cannot avoid crossing the singularity if z_0 moves to either one of the endpoints. This kind of a singularity of an integral is called the *endpoint singularity*.

Note that although in the simple case of (2.136) we could discern the complete analytic structure of $I(z_0)$ directly from its explicit solution (2.137) and (2.138), in more complicated cases an explicit solution may either be unknown to us or it may be known but too complex to study directly. It is then easier to study the problem in terms of the methods that we just introduced. That is, it may be easier to study the analytic structure in terms of the positions of singularities and allowed contour deformations in the complex space of integration.[32]

This representation of analytic continuations of integrals in terms of contour deformations is also useful in situations when we need to calculate discontinuities across branch cuts. We will illustrate this method on the

[32]Let me add two remarks. First, I think that the more important reason why this formalism is superior to the direct analysis of explicit solutions is that it is universal. Even if we were able to find explicit solutions to each problem of interest, we would (presumably) do so for each such a problem, or at least a class of problems, on an individual basis. That makes it difficult to formulate and prove general observations. (What does the solution (2.137) tell us about general Feynman integrals?) On the other hand, the representation in terms of contour deformations in the integration space is completely general and well-suited for discovery of universal mechanisms behind the structure of singularities. One could argue that it allows us to concentrate on the underlying physics without being distracted by specific details of each particular problem. An important example of this is the reformulation of the Landau conditions in terms of a classical propagation of particles due to Coleman and Norton. This, however, brings me to the second remark. Most of the examples that I provide in this chapter are in a single dimensional complex space. These cases are simple to visualize and one can relatively easily construct very persuasive and transparent proofs of various statements related to this approach. (Such as for instance the sufficiency of Landau conditions for the existence of a pinch in the physical region.) Realistic problems, however, are usually formulated in higher dimensional spaces and working with those is much more difficult. The usual treatments currently present in the physics literature do not really address this issue and as a result much of the transparency of the single-dimensional examples is lost once the methods are applied in more realistic settings.

example (2.136), but before we do so, let us first make use of the fact that in this case we know the explicit solution (2.137), (2.138). We will start by calculating the discontinuity directly from this solution, and only afterwards will we introduce and apply the general method that makes use of contour deformations.

Consider two close values of z_0 of the form

$$z_0^{(\pm)} = x \pm i\epsilon \tag{2.139}$$

where $0 < x < 1$ and ϵ is small and positive. That is, $z_0^{(+)}$ lies right above the branch cut while $z_0^{(-)}$ is located below $z_0^{(+)}$ on the other side of the cut. The discontinuity across the branch cut corresponds to the difference in the values of I at those two points in the limit of $\epsilon \to 0^+$. The explicit solution (2.137) leads to

$$I(x \pm i\epsilon) = \log\left(\frac{x \pm i\epsilon - 1}{x \pm i\epsilon}\right) = \log\left((x \pm i\epsilon - 1)(x \mp i\epsilon)\right) - \log(x^2 + \epsilon^2). \tag{2.140}$$

If we retain only those ϵ-terms that are relevant in the limit of $\epsilon \to 0^+$ this can be written as

$$I(x \pm i\epsilon) = \log\Big(x^2 - x \pm i\epsilon\Big) - 2\log x. \tag{2.141}$$

We use the principal branch of the logarithm (2.138), and in the limit $\epsilon \to 0^+$ we therefore obtain $I(x + i\epsilon) = \log((1-x)/x) + i\pi$ and $I(x - i\epsilon) = \log((1-x)/x) - i\pi$. The discontinuity across the branch cut is therefore

$$I(x + i\epsilon) - I(x - i\epsilon) = 2\pi i. \tag{2.142}$$

Of course, this is just the usual logarithmic discontinuity, but we needed to be careful with the sign due to the non-trivial relation between z_0 and the argument of the logarithm.

Our task now is to obtain an interpretation of this branch cut discontinuity in terms of deformations of the integration contour. The discontinuity (2.142) corresponds to the difference between the values of the function I on the opposite sides of the branch cut. Of course, both these values are taken from the same sheet. But the same discontinuity can also be interpreted as the difference between the values taken on the same side of the branch cut but from different branches! This is because if we start at $x + i\epsilon$ and then move the distance 2ϵ in the negative y-direction across the branch cut, we arrive at the point $x - i\epsilon$ but on the neighboring branch. Let us call this neighboring branch the second branch. Due to the continuity of analytic continuation, the value of $I(x - i\epsilon)$ on the second branch is essentially the same as that of $I(x + i\epsilon)$ on the first branch, at least in the limit of $\epsilon \to 0^+$. Thus, the equation (2.142) tells us that the value of $I(x - i\epsilon)$ on the second

Figure 2.3: *An illustration of how to calculate the discontinuity across the branch cut.* If we start with the original contour of integration according to (2.136) and with z_0 just above the contour as depicted in Fig. 2.1, but then move the singularity across the branch cut into the lower half-plane, we need to push the contour away to avoid the singularity. This is depicted on the left-hand side. That integral therefore corresponds to an analytic continuation from the upper half-plane through the branch cut $[0, 1]$. The value of this continuation is not the same as the value at the same z_0 but on the original sheet — that is, the value of the integral when z_0 is as in the present figure but the contour is the original one, as in Fig. 2.1. To calculate the difference between those two values one just needs to notice that according to the Cauchy's theorem the integral on the left is the same as the integral on the right-hand side. Consequently, the difference between the values on two neighboring sheets is given by the residue of the singularity.

branch, which is given by $I(x + i\epsilon)$ on the first branch, minus $I(x - i\epsilon)$ on the first branch equals $2\pi i$.

When stated this way, the discontinuity equation (2.142) can be interpreted in terms of contour deformations quite easily. This is illustrated in Fig. 2.3. The value of $I(x-i\epsilon)$ on the second sheet can be obtained by starting at $x+i\epsilon$ on the first sheet and moving slightly in the negative y-direction across the branch cut. For this we need to deform the integration contour to avoid the singularity, as depicted on the left-hand side of Fig. 2.3. The integral depicted on that side of the figure therefore corresponds to $I(x-i\epsilon)$ on the second sheet. As we can see, on this sheet the integration contour goes below the singularity. In order to calculate the value of $I(x - i\epsilon)$ on the first sheet we would need to use the original contour of integration, as depicted in Fig. 2.1. On this sheet the contour passes above the singularity. Using the standard manipulations of integration contours and the Cauchy's theorem[33] we can easily see that the difference between these two integrals

[33]To obtain the integral over two contours on the right-hand side of the Fig. 2.3 from the integral on the left-hand side of the same figure, it is enough to add to the deformed contour on the left-hand side the "missing" section of the x-axis above the singularity and integrate over it there and back. (Integrating there and back evaluates to zero and the value of the overall integral therefore remains unchanged.) Now we just reconsider this new contour as consisting of two parts: The first part is composed from those parts of the original contour on the left-hand side that lied on the x-axis as well as the new part integrated in the positive direction of the x-axis. This corresponds to the original contour of integration from Fig. 2.1. The second part consist of the parts of the contour on the left-hand side that were not on the x-axis — that is, of the curve below the x-axis

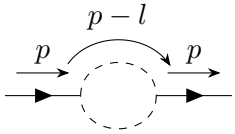


Figure 2.4: A single-loop diagram in a ϕ^3 theory of a real scalar field.

is just the integral over a small circle taken in the counterclockwise direction around the singularity, as depicted on the right-hand side of Fig. 2.3. This means that the difference is $2\pi i$ times the residue of the singularity, which in our case of the integrand $1/(z - z_0)$ is just $2\pi i$. This confirms our earlier calculation (2.142). This method can be generalized and leads to the *Cutkosky rules* for the calculation of discontinuities. (See, e.g. [1].)

We have now reached the point that we have covered everything what we planned to discuss regarding the branch structure of integrals of analytic functions. The example (2.136) allowed us to discuss branch cuts in a very simple setting, but its disadvantage was that it did not possess a structure typical of a Feynman integral. For this reason we will now very briefly discuss a simple example of a Feynman integral. In a theory of a real scalar field ϕ describing a particle of mass m, consider the one-loop diagram depicted in Fig. 2.4, which contains two ϕ^3 interaction vertices and from which we amputate the two external propagators. The loop momentum l flows along the loop in the counterclockwise direction. Except for various constant factors that are currently of no interest to us, the diagram contributes the value[34]

$$I(p) = \int d^4l \frac{1}{l^2 - m^2 + i\epsilon} \frac{1}{(p-l)^2 - m^2 + i\epsilon}. \tag{2.143}$$

This is an example of a simple Feynman integral. An adequate treatment of the analytic structure of this integral would require us to investigate singularity structure and contour deformations in the full complex space of the loop momenta l, which is four complex dimensional or eight real dimensional. For our illustrative purposes, however, we will fix the value of the space components of l at some particular point $\mathbf{l}$ and consider only the integral over the single remaining loop variable, the energy l^0. We will discuss the singularity structure of

$$\mathcal{I}(p, \mathbf{l}) = \int dl^0 \frac{1}{(l^0)^2 - E(\mathbf{l})^2 + i\epsilon} \frac{1}{(p^0 - l^0)^2 - E(\mathbf{p} - \mathbf{l})^2 + i\epsilon}, \tag{2.144}$$

that avoids the singularity — and the new part but integrated in the negative x-direction. Using the Cauchy's theorem this second contour can be deformed — without changing the value of the integral — into the small circle around the singularity, as depicted on the right hand side of Fig. 2.3.

[34]This is a logarithmically divergent integral, but because we are now interested only in singularities that correspond to finite values of loop momenta, we will ignore this issue. Of course, it could be addressed by the usual methods of renormalization.

where $E(\mathbf{k}) = \sqrt{\mathbf{k}^2 + m^2}$ is the energy of a particle of mass m and three-momentum $\mathbf{k}$. Of course, in terms of $\mathcal{I}$ the full integral I can be expressed as

$$I(p) = \int d^3\mathbf{l}\, \mathcal{I}(p, \mathbf{l}),$$

but here we will limit our attention only to the integral (2.144). The integration contour in that integral is the whole real axis of l^0, and the integrand has four singularities, all of them simple poles, at the points[35]

$$l^0 = E(\mathbf{l}) - i\epsilon, \tag{2.145}$$
$$l^0 = -E(\mathbf{l}) + i\epsilon, \tag{2.146}$$
$$l^0 = p^0 - E(\mathbf{p} - \mathbf{l}) + i\epsilon, \tag{2.147}$$
$$l^0 = p^0 + E(\mathbf{p} - \mathbf{l}) - i\epsilon. \tag{2.148}$$

For the case when both p^0 and $E(\mathbf{p} - \mathbf{l})$ are real the situation is illustrated in Fig. 2.5. The first two poles, at $\pm(E(\mathbf{l}) - i\epsilon)$, do not depend on the parameter p and are always positioned just below and above the integration contour, respectively. Note that on their own they would never give rise to a singularity. If they were the only poles present, it would be always possible to deform the integration contour such as to avoid them as $\epsilon \to 0^+$, because there is a finite separation between them in the real-l^0 direction.[36] This changes when we also consider the other two poles. The discussion of these other two poles is complicated by the fact that all the components of p are allowed to be complex. Fortunately, we can use the Lorentz invariance property[37] of the integral $I(p)$, which implies that I must be a scalar function of the only invariant that is present in the problem, the square p^2. (The mass m is another such an invariant, but we consider it fixed now.) Without a loss of generality we may therefore suppose that $\mathbf{p}$ is real, and let p^0 supply the imaginary component of p^2 if needed. In this case we may again use the property that $E(\mathbf{p}-\mathbf{l}) > 0$ to argue that the two poles at $p^0 \pm (E(\mathbf{p}-\mathbf{l}) - i\epsilon)$ cannot create a pinch on their own. The only way a pinch can appear is either between the pole at $E(\mathbf{l}) - i\epsilon$ and the pole at $p^0 - E(\mathbf{p} - \mathbf{l}) + i\epsilon$, or

[35] Recall that the "$i\epsilon$"-terms are meant to indicate the limit $\epsilon \to 0^+$ and their precise value is therefore immaterial. For instance, the following is valid

$$(l^0)^2 - E(\mathbf{l})^2 + i\epsilon = (l^0)^2 - (E(\mathbf{l}) - i\epsilon)^2\,,$$

because the energy is positive.

[36] Here we assume $m > 0$, which means that $E(\mathbf{l}) > 0$ for all $\mathbf{l}$. If the scalar particle was massless, then there would be a pinch at $\mathbf{l} = 0$, $l^0 = 0$, leading to a so-called soft singularity.

[37] For real values of p the Lorentz invariance of (2.143) is easy to show. The extension of this property to complex values of p, however, requires a little bit more work because the integration contour is real. Nonetheless, as we mentioned in Sec. 2.3, one can show that the desired property holds quite generally by considerations of transformation properties under proper complex Lorentz transformations.

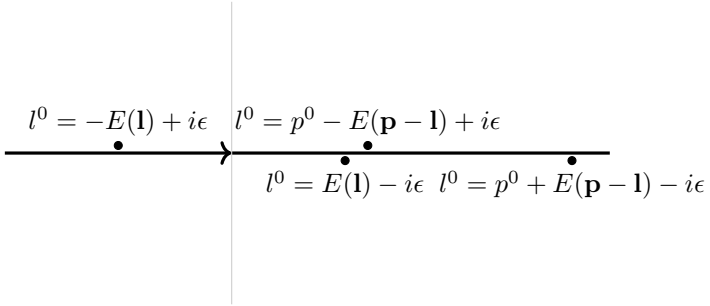


Figure 2.5: The pole structure of the integrand of (2.144) when both $\mathbf{p}$ and p^0 are real.

between the pole at $-E(\mathbf{l}) + i\epsilon$ and the pole at $p^0 + E(\mathbf{p}-\mathbf{l}) - i\epsilon$. For this to happen p^0 must be real. Furthermore, the two poles that participate in the pinch must be located right on top of each other, with the integration contour running in between them. This means that if the pinch is between the poles at $E(\mathbf{l}) - i\epsilon$ and $p^0 - E(\mathbf{p}-\mathbf{l}) + i\epsilon$, then the pinch appears at $l^0 = E(\mathbf{l}) = p^0 - E(\mathbf{p}-\mathbf{l})$. This implies that $l^0 = E(\mathbf{l})$ and $p^0 - l^0 = E(\mathbf{p}-\mathbf{l})$. If, on the other hand, the pinch is between $-E(\mathbf{l}) + i\epsilon$ and $p^0 + E(\mathbf{p}-\mathbf{l}) - i\epsilon$, then $l^0 = -E(\mathbf{l}) = p^0 + E(\mathbf{p}-\mathbf{l})$, or $-l^0 = E(\mathbf{l})$ and $-(p^0 - l^0) = E(\mathbf{p}-\mathbf{l})$. Thus, as we can see, the second case corresponds to the reversal of all time components in the first case. For this reason we will not address the second case separately.

The condition that $l^0 = E(\mathbf{l})$ and $p^0 - l^0 = E(\mathbf{p}-\mathbf{l})$ says that both the virtual particles in the loop are on the mass-shell. If we considered also the space dimensions of l we would find that most of these solutions are not real pinches — the full integration contour can usually avoid singularities by suitable deformations into the imaginary directions of the space components of $\mathbf{l}$. It turns out that only the lowest-energy solution (for a given value of $\mathbf{p}$) of this condition corresponds to the real pinch. That is, the pinch occurs only when $\mathbf{l} = \mathbf{p} - \mathbf{l}$, which implies that $\mathbf{p} = 2\mathbf{l}$, $p^0 = l^0 + E(\mathbf{l}) = 2E(\mathbf{l})$, and therefore $p^2 = 4m^2$. In fact, this singularity is a branch point and corresponds to a normal threshold.

[Let us briefly illustrate how to determine if a pinch is present. Feynman integrals usually entail integration over several loop variables, so instead of a single-dimensional contour of integration we integrate over a multi-dimensional surface. When we integrate over $4L$ loop variables then the singularities typically form surfaces of $4L-1$ complex dimensions, instead of being located at single points as they were in our single-dimensional examples. This is because they can typically be characterized as solutions of some particular equations of the form $\chi(l) = 0$, where l represents all $4L$ integration variables. Recall that our main goal is to investigate points on the integration surface that become singular when the "$i\epsilon$"-terms in denom-

inators of the propagators decrease to zero. It is often possible to avoid such singularities by suitable deformations of the integration surface. This is actually more subtle that may appear at a first glance and it will prove worthwhile to describe the process of avoiding singularities more carefully. While doing so, we will also introduce some useful terminology.[38]

At the beginning the parameter ϵ is fixed at some positive value and the surface of integration has not yet been deformed in any way. For the integrals that concern us here we assume that for this positive value of ϵ none of the singular surfaces crosses the original integration surface, that is, the integrand is regular everywhere. However, when ϵ is set to zero, some of the parts of the surface of integration might become singular. If this happens we need to consider the possibility that the integral can be analytically continued to that point,[39] so that the apparent singularity can be removed. For this reason, our next step — with ϵ still fixed and positive — is to continuously deform the integration surface away from the singularities. However, in order for us to be able to use the Cauchy's theorem to argue that the integral over the original surface of integration has the same value as the integral over the deformed surface, we must make sure that at no point during the process of deformation the surface crosses any singularity. We say that such a deformation is *allowed*. (That is, allowed deformations do not cross any singular surfaces, and consequently the integral over final such a deformed surface has the same value as the original integral.) Now we are ready for the next step: we continuously decrease the value of $\epsilon > 0$ to zero. If no singularity appears on the (deformed) integration surface during this process we say that the deformed surface *avoids* the singularities. To summarize, in order to analytically continue the integral we need to find an allowed deformation of the integration surface that avoids singularities. If this is not possible we say that the integration surface is *pinched*.

We define the *physical region* of a Feynman integral by requiring that the original (undeformed) surface of integration is the real hyperplane (corresponding to real values of the components of loop momenta) and that the external momenta and all the masses are real. The main result is that the so-called *Landau conditions* are both sufficient and necessary for the existence of a pinch in the physical region. (We already mentioned this in Sec. 1.4. A textbook treatment can be found in [1]. A more careful discussion is provided in [6].)

As we already specified, we collect all $4L$ loop variables into a single (vector) variable l. We consider Feynman integrals of the form

$$I(p) = \int d^{4L}l \frac{1}{[A_1(l;p) + i\epsilon] \dots [A_n(l;p) + i\epsilon]}, \tag{2.149}$$

[38] This terminology, as well as the overall presentation, are motivated by [6].

[39] At this level of detail we can think about this either as an analytic continuation in the external parameters, or also as an analytic continuation in ϵ.

where p represents all the external parameters (in principle including also masses). We assume that the denominator factors A_1, ..., A_n are analytic functions of l and p everywhere, and that their zero surfaces are given by equations $\chi_j(l;p) = 0$, $j = 1, \ldots, n$. For $\epsilon = 0$ the zero surfaces of the denominator factors correspond to singular surfaces of the integrand.

The Landau conditions say that (for some fixed value of p) the integration surface is pinched at $l = l_0$ if and only if there is a subset of the denominator factors — say, for simplicity, that this subset consists of the factors A_1, ..., A_m for some $m \leq n$ — such that all of those factors vanish at l_0,

$$A_1(l_0;p) = \cdots = A_m(l_0;p) = 0, \tag{2.150}$$

and that there are numbers $\alpha_1 \geq 0$, ..., $\alpha_n \geq 0$, not all zero, such that

$$\sum_{i=1}^{m} \alpha_i \, \partial_l A_i(l_0;p) = 0. \tag{2.151}$$

Note that $\partial_l A_i$ are best understood as one-forms. That is, when they act on a $4L$-dimensional vector a, whose components we denote a_k, then the result is

$$\left[\partial_l A_i\right](a) = a \cdot \partial_l A_i = \sum_{k=1}^{4L} a_k \frac{\partial A_i}{\partial l_k}. \tag{2.152}$$

Therefore, we may interpret the second condition (2.151) as simply saying that there exists a vanishing convex combination of the one-forms $\partial_l A_1$, ..., $\partial_l A_m$.

Note that even if there is a pinch at some point $l = l_0$, not all of the denominator factors must participate in it. First of all, as was already said, some of them may not even vanish at $l = l_0$. But even those that do vanish — in our notation A_1, ..., A_m — might not all participate. This happens when there exists an allowed deformation that avoids singularities of some of those factors. (But not all of them, because we assume that there is a pinch.) Factors whose singularities can be avoided in this manner are those for which there is no solution of (2.151) with the corresponding α being non-zero positive. Therefore, we say that the denominator factor A_j *participates* in the pinch at $l = l_0$ if and only if $A_j(l_0;p) = 0$ *and* there is a solution of $\sum_{i=1}^{m} \alpha_i \partial_l A_i(l_0;p) = 0$ with $\alpha_j > 0$ and all other $\alpha_i \geq 0$, and where A_1, ..., A_m represent all the denominator factors that vanish at $l = l_0$.

In the integral (2.143) the denominator factors are $A_1(l;p,m) = l^2 - m^2$ and $A_2(l;p,m) = (p-l)^2 - m^2$. The conditions that these factors vanish, $A_1 = 0$ or $A_2 = 0$, are easy to interpret: the virtual particle that corresponds to the vanishing denominator factor must be on the mass-shell. Furthermore, we have $\partial_l A_1 = 2l$ and $\partial_l A_2 = 2(l-p)$. Is it possible for either one of these denominator factors to pinch the integration surface on its own? If A_1 was to pinch the surface on its own (without A_2 participating) the two equations

$A_1 = 0$ and $\partial_l A_1 = 0$ would need to be satisfied simultaneously. (The second equation follows from the condition that $\alpha_1 \partial_l A_1 = 0$ for $\alpha_1 > 0$.) But if $\partial_l A_1 = 0$ then $l = 0$, which means that $A_1 \neq 0$. The case of A_2 is similar. (Recall that we assume that $m > 0$. Otherwise, there indeed would be a soft pinch at $l = 0$ from the denominator factor A_1, and another soft pinch, at $l = p$, from the denominator factor A_2.) Hence, to create a pinch, the two denominators A_1 and A_2 must both participate. The condition that $\alpha_1 2l + \alpha_2 2(l - p) = 0$, with both α_1 and α_2 positive, then requires that

$$l = \frac{\alpha_2}{\alpha_1 + \alpha_2} p. \tag{2.153}$$

But this implies that also

$$p - l = \frac{\alpha_1}{\alpha_1 + \alpha_2} p, \tag{2.154}$$

and if both l and $p-l$ are to be on the mass-shell, we necessarily have (recall that both α_1 and α_2 are positive)

$$\alpha_1 = \alpha_2. \tag{2.155}$$

But this implies that $l = p/2$, and since $l^2 = m^2$ by the first Landau condition, we find that the pinch appears if

$$p^2 = 4m^2. \tag{2.156}$$

In fact, one can show that if there is no solution of $\sum_j \alpha_j \partial_l A_j = 0$ with $\alpha_j \geq 0$ not all zero, then there exists a vector v such that $v \cdot \partial A_j > 0$ for all j. (See [6].) In such a case, if we deform the integration surface continuously in such a way that at l_0 the surface is moved to $l_0 + iv\delta$, then for a sufficiently small $\delta > 0$ the deformation is allowed, and when $\epsilon \to 0^+$ neither of the denominator factors $A_j + i\epsilon$ vanishes on the deformed surface of integration. Hence, we escape the pinch. If presented in a greater detail, this would serve as a proof that the Landau conditions are necessary for the existence of a pinch. However, the implication in the other direction, that in the physical region the Landau conditions are also sufficient, seems to be much more difficult to prove.

And lastly, it may useful to return to our discussion of the integral (2.144), where the integration has been restricted only to the l^0-component with the other three components of l fixed, and show how to choose a suitable vector v to escape the pinch that seems to appear in the l^0-plane between the two denominators A_1 and A_2 of (2.143). In that discussion we found that in the l^0-plane the pinch occurs only if $E(\mathbf{l}) = E(\mathbf{p} - \mathbf{l})$. That means that in the real three-dimensional space the vectors $\mathbf{l}$ and $\mathbf{p} - \mathbf{l}$ must be of the same length. The vectors $\mathbf{p}$, $\mathbf{l}$ and $\mathbf{l} - \mathbf{p} = -(\mathbf{p} - \mathbf{l})$, when translated appropriately, form a triangle in which $\mathbf{l}$ and $\mathbf{l} - \mathbf{p}$ start at different endpoints of $\mathbf{p}$

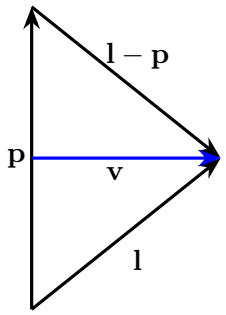


Figure 2.6: An illustration of how to choose the vector $v = (0, \mathbf{v})$ to escape the apparent pinch in the l^0-plane of (2.144). The pinch truly appears only when $\mathbf{l} = \mathbf{p}/2$, in which case the vector $\mathbf{v}$ illustrated above is of zero length.

and end at the same point. (See Fig. 2.6.) This configuration is possible only if their length is at least $|\mathbf{p}|/2$ each, in which case they are both parallel to $\mathbf{p}$ and meet in the middle of that vector. (This, of course, corresponds to the minimal energy solutions, and to the only case when the pinch is really present. In this case a suitable v does not exist.) If their lengths are larger than $|\mathbf{p}|/2$ then a full non-degenerate triangle is formed. In those cases, if we define $\mathbf{v}$ to be the vector that is orthogonal to $\mathbf{p}$ and spans the interval from the middle point of $\mathbf{p}$ to the vertex where $\mathbf{l}$ and $\mathbf{l} - \mathbf{p}$ meet, then $v = (0, \mathbf{v})$ satisfies $v \cdot \partial A_1 = v \cdot 2l > 0$ and $v \cdot \partial A_2 = v \cdot 2(l - p) > 0$. Deforming the integration surface (near the singularity) in the direction iv therefore escapes the pinch.]

The discussion presented here does by no means provide an adequate treatment of the general singularity structure of Feynman integrals. Unfortunately, such a complete treatment would take us far too away from our present goals and we will therefore skip it. A reader interested in this topic can find a good exposition of the relevant techniques for instance in [1, 6]. The remaining part of this section is devoted to a brief discussion of a special case of a branch point, the normal threshold.

2.5.2 Unitarity and normal thresholds

As we are going to see, the rough idea of how the unitarity implies the existence of normal thresholds is quite straightforward. It would be, however, much more difficult to supply all the missing details and to state the result in a reasonable generality. For this reason we will not provide here such a complete treatment. Rather, we will present a simple but very limited argument that makes the main idea apparent. Then we will briefly discuss some of the weak points and inadequacies of that explanation, but for a more complete treatment the reader is advised to consult [1].

One last remark before we begin. Although our discussion is going to be

limited only to their appearance in the S-matrix, the normal thresholds exist also in other matrix elements of products of field operators. For example, to study them in general time-ordered Green's functions we could employ an approach similar to what we used in Sec. 2.4 when we studied particle poles.

In our argument we are going to present normal thresholds as a consequence of the unitarity of the S-matrix. Recall that the S-matrix element $S_{\beta\alpha}$ is defined as

$$S_{\beta\alpha} = \left(\Psi_\beta^{(\mathrm{out})}, \Psi_\alpha^{(\mathrm{in})}\right), \tag{2.157}$$

where $\Psi_\beta^{(\mathrm{out})}$ is the asymptotic "out"-state consisting of particles described by the multi-index β, and $\Psi_\alpha^{(\mathrm{in})}$ is the asymptotic "in"-state whose particle content is specified by the multi-index α. The unitarity relation for $S_{\beta\alpha}$ then follows from the completeness of the asymptotic states,[40]

$$\begin{aligned}\int d\gamma\; S^*_{\gamma\beta} S_{\gamma\alpha} &= \int d\gamma \left(\Psi_\beta^{(\mathrm{in})}, \Psi_\gamma^{(\mathrm{out})}\right)\left(\Psi_\gamma^{(\mathrm{out})}, \Psi_\alpha^{(\mathrm{in})}\right) \\ &= \left(\Psi_\beta^{(\mathrm{in})}, \Psi_\alpha^{(\mathrm{in})}\right) = \delta\left(\beta - \alpha\right). \end{aligned} \tag{2.158}$$

Here we assumed that the asymptotic states were normalized "to delta functions", as they were in Sec. 2.4. The symbol $\delta\left(\beta - \alpha\right)$ represents the appropriate product of delta functions in continuous labels that are present in the multi-indices α, β, and of the Kronecker delta symbols in their discrete labels. Also, just as before, the symbol "$\int d\gamma$" involves integration over all the continuous labels together with sums over all the discrete labels in γ. (It usually also involves extra combinatorial or other factors, depending on what exact normalization condition we employ for multi-particle states. Of course, these factors are fixed by the condition $\int d\gamma \left|\Psi_\gamma\right\rangle \left\langle\Psi_\gamma\right| = \mathbf{1}$.) The symbol $S^*_{\gamma\beta}$ stands for the complex conjugate of $S_{\gamma\beta}$. In a similar way, but this time using the completeness of the asymptotic "in"-states, we can also obtain

$$\int d\gamma\; S_{\beta\gamma} S^*_{\alpha\gamma} = \delta\left(\beta - \alpha\right). \tag{2.159}$$

There is an important property of the S-matrix related to its connectedness structure, which we have not yet mentioned: the S-matrix is assumed to satisfy the *cluster decomposition principle*. This roughly corresponds to the basic assumption which is implicitly present in the scientific method and which says that the experiments that we carry out here and now are (in some sense) independent of what is happening elsewhere in the universe

[40] This assumption is called the *asymptotic completeness* and it states that the asymptotic states $\{\Psi_\alpha^{(\mathrm{out})}\}$, as well as the states $\{\Psi_\alpha^{(\mathrm{in})}\}$, form a complete basis for physical states.

at a distant location.[41] Roughly speaking, for the S-matrix element $S_{\beta\alpha}$ the cluster decomposition principle says that if the "in"-state described by the multi-index α and the "out"-state described by β consist of particles that can be divided into two separate sets, one described by the multi-indices α_1 and β_1 and the other by α_2 and β_2, such that all the particles in the first set are far away from all the particles in the second set, then as we gradually increase the distance between these two sets of particles the S-matrix factorizes,

$$S_{\beta\alpha} \to S_{\beta_2\alpha_2} S_{\beta_1\alpha_1}. \tag{2.160}$$

Note that in order to speak about large space-like separations of various subsets of particles in the asymptotic states one needs to pass from the momentum representation to the coordinate representation. Furthermore, any complete statement of (2.160) would need to specify what kind of convergence is implied by the "$\to$" sign in that condition. We will not go into those details here. Instead, what we would at this point like to take away from (2.160) is that a coordinate-representation S-matrix that describes particles interacting in several clusters that are separated from each other by large space-like intervals factorizes into a product of S-matrices, one for each such a cluster. This implies that the scattering in each such a cluster is independent of the processes that happen in the other clusters. Furthermore, since the asymptotic states each contain a finite number of particles, this kind of factorization cannot continue indefinitely, and for any given process we must eventually reach a point when none of the clusters can be further decomposed into widely separated sub-clusters. Within such irreducible "atomic" clusters each particle interacts with all other particles in that same cluster.[42] The S-matrices that correspond to processes occurring in such irreducible clusters are called *connected* S-matrices.

In the momentum representation the S-matrix elements carry no information about the space-time location of the individual particles, because those particles are of definite momenta. Nonetheless, the overall momentum-space S-matrix element can still be separated into the sum of terms, one for each possible partition of particles into irreducible clusters, and each such term consisting of an appropriate product of the momentum-space versions of connected S-matrices. Connected S-matrices expressed in momentum representation are characterized by the property that except for a single delta function that ensures the conservation of the overall four-momentum in the cluster they do not contain any other delta-function factors. In terms of

[41] This is a complex topic, and I will not try to address it in the present work. Limiting our attention to the cluster decomposition principle itself, a readable exposition of this property can be found, for instance, in Chapter 4 of [9].

[42] There are various issues that we would need to explicitly discuss in order to make this statement precise. For our purposes, let us interpret the phrase that 'each particle interacts with all other particles' as saying that in perturbation theory the reaction is calculated by considering only connected Feynman diagrams.

Feynman diagrams, the connected S-matrices are represented by the sums of all appropriate connected diagrams.

Let us illustrate this on a simple example of elastic scattering of three distinct particles labeled by numbers 1, 2, 3 when they are in the "in"-state and by 1', 2', 3' when they are in the "out"-state. Let us denote the overall momentum-space S-matrix element by $S\,(123 \to 1'2'3')$. This contains contributions from processes when all the particles interact with each other — that is, when the particles cannot be separated into several distant clusters — as well as from processes in which there exist two or more disconnected clusters. Let us denote the contribution from the connected part by $S^{\rm C}\,(123 \to 1'2'3')$. Next, we need to characterize all possible ways in which the particles can be partitioned into several clusters. We may start by considering processes in which the particle 3 is far away from all the other particles. This leads to the overall contribution $S(12 \to 1'2')S^{\rm C}(3 \to 3')$. Now, the term $S(12 \to 1'2')$ contains the contribution from the case when 1 and 2 interact, $S(12 \to 1'2')^{\rm C}$, as well as from the case when they are separated, $S^{\rm C}(1 \to 1')S^{\rm C}(2 \to 2')$. Thus, we have so far considered all the cases when the particle 3 is far away from particles 1 and 2, and also the case when all three particles form a single irreducible cluster. Hence, there are only two options remaining. Either the particle 3 interacts with the particle 1, in which case the particle 2 must be far away, or the particle 3 interacts with the particle 2, in which case the particle 1 is far away. These two cases correspond to terms $S^{\rm C}(13 \to 1'3')S^{\rm C}(2 \to 2')$ and $S^{\rm C}(23 \to 2'3')S^{\rm C}(1 \to 1')$, respectively. Altogether, we have therefore found that

$$\begin{aligned} S\,(123 \to 1'2'3') = S^{\rm C}\,(123 \to 1'2'3') + S^{\rm C}(12 \to 1'2')S^{\rm C}(3 \to 3') + \\ S^{\rm C}(23 \to 2'3')S^{\rm C}(1 \to 1') + S^{\rm C}(13 \to 1'3')S^{\rm C}(2 \to 2') \\ + S^{\rm C}(1 \to 1')S^{\rm C}(2 \to 2')S^{\rm C}(3 \to 3'). \end{aligned} \quad (2.161)$$

This provides an example of the decomposition of the S-matrix into the sum over all possible partitions into disconnected clusters. Each of the connected S-matrices on the right-hand side contains only a single delta-function factor. For instance, $S^{\rm C}\,(123 \to 1'2'3')$ contains $\delta^4(p_1+p_2+p_3-p_{1'}-p_{2'}-p_{3'})$ and no other delta-function factors, and $S^{\rm C}(13 \to 1'3')$ contains $\delta^4(p_1+p_3-p_{1'}-p_{3'})$ and no other delta functions. Single-particle S-matrices, such as $S^{\rm C}(1 \to 1')$, correspond to a single particle not interacting with any other particles. Under the usual definition of the relative phases of the "in"- and the "out"-states $S^{\rm C}(1 \to 1')$ is therefore equal to $\delta^4(p_1 - p_{1'})$ times the Kronecker delta symbols for the spin and species indices of the particle. The term $S^{\rm C}(1 \to 1')S^{\rm C}(2 \to 2')S^{\rm C}(3 \to 3')$ in (2.161) represents a special case of what we generally denote by $\delta(\beta - \alpha)$.

This structure of the S-matrix is very important in the study of its analytic structure, and if we decided to further study the results which will be mentioned at the end of this section we would make an extensive use of

it. For now, however, we are only going to need its simplest application: the widely-used decomposition of the S-matrix into the following two terms

$$S_{\beta\alpha} = \delta(\beta - \alpha) - 2\pi i\, \delta^4\left(p_\beta - p_\alpha\right) M_{\beta\alpha}. \tag{2.162}$$

Here the first term corresponds to the contribution of the scenario when all the particles are far away from each other and there is no interaction between them. The second term corresponds to all the effects of interaction between the particles. We have factored out the overall four-momentum conservation delta function. From our earlier discussion it then follows that the connected part of $M_{\beta\alpha}$ does not include any other delta-function factors, but in general we should expect there to be other parts of $M_{\beta\alpha}$, corresponding to processes when the particles interact in several disconnected clusters, that contain additional delta-function factors.

Using the decomposition (2.162), the unitarity condition (2.159) reads

$$\begin{aligned}\delta\left(\beta - \alpha\right) = \int d\gamma\ & \Big(\delta(\beta - \gamma) - 2\pi i\, \delta^4\left(p_\beta - p_\gamma\right) M_{\beta\gamma}\Big) \\ & \times \Big(\delta(\alpha - \gamma) + 2\pi i\, \delta^4\left(p_\alpha - p_\gamma\right) M^*_{\alpha\gamma}\Big) \\ = \delta(\beta - \alpha) & - 2\pi i\, \delta^4\left(p_\beta - p_\alpha\right)\Big(M_{\beta\alpha} - M^*_{\alpha\beta}\Big) \\ & + 4\pi^2 \delta^4\left(p_\beta - p_\alpha\right) \int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) M_{\beta\gamma} M^*_{\alpha\gamma}. \end{aligned} \tag{2.163}$$

For the cases when $p_\beta = p_\alpha$ this implies the condition

$$\frac{i}{2\pi}\Big(M_{\beta\alpha} - M^*_{\alpha\beta}\Big) = \int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) M_{\beta\gamma} M^*_{\alpha\gamma}. \tag{2.164}$$

When we apply the decomposition (2.162) to the other version of the unitarity condition, (2.158), we instead obtain

$$\frac{i}{2\pi}\Big(M_{\beta\alpha} - M^*_{\alpha\beta}\Big) = \int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) M^*_{\gamma\beta} M_{\gamma\alpha} \quad (\text{for } p_\alpha = p_\beta). \tag{2.165}$$

These are the unitarity conditions for the M-matrix.

The conditions (2.164), (2.165) become much more simple if we restrict our attention to the case of the forward scattering, $\beta = \alpha$. In that case we obtain the *optical theorem* for the M-matrix

$$-\frac{1}{\pi}\Im\left\{M_{\alpha\alpha}\right\} = \int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) \left|M_{\alpha\gamma}\right|^2 = \int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) \left|M_{\gamma\alpha}\right|^2. \tag{2.166}$$

The integral $\int d\gamma\, \delta^4\left(p_\gamma - p_\alpha\right) \left|M_{\gamma\alpha}\right|^2$ is proportional to the total rate of the reaction $\alpha \to$ 'anything', or, if α describes a two-particle state, to the total cross section, and the matrix element $M_{\alpha\alpha}$ is proportional to the forward

scattering amplitude.[43] The equation (2.166) therefore states that there is a direct correspondence between the imaginary part of the forward scattering amplitude and the total rate of reaction or the total cross section.

We will now argue that the equation (2.166) suggests the existence of normal thresholds in $M_{\alpha\alpha}$. To make the presentation more simple, suppose that there is only one kind of particle present in the theory, and that those particles are of spin zero and have the mass $m > 0$. We will consider states α that consist of two particles and are described in their center-of-mass frame. As a consequence of the rotational symmetry, the only variable on which

[43] It does not really concern us here what are the exact proportionality constants. However, for the reader's convenience, we will collect here all these relations in full detail. We continue using the convention (of, for instance, [9]) that asymptotic states are normalized to delta functions: $\left(\Psi_\beta^{(\mathrm{in})}, \Psi_\alpha^{(\mathrm{in})}\right) = \delta(\beta - \alpha)$, etc. The differential rate of the reaction $d\Gamma(\alpha \to \beta)/d\beta$ from the state described by α into any of the states inside the infinitesimal volume $d\beta$ near the state β, also known as the *differential transition rate*, can be calculated as

$$\frac{d\Gamma(\alpha \to \beta)}{d\beta} = (2\pi)^{3N_\alpha - 2} V^{1-N_\alpha} \left|M_{\beta\alpha}\right|^2 \delta^4(p_\beta - p_\alpha), \tag{2.167}$$

where N_α is the number of particles in the state α, and V is the volume of the system ("box") in which the reaction takes place. (Note that the rate here corresponds to probability per unit of time.) If $N_\alpha = 2$ we can think of the reaction probability in terms of the cross section. The differential cross section $d\sigma/d\beta$ can then be given as

$$\frac{d\sigma(\alpha \to \beta)}{d\beta} = (2\pi)^4 \frac{1}{u_\alpha} \left|M_{\beta\alpha}\right|^2 \delta^4(p_\beta - p_\alpha), \tag{2.168}$$

where u_α is the velocity of either of the two incoming particles in the frame in which the other particle is at rest. (The generalization to other frames is fixed by the usual convention that the cross section, when summed over the spins of all particles, is a Lorentz invariant quantity. Then $u_\alpha = \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}/E_1 E_2$, where p_1 and p_2 are four-momenta of the two particles in the state α and E_1, E_2 are their energies.) In the center-of-mass frame we define the scattering amplitude for the scattering of a two-particle state α into another two-particle state β as

$$f(\alpha \to \beta) = -\frac{4\pi^2}{E} \sqrt{\frac{k'}{k} E_1 E_2 E_1' E_2'}\, M_{\beta\alpha}, \tag{2.169}$$

where E is the total energy, E_1 and E_2 are the energies of the particles in the initial state, E_1', E_2' energies of the particles in the final state, and k, k' the magnitudes of the (center-of-mass) three-momenta of the initial and final particles, respectively. For the forward scattering this reads $f(\alpha \to \alpha) = -\frac{4\pi^2 E_1 E_2}{E} M_{\alpha\alpha}$. Furthermore, for $N_\alpha = 2$ the optical theorem (2.166) reads $\Im\{M_{\alpha\alpha}\} = -u_\alpha \sigma_\alpha^{(\mathrm{total})}/16\pi^3$. Combining this with the above expression for $f(\alpha \to \alpha)$ (and using the fact that in the center-of-mass frame $u_\alpha = kE/E_1 E_2$) then yields the familiar form of the optical theorem, which is often derived in non-relativistic quantum mechanics,

$$\Im\{f(\alpha \to \alpha)\} = \frac{k}{4\pi} \sigma_\alpha^{(\mathrm{total})}. \tag{2.170}$$

such $M_{\alpha\alpha}$ can depend is the total energy E. Now, notice that the integral

$$\int d\gamma\ \delta^4\left(p_\gamma - p_\alpha\right)\left|M_{\gamma\alpha}\right|^2, \tag{2.171}$$

which appears on the right-hand side of (2.166), runs over all intermediate states (labeled by γ) that have the same total four-momentum as the state $\Psi_\alpha^{(\mathrm{in})}$. That is, these are states of zero total three-momentum and the total energy E. The range of possible values of E is determined by the energy spectrum of the physical states $\Psi_\alpha^{(\mathrm{in})}$, which starts at $E = 2m$ and continues to infinity. If we start at $E = 2m$ and gradually increase the energy we expect the integral (2.171) to change, both because $M_{\gamma\alpha}$ can in general depend on the energy and also as a consequence of the gradual change in the size of the kinematic space of integration. The optical theorem (2.166) then implies that the imaginary part of $M_{\alpha\alpha}$ also changes with energy. Note, however, that when we reach the value $E = 3m$, the character of the integral (2.171) suddenly changes. While for energies $2m < E < 3m$ only two-particle intermediate states γ were permissible, when the energy reaches the threshold value of $3m$ a new channel for intermediate states opens. For $E > 3m$, the integral (2.171) runs not only over two-particle intermediate states, but also over three-particle intermediate states. Because of this abrupt change in the character of the right-hand side of (2.166) at $E = 3m$ we would expect some kind of irregularity[44] of the imaginary part of $M_{\alpha\alpha}$ to appear. Therefore, we would expect $M_{\alpha\alpha}$ to not be analytic at $E = 3m$, and under normal circumstance this turns out to be correct. This kind of a threshold is called a *normal threshold*, and it is usually a branch point. As the energy increases, new and new channels open, and at each such a point a normal threshold singularity is expected to appear.

This argument may be intuitive but it is also unsatisfactory in several regards. Among those, the most glaring are likely the following:

1. Although the opening of a new channel on the right-hand side of (2.166) strongly suggests that the matrix element $M_{\alpha\alpha}$ is not analytic there, nothing that we have said so far proves that there must indeed be a singularity.

2. In our argument we considered only the case of the forward-scattering matrix element $M_{\alpha\alpha}$. That is too restrictive, and we would like to

[44] Note that the opening of a new channel does not imply that $\Im\{M_{\alpha\alpha}\}$ must be discontinuous there. In fact, under usual circumstances if there is a reaction that is possible only above some threshold E_0 we would expect the total cross section for that reaction to depend on the energy approximately as $\sqrt{E - E_0}$ when E is slightly above the threshold. (This is under assumption that the final state is allowed to have zero orbital angular momentum. Otherwise, we would instead expect $\sigma \propto (E - E_0)^{\frac{2l+1}{2}}$, where l is the lowest value of the orbital angular momentum present in the final state.) If the contribution of the new channel above the threshold had this character (with $l = 0$) then $\Im\{M_{\alpha\alpha}\}$ would be continuous but not differentiable at the threshold.

show that normal thresholds are present in general matrix elements $M_{\beta\alpha}$.

3. Unless α corresponds to a single stable particle, in which case the matrix $M_{\beta\alpha}$ is identically zero, the energy of a scattering state α is always at or above the lowest threshold. (That is, in our case $E_\alpha \geq 2m$.) But this means that our argument cannot be used to say anything about the nature of $M_{\alpha\alpha}$ at the lowest normal threshold, which in our case was at $E = 2m$.

In the remainder of this section we will provide an incomplete discussion of some of these objections. A more complete treatment can be found in [1] and references cited therein.

Let us start with the second objection, that our argument was of very limited scope. When we decided to restrict our attention to the case $\alpha = \beta$, the unitarity condition (2.165)

$$\frac{i}{2\pi}\left(M_{\beta\alpha} - M^*_{\alpha\beta}\right) = \int d\gamma\ \delta^4\left(p_\gamma - p_\alpha\right) M^*_{\gamma\beta} M_{\gamma\alpha} \quad (\text{for } p_\alpha = p_\beta),$$

simplified at two places. First, the factor $M^*_{\gamma\beta}M_{\gamma\alpha}$ on the right-hand side became $|M_{\gamma\alpha}|^2$. This allowed us to express the right-hand side in terms of the total transition rate or the total cross section. For our purposes, however, this was not very important. Our argument relied only on the fact that at normal thresholds new channels for intermediate states open, and this happens for the product $M^*_{\gamma\beta}M_{\gamma\alpha}$ in pretty much the same way as for $|M_{\gamma\alpha}|^2$. It was the second simplification, that allowed us to replace $M_{\beta\alpha} - M^*_{\alpha\beta}$ on the left-hand side with $2i\Im\{M_{\alpha\alpha}\}$, that was of key importance. This is because for $\alpha \neq \beta$ there is no simple relation between $M_{\beta\alpha}$ and $M_{\alpha\beta}$.[45]

There is a special case when this complication can be addressed by a simple application of the rotational covariance of the S-matrix, and that is the case of elastic scattering of two spinless particles. Suppose that in the center-of-mass frame the initial state α consists of a spinless particle of mass m_1 and three-momentum $\mathbf{p}_1$ and a spinless particle of m_2 and three-momentum $\mathbf{p}_2$. Since we assume that the scattering is elastic, the final state β contains the same particles, and we denote the final three-momentum of the particle of mass m_1 by $\mathbf{p}'_1$, and the final three-momentum of the particle

[45] To exchange multi-indices α and β in the S-matrix element $S_{\beta\alpha}$ one could try to apply the operation T of time inversion which transforms asymptotic "in"-states into asymptotic "out"-states and vice versa. In fact, one would need to consider the product CPT of the operators of charge conjugation, space inversion and time inversion, because a general relativistic quantum field theory is symmetric under CPT, but not under T alone. There is, however, a complication that CPT relates the "in"-state ("out"-state) Ψ_α to the "out"-state ("in"-state) $\Psi_{\alpha'}$ where α' does not correspond to the same particle content as α, but instead to the content of α with all particles exchanged for their antiparticles and their spin z-components and helicities reversed.

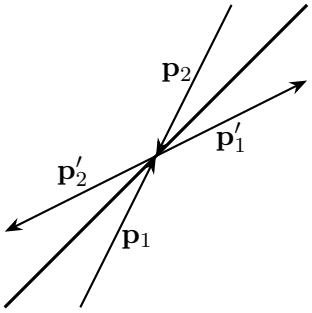


Figure 2.7: A rotation by π about the axis depicted in the figure by the thick line carries the vector $\mathbf{p}_1$ into the vector $\mathbf{p}'_1$ (we are considering only the vectors themselves — their directions and lengths — and not their positions) and $\mathbf{p}'_1$ into $\mathbf{p}_1$. It also exchanges $\mathbf{p}_2$ and $\mathbf{p}'_2$.

of mass m_2 by $\mathbf{p}'_2$. Consider now the axis parallel to the vector $\mathbf{p}_1 + \mathbf{p}'_1$ (or, equivalently, parallel to $\mathbf{p}_2 + \mathbf{p}'_2$). (See Fig. 2.7.) A rotation by π about this axis transforms the momentum $\mathbf{p}_1$ into $\mathbf{p}'_1$ (and vice versa) and $\mathbf{p}_2$ into $\mathbf{p}'_2$ (and vice versa). Hence, such a rotation exchanges the states α and β. But because the particles were assumed to be spinless the rotational covariance of the S-matrix implies that $M_{\beta\alpha} = M_{\alpha\beta}$ and the unitarity conditions (2.164) and (2.165) therefore simplify into

$$-\frac{1}{\pi}\Im\{M_{\beta\alpha}\} = \int d\gamma\ \delta^4(p_\gamma - p_\alpha)\, M_{\beta\gamma}M^*_{\alpha\gamma} = \int d\gamma\ \delta^4(p_\gamma - p_\alpha)\, M^*_{\gamma\beta}M_{\gamma\alpha}. \tag{2.172}$$

We can now use the same argument as before to argue that for an elastic scattering of two spinless particles $M_{\beta\alpha}$ should be singular at normal thresholds.

The result (2.172) is encouraging, but because of its limited applicability it is still not very satisfying. If one assumes the principle of maximal analyticity, however, then it is possible to obtain much more general results from the unitarity condition for the S-matrix combined with the cluster decomposition principle, which dictates the connectedness structure of the S-matrix. In particular, one can provide strong arguments in favor of the so-called principle of *generalized unitarity*, which says that we can continue the unitarity relations (2.164), (2.165) to energies outside of the physical spectrum of the scattering states α, β. To illustrate this, let us return to our earlier example of a theory that contains only one kind of particle, of the mass $m > 0$. Suppose that α describes a two-particle state and β describes a state consisting of three particles. The unitarity relations (2.164) and (2.165), as they were derived, can therefore be applied only if $E \geq 3m$, because that is the minimal possible energy of a physically permissible three-particle state β. (Of course, the energy spectrum of two-particle states α goes as low as $2m$, but for the unitarity condition to be applicable both α and β must

describe physical states.) But if the condition of generalized unitarity holds we can continue the relations (2.164) and (2.165) out of their physical region to energies $E < 3m$. The integrals over the intermediate states in those relations, however, are still restricted to those states γ that are allowed by the energy-momentum conservation condition. This can then be used to prove the property called *Hermitian analyticity*, which says that $M_{\alpha\beta}$ and $M^*_{\beta\alpha}$ are analytic continuations of the same analytic function, which is real below the lowest threshold. To reach $M_{\alpha\beta}$ we must continue above the normal thresholds and to reach $M^*_{\beta\alpha}$ we continue below the thresholds. This then provides an argument that the normal thresholds are branch points that is both more general and more convincing than what we have presented. For details, the reader is advised to consult Chapter 4 of [1].

2.6 Poles of unstable particles

At this point we are approaching the end of the first part of this thesis. We have discussed some of the fundamental reasons why Green's functions and other matrix elements can be expected to have some particular analytic properties. We have mentioned that Laplace transforms, if they exist, are always analytic, and that the existence of Laplace transforms of tempered distributions in certain tubes is equivalent to the presence of particular restrictions on the support of those distributions. We also briefly mentioned two basic physical principles that lead to such restrictions in quantum field theory: the spectral condition in case of the vacuum expectation values of fixed-order products of fields in momentum-space, and the condition of microscopic causality in the case of coordinate-space (time-ordered) Green's functions. Then we moved on to explain why stable particles correspond to poles in Green's functions and to introduce the concept of a normal threshold. The primary value that I see in this discussion is that it may serve as a gentle introduction to this field of study. But there was also another, more specific reason behind the selection of topics in that discussion: I wanted to introduce the main concepts that are employed in the Unitary and Analytic model. As we will see later, these are the concepts of maximal analyticity, and of normal threshold and particle pole singularities. However, the particles that play role in the Unitary and Analytic model are not stable, they are resonances, and for this reason we need to discuss one more topic, that of poles of unstable particles.

We have seen that stable particles give rise to poles in Green's functions and a more careful analysis would show that these poles lie on the physical sheet. But what about unstable particles? On the one hand, unstable particles do not appear among asymptotic states which suggests that they should not lead to poles on the physical sheet. But on the other hand, it is difficult in practice to distinguish between a sufficiently long-lived unstable particle

and a particle that is truly stable, and one would therefore expect there to be some kind of poles also for unstable particles. As it turns out, unstable particles do indeed seem to result in poles, but those poles are not on the physical sheet. Instead, they are hidden behind normal threshold branch cuts. It is possible to give various arguments in favor of that statement, but here we will limit ourselves only to a simple heuristic description of those poles. Afterwards we will present a depiction of how the relation between poles of unstable particles and poles of stable particles can be understood. In that we will follow the treatment of [16]. (For a wider and more detailed discussion see [1] and references cited therein.)

As is well-known from the usual derivation of the Breit-Wigner formula, a term with energy dependence of the form

$$\frac{1}{E - \left(M_r - i\frac{\Gamma_r}{2}\right)} \tag{2.173}$$

in a scattering amplitude has the effect that seems to correspond to the situation when there is present a resonant intermediate state with mass spectrum centered near M_r and with decay rate Γ_r. To see this, consider a superposition $\int dE\ g(E)\Psi_E$ of energy eigenstates Ψ_E where the coefficient function g is sufficiently smooth and centered near E_r. Then it may easily happen that the scattering is dominated by the pole term (2.173), and if that is the case then the time dependence of the scattered state is approximately[46]

$$\int dE\ g(E)\frac{\mathrm{e}^{-iEt}}{E - M_r + i\frac{\Gamma_r}{2}} \to -2\pi i\ g\left(M_r - i\frac{\Gamma_r}{2}\right)\ \mathrm{e}^{-iM_r t}\mathrm{e}^{-\Gamma_r t/2}. \tag{2.174}$$

Now, since the probability can be calculated from the amplitude by taking the square of its modulus, the result (2.174) corresponds to the presence of an intermediate state that decays as $\exp(-\Gamma_r t)$. This heuristic argument therefore suggests that the effect of an unstable particle can be approximated

[46] Let us unpack the derivation of this result. First of all, we assume that near $E = M_r$ the scattering amplitude is entirely dominated by the contribution of the pole at $E = M_r - i\Gamma_r/2$. This allows us to claim that we can approximate its energy dependence as being proportional to $1/\left[E - (M_r - i\Gamma_r/2)\right]$. In order for this assumption to be reasonable, we must require that the wave packet be centered near $E = M_r$, because otherwise we would need to explicitly consider also contributions from distant energies, where the resonance pole can no longer be expected to dominate. In particular, this implies that the wave packet must be negligibly small near other singularities, such as normal thresholds or other resonances. Second, we assume that g is smooth enough because we want to calculate the integral by using the residue theorem. We are interested in large positive times, which means that we can close the contour of integration by a large semi-circle in the lower half-plane. And lastly, the limit in (2.174) corresponds to $t \to +\infty$. This is because generally there will also be contributions from poles of g, which might be present in the lower half-plane. However, if the resonance is sufficiently long-lived, then the poles of g will all lie much further down the imaginary axis than the resonance pole, and will therefore not contribute in that limit.

by a pole at $M_r - i\Gamma_r/2$. But is this pole on the physical sheet? The answer is no. First of all, if the particle is unstable, then M_r must necessarily lie above the lowest threshold, for otherwise there would be no scattering state into which it could decay. Although we did not really explain it properly here, the physical region of a normal threshold branch cut can be reached by approaching the branch cut on the physical sheet from above. That is, if the lowest threshold is at E_0 then for $E > E_0$ the physical region corresponds to points $E = \Re\{E\} + i\epsilon$ (with $\epsilon > 0$) on the physical sheet in the limit $\epsilon \to 0^+$. But the resonance pole (2.173) is accessed by starting at $E = M_r + i\epsilon$ in the physical region — right above the branch cut — and then going downwards parallel to the imaginary axis the distance $\Gamma_r/2$. Hence, to reach the pole we start on the physical sheet and cross the branch cut from above. The pole is therefore located on another sheet, hidden behind the branch cut.

Note that since the amplitude is real below the lowest threshold, the Schwarz reflection principle implies that the resonance pole at $E = M_r - i\Gamma_r/2$, which is on the sheet that can be accessed from the physical sheet by crossing the branch cut from above, must be accompanied by another, complex conjugate "shadow" pole at $E = M_r + i\Gamma_r/2$, and this pole is on the sheet that is accessed from the physical sheet by crossing the branch cut from below. The shadow pole is therefore quite distant from the physical region, and its contribution is neglected in the Breit-Wigner formula.[47]

In the remaining part of this section we will suggest a way of how to understand the relation between poles of stable particles and those of unstable particles. We will follow the treatment of [16, 1]. For simplicity, we will return to our earlier example of a theory that contains only a single kind of particle, of spin zero and mass $m > 0$. Let us suppose that there is some real parameter g, corresponding to a coupling constant, which can be continuously adjusted, and that as we vary the coupling constant the solution of the theory — and, in particular, all the amplitudes — changes analytically with g. Furthermore, we suppose that at some initial value $g = g_0$ there are no true bound states present in the theory but that there is an unstable bound state that decays into two particles of mass m, and we also suppose that there is some higher value $g_1 > g_0$ of g, at which the coupling is strong enough to make the unstable bound state stable. As we gradually increase g_0 to g_1 we will encounter a point when the unstable state becomes stable. Let us denote this threshold value of g by g_{th}. Clearly, $g_0 < g_{\text{th}} < g_1$. As we will see, when $g = g_{\text{th}}$ the particle pole meets the normal threshold, and that renders the description less transparent. For that reason, we will also assume that the solution can be continued into some complex neighborhood of the real axis of g, so that we can avoid the threshold value $g = g_{\text{th}}$ by

[47] This of course leads to the question of whether the usual Breit-Wigner formula needs to be adjusted for near-threshold resonances. We will encounter this question again in the section 6.2, but we will not really address it in the current work.

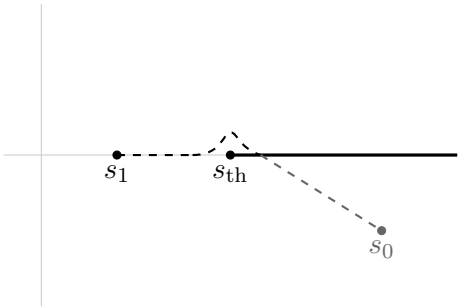


Figure 2.8: *An illustration of how a stable particle pole can change into a resonance pole.* We suppose that for $g = g_1$ there is a (stable) bound state, whose pole is located at $s = s_1$ on the physical sheet. This pole lies on the real axis below the lowest threshold, which is at $s_{\rm th}$. As we decrease the strength of the coupling g the energy of the bound state increases and the corresponding pole moves towards the threshold. For $g = g_{\rm th}$ the particle pole and the lowest threshold coincide, and to avoid that situation we give g a small imaginary component when it is near $g_{\rm th}$. This moves the pole off the real axis. For $g < g_{\rm th}$ the bound state turns into a resonance, and as we make the coupling constant real again the pole crosses the branch cut from above, and appears on another sheet. It then continues to move on that sheet until it reaches the position $s = s_0$ for $g = g_0$. In the figure, the part of the pole's path that is on the physical sheet is depicted by black color, and the path on the unphysical sheet is depicted by gray.

going slightly above or slightly below it. Note that when g is complex the theory loses some of its physical properties — in particular, the condition of Hermitian analyticity does no longer apply, which means that a resonance pole and its "shadow" pole might not occupy complex conjugated positions.

To be specific, let us consider the amplitude for elastic scattering of two particles (of mass m). This is the simplest amplitude that should contain the pole for the bound state of those two particles. We will consider this amplitude as a function of the variable s, the center-of-mass energy squared.[48] The path that the pole takes when we vary the coupling g between the values g_0 and g_1 is depicted in Fig. 2.8. If we start at the value $g = g_1$, the bound state is stable, and there should be a corresponding pole on the physical sheet, lying on the real axis below the lowest threshold. In the figure, we labeled its position with s_1. Now, as the coupling gradually decreases, the binding energy of the bound state gets smaller, and the pole approaches the lowest threshold at $s = s_{\rm th}$. When $g = g_{\rm th}$ the pole is right at the lowest threshold because it is at this value of s that the stable state turns into an unstable one. However, if the pole is right at the threshold, then as we

[48]This is a more usual choice of the variable for such a case, rather than the total center-of-mass energy E that we used in the previous section.

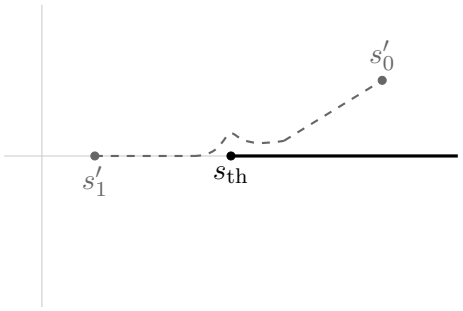


Figure 2.9: *The path of the "shadow" pole as the coupling constant changes between* g_0 *and* g_1. In Fig. 2.8 we depicted how a pole corresponding to a bound state moves on the physical sheet and eventually crosses the branch cut when the coupling gets weak enough for the state to become unstable. Here we depict the corresponding path of its "shadow" pole. (But for the ease of presentation we describe it for g starting at g_0 and being gradually increased.) The pole starts at $s = s'_0$. As the coupling increases from g_0 to g_1, and the resonance pole crosses the branch cut and moves onto the physical sheet, the "shadow" pole must remain on the "unphysical" sheet. Hence, near $g = g_{\rm th}$ when g has a small imaginary part, the "shadow" pole must move as indicated in the figure, because it must not cross the branch cut. When the coupling constant $g > g_{\rm th}$ becomes real again, the "shadow" pole moves onto the real axis but stays on the "unphysical" sheet.

further vary g the pole can move either to the original sheet or cross the branch cut — both of those options are possible. To avoid this complication we give g a small imaginary component when it is close to $g_{\rm th}$. This way it will avoid the threshold, making it easy to determine if the pole crosses the branch cut or not. In the figure this is illustrated as a small detour into the upper-half plane, which the pole takes around the point $s = s_{\rm th}$. (Note that we do not know that the pole must go into the upper half-plane. The precise way in which the pole avoids the threshold depends on what exact theory do we work with, as well as the sign of the imaginary component of g.) Afterwards, when the real part of g is smaller than $g_{\rm th}$, we can make the coupling real again and the pole crosses the branch cut, as indicated in the figure, and moves to another sheet. And finally, when g decreases to g_0 the pole assumes its position at $s = s_0$ on this other, "unphysical" sheet. Note, however, that since the overall amplitude must be real below the lowest threshold the presence of the pole at the complex position s_0 in Fig. 2.8 necessitates a presence of another pole at a complex conjugated position. This "shadow" pole is on the sheet that can be accessed from the physical sheet by crossing the branch cut from below. Note that since in our example the threshold is assumed to be the lowest threshold, it must be of the square-root type, which means that if one crosses it once, then goes around the branch point and crosses the cut the second time in the same direction, then one returns to the original sheet — just as if the function was $\sqrt{s}$. Therefore, in our case the resonance pole and its "shadow" pole both lie on the same "unphysical" sheet. Under more general circumstances, though, that might no longer be true.

It is important to recognize that when the coupling is strong enough so that the resonance turns into a stable state and its pole is on the real axis on the physical sheet, the "shadow" pole must be located on another sheet. This is because at any given time there should be only a single pole corresponding to the bound state on the physical sheet. For that reason, when g is varied in a way that corresponds to the path depicted in Fig. 2.8, the "shadow" pole must move on the path illustrated in Fig. 2.9. In particular, when g is near $g_{\rm th}$ and is complex, the shadow pole takes the path as depicted in figure: it remains in the upper half-plane and ceases to occupy the position that is complex conjugate to the resonance pole, for it must not cross the branch cut. For real $g > g_{\rm th}$ both the resonance pole and the "shadow" pole are on the real axis as required by the Hermitian analyticity, but each on a different sheet.

Before we close, let us mention that decreasing the coupling is not guaranteed to turn a bound state into a resonance. It might instead become a virtual state. In such a case, as we decrease g from g_1 to g_0 the pole at first approaches the branch point from the physical sheet, just as in Fig. 2.8, but after it crosses the branch cut and moves onto the other sheet, and after the coupling becomes real again (and $g < g_{\rm th}$) it does not remain complex

but instead returns onto the real axis below the threshold, remaining on the "unphysical" sheet. In this case there is no need for a "shadow" pole.

Part II

Electromagnetic form factors and the Unitary and Analytic model

Chapter 3

Electromagnetic form factors

Electromagnetic form factors appear when one considers the scattering of a charged particle by a virtual photon. As we will see, the corresponding matrix elements are fully determined by their properties under Lorentz transformations and by the conservation condition for the electromagnetic current except for a number of unknown coefficient functions that depend on kinematic invariants of the scattering. These functions are called electromagnetic form factors, and are of great importance when one needs to make a contact between theory and experiment. For the purposes of this thesis electromagnetic form factors are relevant in the following three ways. First, the main aim of this second part of the thesis is to introduce the Unitary and Analytic model, which is a model of electromagnetic form factors of hadrons. Second, one of the more practical topics that this thesis touches upon is the evaluation of the muon's magnetic anomaly, and this quantity is directly related to electromagnetic form factors of the muon. And thirdly, electromagnetic form factors of hadrons, and their representation by the Unitary and Analytic model, will play an important role in several research topics discussed in the third part of this thesis. A reader interested in the general theory of form factors can find additional information for example in the chapter 10 of [9] or the chapter 26 of [17], which were also employed as main references for the treatment presented here.

3.1 Initial considerations

Suppose that we wish to describe the scattering of an electrically charged particle by a single virtual photon, and that we want to include all self-interactions of that particle. Another way of saying this is that we are describing the interaction of that charged particle with an external electromagnetic field (or with some other charged particle), but only to the first order in that external field (or in the charge of that other particle), and to all orders — or, in fact, non-perturbatively — in all other interactions. In

terms of Feynman diagrams this corresponds to the sum of all the diagrams with one external line corresponding to the charged particle entering the diagram, another external line corresponding to the charged particle exiting the diagram, and a single external vertex corresponding to the electromagnetic current. We can consider the matrix element of the operator of the current, instead of the matrix element of the gauge field, because the external photon gauge field is assumed to be coupled to the current. (One of the advantages of considering the current instead of the gauge field is that matrix elements of the gauge field transform under Lorentz transformations in a more complicated way than those of the current.) This sum of Feynman diagrams corresponds to the matrix element

$$\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, J^{\mu}(x)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right), \tag{3.1}$$

where $\Psi^{\text{in}}_{\mathbf{p},\sigma}$ and $\Psi^{\text{out}}_{\mathbf{p}',\sigma'}$ are the "in" and "out" states of the charged particle and $J^{\mu}(x)$ is the electromagnetic current (density) expressed in the Heisenberg picture.

Before we continue, let me add one remark. People sometimes say that the non-triviality of electromagnetic form factors of some particles, such as the proton, corresponds to the fact that those particles are not point particles. If the form factors are understood in the way as we introduced them here, that statement is simply wrong. In quantum field theory, the distinction between elementary and composite particles is far from obvious, but even if we sidestep this issue by adopting the definition that elementary particles are those whose fields appear in the Lagrangian, electromagnetic form factors of charged particles are never trivial if we include effects of their electromagnetic self-interaction. However, there is a good sense in which such statements are meaningful, and that is when we consider the matrix elements such as (3.1) not in the full theory, but in a theory with some of the interactions switched off. In particular, it is very useful to consider the matrix element (3.1) evaluated with only the strong force switched on. This corresponds to the calculation of the scattering in the external electromagnetic field to all orders in the strong interaction but only to the lowest order in the electromagnetic charge e. In that approach the electromagnetic form factor of, say, the electron does indeed turn out to be trivial, since the electron does not interact strongly and we decided to ignore its electromagnetic (and weak) self-interactions. In this sense the statements such as that the electron is a point-like particle while the proton is not simply mean that to our required precision — the lowest order in e — the electron's self-interactions can be neglected but those of the proton cannot, and this is a direct consequence of the fact that at the relevant energy scale the strong force is much stronger than the electromagnetism.[1] In fact, even in this thesis, when we will later speak about the electromagnetic form factors of

[1] Of course, there is also (at least one) other sense in which statements such as that

hadrons and their representations by the Unitary and Analytic model, what we will have in our mind are exactly these "strong-interactions-only" electromagnetic form factors: form factors that correspond to (3.1) evaluated with only the strong force switched on.

Let us now start exploring the amplitude (3.1). First, transformation properties (2.104) under translations allow us to explicitly factor out its x-dependence,

$$\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, J^{\mu}(x)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right) = \mathrm{e}^{ix(p'-p)}\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, J^{\mu}(0)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right). \tag{3.2}$$

If we now apply to both sides the derivative $\partial/\partial x^{\mu}$, and use the current-conservation condition $\partial_{\mu}J^{\mu} = 0$, we obtain the important restriction

$$(p'-p)_{\mu}\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, J^{\mu}(0)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right) = 0. \tag{3.3}$$

Another restriction on (3.1) that can be obtained using properties of the conserved current J^{μ} is a kind of normalization condition. Before we derive this condition let us briefly discuss the required properties of J^{μ}.

From any conserved current, electromagnetic or other, we can construct the corresponding charge operator

$$Q = \int d^3\mathbf{x}\; J^0(\mathbf{x}, t). \tag{3.4}$$

The current-conservation condition $\partial_{\mu}J^{\mu}$ implies that Q is conserved,

$$\frac{d}{dt}Q = \int d^3\mathbf{x}\; \partial_0 J^0(\mathbf{x}, t) = -\int d^3\mathbf{x}\; \nabla\cdot\mathbf{J}(\mathbf{x}, t) = 0, \tag{3.5}$$

because we can use the Gauss's theorem to express the last integral in terms of the integral of the flux of $\mathbf{J}$ over the boundary of the volume of integration, which is zero. (As usual, we assume here that the fields vanish as $|\mathbf{x}| \to \infty$.) The fact that Q is an integral of J^0 over the whole space means that Q is invariant under space translations. This, together with $dQ/dt = 0$, means that Q commutes with generators of space-time translations P^{μ},

$$[Q, P^{\mu}] = 0. \tag{3.6}$$

By the same argument we can see that Q also commutes with the generators of rotations. It is, however, a slightly more difficult to show that Q also commutes with boost generators. The problem is that the definition (3.4) is not manifestly covariant and appears to depend on the definition of the time coordinate. Boosts modify both the hyperplanes of constant

proton is not a point-like particle can be interpreted: hadrons are bound states in an asymptotically free theory and it is therefore quite meaningful to speak about its internal structure when considering high-energy processes such as, for instance, deep inelastic scattering.

time and the time component J^0. In order to explore this issue let us first rewrite the definition (3.4) in a form in which its covariance will become more apparent. Consider the Lorentz-covariant three-form $\epsilon_{\mu\nu\rho\sigma}J^\sigma$. When we integrate it over a hyperplane of constant time, only the time component of J^σ contributes. Hence, one can see that the definition (3.4) is equivalent to[2]

$$Q = \int \epsilon_{\mu\nu\rho\sigma}J^\sigma, \tag{3.7}$$

where the integrand is interpreted as a differential form and the integration surface must be a hyperplane of constant time in some given coordinate frame. Let us now fix a reference frame, denote the time coordinate in this reference frame by t, and apply (3.7) to obtain the value Q of the charge by integrating over the hypersurface $t = 0$. Let us next boost to a different frame of reference and denote the new time coordinate by t'. If we now apply (3.7) in this new reference frame we obtain a new value Q' of the charge operator, and it is not obvious that $Q = Q'$, because when inspected in the old reference frame the quantity Q' is obtained by integrating the three-form $\epsilon_{\mu\nu\rho\sigma}J^\sigma$ over the hypersurface $t' = \text{const}$, which is *not* a hypersurface of constant time in the old reference frame. However, by the generalized Stokes' theorem the difference between Q and Q' equals the integral of the differential of $\epsilon_{\mu\nu\rho\sigma}J^\sigma$ over the oriented volume between the hyperplanes $t = 0$ and $t' = 0$.[3] But the three-form $\epsilon_{\mu\nu\rho\sigma}J^\sigma$ is closed,

$$d_\alpha\left(\epsilon_{\mu\nu\rho\sigma}J^\sigma\right) \propto \epsilon_{\alpha\mu\nu\rho}\partial_\mu J^\mu = 0. \tag{3.8}$$

Hence $Q = Q'$, and we have demonstrated that the operator Q commutes with boosts. (In fact, we have actually demonstrated that Q can be calculated by the integral (3.7) over any three-surface that stretches to the infinity in all space directions.) Thus, we see now that Q commutes with the generators of translations and also all the generators of Lorentz transformations. We can use these commutation properties to constrain the possible values of $Q\Psi_{\mathbf{p},\sigma,n}$. First, since $[Q, P^\mu] = 0$ we know that $Q\Psi_{\mathbf{p},\sigma,n}$ has the same value of four-momentum as $\Psi_{\mathbf{p},\sigma,n}$. Barring any degeneracy in particle masses this also means that the species of $Q\Psi_{\mathbf{p},\sigma,n}$ is still n.[4] Furthermore, since Q also commutes with the generators of Lorentz transformations, it preserves the spin or helicity value σ. In other words, $Q\Psi_{\mathbf{p},\sigma,n}$ must be proportional to $\Psi_{\mathbf{p},\sigma,n}$,

$$Q\Psi_{\mathbf{p},\sigma,n} = q_n\Psi_{\mathbf{p},\sigma,n}. \tag{3.9}$$

[2]Perhaps up to the overall sign, depending on the exact definition of $\epsilon_{\mu\nu\rho\sigma}$. Here, we use $\epsilon_{0123} = -1$, but the overall sign is not important in our current argument.

[3]Here we again use the assumption that J^μ vanishes as $|\mathbf{x}|$ goes to infinity, so that we do not need to consider the integral of $\epsilon_{\mu\nu\rho\sigma}J^\sigma$ over the rest of the boundary of the volume between $t = 0$ and the hyperplane of the constant t'.

[4]In fact, to exclude the possibility that Q transforms the particle into its antiparticle would require a further discussion. Let us just assume here that this does not happen.

The proportionality coefficient q_n cannot depend on $\mathbf{p}$ or σ; again because of the commutation properties of Q with the generators of Lorentz transformations. We call the coefficient q_n the charge of the particle n. (In our case q_n is its electric charge.)

Now we can return to (3.1) and derive the normalization condition. Integrating the zeroth component of (3.2) over the whole space $t = 0$ yields

$$\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, Q\Psi^{\text{in}}_{\mathbf{p},\sigma}\right) = (2\pi)^3\delta^3\left(\mathbf{p}' - \mathbf{p}\right)\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, J^0(0)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right). \tag{3.10}$$

We can now use (3.9) and the normalization condition $(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, \Psi^{\text{in}}_{\mathbf{p},\sigma}) = \delta^3\left(\mathbf{p}' - \mathbf{p}\right)\delta_{\sigma'\sigma}$ to obtain

$$\frac{1}{(2\pi)^3}\, q\, \delta_{\sigma'\sigma} = \left(\Psi^{\text{out}}_{\mathbf{p},\sigma'}, J^0(0)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right), \tag{3.11}$$

where we denoted the electric charge of the particle by q. Notice that we had to change $\mathbf{p}'$ to $\mathbf{p}$ because we were comparing here the coefficients of two three-momentum delta functions. The normalization condition (3.11) together with the current-conservation condition (3.3) are the two main conditions resulting from the properties of the current operator J^μ that we are going to use in our investigation of the matrix elements (3.1). Furthermore, we are also going to employ the fact that J^μ is a Hermitian operator. The only remaining restriction that we are going to use is the requirement that (3.1) must be Lorentz-covariant.

3.2 Spin zero particle

Electromagnetic form factors can be defined for a particle of any spin, but we will be mainly interested in the cases of spin 0 and spin $\frac{1}{2}$, which we are now going to treat in a greater detail. We start with the case of spin zero. That is, we investigate the amplitude

$$\left(\Psi^{\text{out}}_{\mathbf{p}'}, J^\mu(0)\Psi^{\text{in}}_{\mathbf{p}}\right), \tag{3.12}$$

where both the "in" and the "out" states transform as

$$U(\Lambda)\Psi_{\mathbf{p}} = \sqrt{\frac{(\Lambda p)^0}{p^0}}\Psi_{\mathbf{p}_\Lambda}, \tag{3.13}$$

where we have denoted the three-vector part of Λp by $\mathbf{p}_\Lambda$. Let us now write (3.12) as

$$\left(\Psi^{\text{out}}_{\mathbf{p}'}, J^\mu(0)\Psi^{\text{in}}_{\mathbf{p}}\right) = \frac{q}{(2\pi)^3}A^\mu\left(p', p\right), \tag{3.14}$$

where we have extracted the factor $q/(2\pi)^3$ according to the normalization condition (3.11). Using transformation properties of the vector operator J^μ

and of the spin zero particle, we obtain the following restriction on A

$$\begin{aligned}\frac{q}{(2\pi)^3}\Lambda^\mu{}_\nu A^\nu(p',p) &= \left(\Psi^{\rm out}_{\mathbf{p}'}, U(\Lambda)^{-1}J^\mu(0)U(\Lambda)\Psi^{\rm in}_{\mathbf{p}}\right)\\ &= \left(\sqrt{\frac{(\Lambda p')^0}{p'^0}}\Psi^{\rm out}_{\mathbf{p}'_\Lambda}, J^\mu(0)\sqrt{\frac{(\Lambda p)^0}{p^0}}\Psi^{\rm in}_{\mathbf{p}_\Lambda}\right) =\\ &= \frac{q}{(2\pi)^3}\sqrt{\frac{(\Lambda p')^0(\Lambda p)^0}{p'^0 p^0}}A^\mu(\Lambda p',\Lambda p), \quad (3.15)\end{aligned}$$

or, more simply,

$$\sqrt{p'^0 p^0}\;\Lambda A(p',p) = \sqrt{(\Lambda p')^0(\Lambda p)^0}\;A(\Lambda p',\Lambda p). \quad (3.16)$$

We can solve this by choosing A^μ of the form

$$A^\mu(p',p) = \frac{1}{\sqrt{2p'^0\;2p^0}}\mathcal{J}^\mu(p',p), \quad (3.17)$$

with $\mathcal{J}$ a vector function of p' and p. (We have also extracted an additional factor $1/2$ for later convenience.) Now, the only way to construct a four-vector $\mathcal{J}^\mu(p',p)$ from p and p' is to add them in a linear combination with coefficients that are scalar functions of p and p'. (Only objects that are at our disposal are the two four-vectors p^μ and p'^ν, the metric tensor $\eta_{\mu\nu}$, and the anti-symmetric pseudo-tensor $\epsilon_{\mu\nu\rho\sigma}$.) Of course, instead of considering linear combinations of p and p' directly, we can equivalently also consider linear combinations of any two independent four-vectors that can be constructed from p and p'. It turns out that the most convenient approach is to consider linear combinations of $p'+p$ and $p'-p$. This is because that choice makes it easy to apply the current-conservation condition (3.3). As we said, the coefficients in that linear combination must be scalar functions of p and p', and we can construct only three independent scalars from those two vectors, for example, the scalars p^2, $p'_\mu p^\mu$, and p'^2. However, the four-momenta p' and p are on the mass shell, so we have $p^2 = p'^2 = m^2$. This means that there is only single non-trivial scalar left, which we can choose as $p'_\mu p^\mu$, or even more conveniently, $k^2 = (p'-p)^2$, the square of the transferred momentum. Thus, the most general solution of (3.15) can be written in the form

$$\left(\Psi^{\rm out}_{\mathbf{p}'}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p}}\right) = \frac{q}{(2\pi)^3}\frac{1}{\sqrt{2p'^0\;2p^0}}\left(F(k^2)(p'+p)^\mu + i(p'-p)^\mu G(k^2)\right). \quad (3.18)$$

This is a general Lorentz-covariant form of the left-hand side for a spin zero particle. Notice that we have extracted the factor i from the function G. This is to make the functions F and G both real, which follows from the fact that J^μ is Hermitian. To see this, consider the identity

$$\left(\Psi^{\rm out}_{\mathbf{p}'}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p}}\right)^* = \left(\Psi^{\rm out}_{\mathbf{p}}, (J^\dagger)^\mu(0)\Psi^{\rm in}_{\mathbf{p}'}\right), \quad (3.19)$$

where we have used the fact that for single-particle states $\Psi_{\mathbf{p}}^{\text{in}} = \Psi_{\mathbf{p}}^{\text{out}}$. The reality of F and G then follows from $J = J^\dagger$ and the equation (3.18).

Our next step is to employ the constraints (3.3) and (3.11). Applying the current-conservation condition (3.3) to (3.18) yields

$$0 = F(k^2)(p' + p)^\mu (p' - p)_\mu + ik^2 G(k^2). \tag{3.20}$$

Because p and p' are both on the mass shell, we have $(p' + p)^\mu (p' - p)_\mu = 0$, and the condition (3.20) reduces to $k^2 G(k^2) = 0$. This must hold for all allowed values of k^2. Both p and p' are future-oriented and on the (same) mass shell, which means that $p' - p$ must be space-like, as one can easily verify by inspecting this quantity in the center-of-mass frame. Possible values of k^2 therefore form the interval $-\infty < k^2 \leq 0$. Normally, we would interpret the condition $k^2 G(k^2) = 0$ as implying that G is zero everywhere (on the allowed interval), but since k^2 can be zero we should also consider the option when $G(k^2)$ is proportional to $\delta(k^2)$. But for k space-like, $k^2 = 0$ only if $k^\mu = 0$, and because in (3.18) $G(k^2)$ appears in a product with k^μ, the term containing $G(k^2)$ would still not contribute even if $G(k^2)$ was proportional to $\delta(k^2)$. Hence, the current-conservation condition implies that we may drop the term containing G, so that

$$\left(\Psi_{\mathbf{p}'}^{\text{out}}, J^\mu(0)\Psi_{\mathbf{p}}^{\text{in}}\right) = \frac{q}{(2\pi)^3} \frac{1}{\sqrt{2p'^0\, 2p^0}} F(k^2)(p' + p)^\mu. \tag{3.21}$$

The function F is called the *electromagnetic form factor* of the particle. The normalization condition (3.11) now reads

$$\frac{1}{(2\pi)^3}\, q = \left(\Psi_{\mathbf{p}}^{\text{out}}, J^0(0)\Psi_{\mathbf{p}}^{\text{in}}\right) = \frac{q}{(2\pi)^3} \frac{1}{2p^0} F(0) 2p^0,$$

which immediately implies that

$$F(0) = 1. \tag{3.22}$$

Before we close our discussion of the spin 0 case, let us note that the above results were derived under the normalization condition (2.109), which we have followed throughout the first part of this thesis. If the states $\Psi_{\mathbf{p}}$ and $\Psi_{\mathbf{p}'}$ describe the same spin 0 particle, this condition reads

$$(\Psi_{\mathbf{p}'}, \Psi_{\mathbf{p}}) = \delta^3 (\mathbf{p}' - \mathbf{p}). \tag{3.23}$$

There are also other widespread normalization conventions. Perhaps the most common among them is what is sometimes called the covariant normalization, and which (for a spin 0 particle) reads

$$(\Psi_{\mathbf{p}'}, \Psi_{\mathbf{p}}) = 2p^0 (2\pi)^3 \delta^3 (\mathbf{p}' - \mathbf{p}). \tag{3.24}$$

With this normalization, single-particle states transform simply as

$$U(\Lambda)\Psi_{\mathbf{p}} = \Psi_{\mathbf{p}_\Lambda}, \tag{3.25}$$

instead of (3.13). This simplifies properties of the matrix element (3.12) under Lorentz transformations, and instead of (3.15) we simply get $\Lambda A(p', p) = A(\Lambda p', \Lambda p)$. We can therefore put $A^\mu = \mathcal{J}^\mu$, with $\mathcal{J}^\mu$ defined just as before. Thus, after applying the current-conservation condition, instead of (3.21) we obtain

$$\left(\Psi^{\text{out}}_{\mathbf{p}'}, J^\mu(0)\Psi^{\text{in}}_{\mathbf{p}}\right) = \frac{q}{(2\pi)^3} F(k^2)(p' + p)^\mu. \tag{3.26}$$

However, because now we are using a different normalization of state vectors, the charge normalization condition (3.11) has also changed, and now reads

$$\left(\Psi^{\text{out}}_{\mathbf{p},\sigma'}, J^0(0)\Psi^{\text{in}}_{\mathbf{p},\sigma}\right) = 2p^0\, q\, \delta_{\sigma'\sigma}. \tag{3.27}$$

For F defined as in (3.26) this implies $F(0) = (2\pi)^3$. Hence, if we redefine $F_{\text{new}} = (2\pi)^{-3} F_{\text{old}}$, the equation (3.26) becomes

$$\left(\Psi^{\text{out}}_{\mathbf{p}'}, J^\mu(0)\Psi^{\text{in}}_{\mathbf{p}}\right) = q\, F(k^2)\, (p' + p)^\mu, \tag{3.28}$$

where $F(0) = 1$. This is the usual relation for the electromagnetic form factor of a spin 0 particle, when expressed in the covariant normalization. In fact, in Part III of this thesis we are going to work in that normalization. In the remainder of this chapter, however, we continue using the normalization (2.109).

3.3 Spin zero: Related diagrams

In the previous subsection we have arrived at the equation (3.21), which expresses the matrix element (3.12) in terms of a single unknown real function $F(k^2)$. This function is the electromagnetic form factor of that particle, and the equation (3.21) defines it only for zero or negative values of k^2. This is because the four-momenta p' and p of (3.12) must be on their mass shell and, furthermore, $p'^0 > 0$ and $p^0 > 0$, which means that the transferred four-momentum $k = p' - p$ is either zero or space-like.[5] However, it is very useful to consider the function F also for other values of k^2. To motivate this extension let us first consider two other amplitudes closely related to (3.12).

[5] As we mentioned in the previous subsection, this is most obvious in the center-of-mass frame. Another simple way to show this is to transform into the rest frame of the incoming particle, in which $p = (m, \mathbf{0})$ and $p' = (\sqrt{\mathbf{p}'^2 + m^2}, \mathbf{p}')$. Hence,

$$k^2 = \left(\sqrt{\mathbf{p}'^2 + m^2} - m\right)^2 - \mathbf{p}'^2 = -2m\left(\sqrt{\mathbf{p}'^2 + m^2} - m\right) \leq 0.$$

The first of them is

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^\mu(x)\Psi_0\right). \tag{3.29}$$

Here we have explicitly depicted the particle species labels to show that the particle with the four-momentum $\bar{p}$ is the antiparticle of the particle with the four-momentum p'. In (3.12) both particles were of the type n. We have

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^\mu(x)\Psi_0\right) = {\rm e}^{i(p'+\bar{p})x}\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^\mu(0)\Psi_0\right), \tag{3.30}$$

which means that the condition $\partial_\mu J^\mu = 0$ implies

$$(p'+\bar{p})_\mu\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^\mu(0)\Psi_0\right) = 0. \tag{3.31}$$

We can also try to apply the approach that yielded the normalization condition (3.11) for the original matrix element. Integrating the zeroth component of (3.30) over the hyperplane $t = 0$ leads to

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, Q\Psi_0\right) = (2\pi)^3\delta^3\left(\mathbf{p}'+\bar{\mathbf{p}}\right)\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^0(0)\Psi_0\right). \tag{3.32}$$

Since $Q\Psi_0 = 0$,[6] we obtain the requirement

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;-\mathbf{p}',n^{\rm c}},\, J^0(0)\Psi_0\right) = 0. \tag{3.33}$$

However, as we will see below, this condition turns out to be of little use.

Now, to make use of the Lorentz-covariance properties of (3.29) we can proceed in the same way as we did to obtain (3.18). The result is that the most general Lorentz-covariant form of the amplitude (3.29) can be expressed in the form

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}},\, J^\mu(0)\Psi_0\right) = \frac{q}{(2\pi)^3}\frac{1}{\sqrt{2p'^0\;2\bar{p}^0}}\left(X(k^2)\,(p'+\bar{p})^\mu + Y(k^2)\,(p'-\bar{p})^\mu\right). \tag{3.34}$$

Here, X and Y are some scalar functions of the on-shell four-momenta p' and $\bar{p}$. Unlike in the case of (3.18) we have not extracted the imaginary unit i from any of the functions, since now it is not possible to use the Hermiticity of J^μ in the way as we did before to derive reality properties of X or Y. Also, as before, there is only one independent non-trivial scalar that can be formed from p' and $\bar{p}$, which we may choose as $p'_\mu\bar{p}^\mu$. However, just as we used $(p'-p)^2$ instead of $p'_\mu p^\mu$ in (3.18), we will make a similar choice

[6] By the arguments that we used to obtain (3.9) we find that $Q\Psi_0$ must be proportional to Ψ_0. However, we can also see that $(\Psi_0, J^\mu(x)\Psi_0)$ must vanish as a result of the Lorentz-transformations properties of J^μ and Ψ_0. (Or, one can also say as a consequence of the Wigner-Eckart theorem.) Hence, the constant of proportionality between $Q\Psi_0$ and Ψ_0 must be zero.

here. We could choose to use $(p' - \bar{p})^2$ in the direct formal analogy with (3.18), but as we will see below, it is better to use $(p' + \bar{p})^2$ instead. That is, we define $k = p' + \bar{p}$ here. The current-conservation condition (3.31) now implies that $X(k^2) = 0$. (Note that the allowed values of k^2 are now those for which $k^2 \geq 4m^2$, where m is the mass of the particle.[7] Hence, there is no need to pay special attention to the case of $k^2 = 0$.) We then obtain

$$\left(\Psi^{\rm out}_{\mathbf{p}',n\,;\bar{\mathbf{p}},n^{\rm c}}, J^\mu(0)\Psi_0\right) = \frac{q}{(2\pi)^3}\frac{1}{\sqrt{2p'^0\ 2\bar{p}^0}}(p' - \bar{p})^\mu Y(k^2). \tag{3.35}$$

We also can now see why the "normalization" condition (3.33) is of no help here. If $\mathbf{p}' = -\bar{\mathbf{p}}$ then $p'^0 = \bar{p}^0$ and the right hand side of (3.35) satisfies (3.33) automatically, without any additional requirements on the function Y.

The second amplitude that we wish to investigate is

$$\left(\Psi_0, J^\mu(x)\Psi^{\rm in}_{\bar{\mathbf{p}}',n^{\rm c};\mathbf{p},n}\right). \tag{3.36}$$

Repeating the steps that we took to obtain (3.35) we arrive at

$$\left(\Psi_0, J^\mu(0)\Psi^{\rm in}_{\bar{\mathbf{p}}',n^{\rm c};\mathbf{p},n}\right) = \frac{q}{(2\pi)^3}\frac{1}{\sqrt{2\bar{p}'^0\ 2p^0}}(-\bar{p}' + p)^\mu Z(k^2), \tag{3.37}$$

where Z is a scalar function of $k^2 = (\bar{p}' + p)^2$.

Now we are ready to look more closely at the question of what is the relation between the functions F, Y and Z. As we will see, they can be thought of as continuations of the same function of complex k^2 into different regions. First, note that for on-shell momenta the squares $k^2 = (p' + \bar{p})^2$ of Y and $k^2 = (\bar{p}' + p)^2$ of Z are always positive, and satisfy $k^2 \geq 4m^2$. To summarize, the function F is defined for values $k^2 \leq 0$ of its argument, while the functions Y and Z are both defined only for values $k^2 \geq 4m^2$ of their arguments. Since the functions Y and Z seem to share the same domain, let us first restrict our attention only to them. To relate the amplitudes (3.29) and (3.36) we can use the operator CPT, the product of the operators of charge conjugation, space inversion, and time reversal[8]. This operator is

[7] This is easy to see in the center-of-mass frame, when $p' = \left(\sqrt{\bar{\mathbf{p}}^2 + m^2}, -\bar{\mathbf{p}}\right)$ and $\bar{p} = \left(\sqrt{\bar{\mathbf{p}}^2 + m^2}, \bar{\mathbf{p}}\right)$. We have

$$k^2 = \left(2\sqrt{\bar{\mathbf{p}}^2 + m^2}, \mathbf{0}\right)^2 = 4\left(\bar{\mathbf{p}}^2 + m^2\right) \geq 4m^2.$$

[8] Sometimes people try to do this by arguing from the Hermiticity of J^μ and pretending that something like the following formula is correct: $\left(\Psi^{\rm out}_{\mathbf{p}';\bar{\mathbf{p}}}, J^\mu(0)\Psi_0\right)^* = \left(\Psi_0, J^\mu(0)\Psi^{\rm in}_{\mathbf{p};\bar{\mathbf{p}}}\right)$. This is similar to what we did in (3.19) but I do not think it works here, since for multi-particle states we no longer have relations such as $\Psi^{\rm in}_{\mathbf{p};\bar{\mathbf{p}}} = \Psi^{\rm out}_{\mathbf{p};\bar{\mathbf{p}}}$.

conserved in any quantum field theory. When applied to a state $\Psi_{\mathbf{p},\sigma,n}$ that corresponds to a single massive particle, the operator P of the space inversion reverses the particle's three-momentum and multiplies the state vector by the intrinsic parity of that particle η_n,

$$\mathrm{P}\Psi_{\mathbf{p},\sigma,n} = \eta_n \Psi_{-\mathbf{p},\sigma,n}. \tag{3.38}$$

The effect of the operator T of the time reversal is to reverse both the particle's three-momentum and its spin z-component, multiply the state by its time-reversal phase factor ζ_n, and also to change the asymptotic "in" state into the appropriate "out" state and vice versa,

$$\begin{aligned} \mathrm{T}\Psi^{\mathrm{in}}_{\mathbf{p},\sigma,n} &= \zeta_n \Psi^{\mathrm{out}}_{-\mathbf{p},-\sigma,n}, \\ \mathrm{T}\Psi^{\mathrm{out}}_{\mathbf{p},\sigma,n} &= \zeta_n \Psi^{\mathrm{in}}_{-\mathbf{p},-\sigma,n}. \end{aligned} \tag{3.39}$$

And, finally, the operation of charge conjugation changes particles into their antiparticles and multiplies the state by its charge-conjugation phase factor ξ_n,

$$\mathrm{C}\Psi_{\mathbf{p},\sigma,n} = \xi_n \Psi_{\mathbf{p},\sigma,n^{\mathrm{c}}}. \tag{3.40}$$

Hence, altogether, the effect of CPT on a single-particle state of a massive particle of spin zero is

$$\begin{aligned} \mathrm{CPT}\Psi^{\mathrm{in}}_{\mathbf{p},n} &= \xi_n \eta_n \zeta_n \Psi^{\mathrm{out}}_{\mathbf{p},n^{\mathrm{c}}}, \\ \mathrm{CPT}\Psi^{\mathrm{out}}_{\mathbf{p},n} &= \xi_n \eta_n \zeta_n \Psi^{\mathrm{in}}_{\mathbf{p},n^{\mathrm{c}}}. \end{aligned} \tag{3.41}$$

It is possible to show[9] that in a theory that conserves the CPT operator one must have

$$\xi^*_n \eta^*_n \zeta^*_n = \xi_{n^{\mathrm{c}}} \eta_{n^{\mathrm{c}}} \zeta_{n^{\mathrm{c}}}. \tag{3.42}$$

When we apply CPT to multi-particle asymptotic states, they transform as direct products of corresponding single-particle states. Thus, if we consider two-particle states consisting of particle–anti-particle pairs, and if we use the fact that the parities η, ζ and ξ are all just phase factors, we find

$$\begin{aligned} \mathrm{CPT}\Psi^{\mathrm{in}}_{\bar{\mathbf{p}}',n^{\mathrm{c}};\mathbf{p},n} &= \Psi^{\mathrm{out}}_{\bar{\mathbf{p}}',n;\mathbf{p},n^{\mathrm{c}}}, \\ \mathrm{CPT}\Psi^{\mathrm{out}}_{\bar{\mathbf{p}}',n;\mathbf{p},n^{\mathrm{c}}} &= \Psi^{\mathrm{in}}_{\bar{\mathbf{p}}',n^{\mathrm{c}};\mathbf{p},n}. \end{aligned} \tag{3.43}$$

Before we use this property we need to recall the important fact that although the operators P and C are linear and unitary, the operator of time inversion is anti-linear and anti-unitary. Anti-linearity means that for any complex number α we have $\mathrm{T}\alpha = \alpha^*\mathrm{T}$. This property right away implies that T cannot be unitary, because then we would have

$$(\Psi, \mathrm{T}\Phi) = \left(\mathrm{T}^{-1}\Psi, \Phi\right),$$

[9] See, e.g., Chapter 5 of [9].

the left-hand side of which would be anti-linear in Φ (because of the anti-linearity of T) and anti-linear in Ψ (because of the property of the scalar product), while the right-hand side would be linear both in Φ and Ψ. In fact, T is an anti-unitary operator, which means that it satisfies

$$(\Psi, \mathrm{T}\Phi) = \left(\mathrm{T}^{-1}\Psi, \Phi\right)^*. \tag{3.44}$$

Thus, we can write

$$\begin{aligned}\left(\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, (\mathrm{CPT})^{-1}J^\mu(x)(\mathrm{CPT})\Psi_0\right)\\ = \left((\mathrm{CPT})\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, J^\mu(x)(\mathrm{CPT})\Psi_0\right)^* = \left(\Psi^{\mathrm{in}}_{\mathbf{p}',n^{\mathrm{c}};\,\bar{\mathbf{p}},n}, J^\mu(x)\Psi_0\right)^*\\ = \left(\Psi_0, J^\mu(x)\Psi^{\mathrm{in}}_{\mathbf{p}',n^{\mathrm{c}};\,\bar{\mathbf{p}},n}\right).\end{aligned} \tag{3.45}$$

One can also show that the electric current operator J^μ transforms as[10]

$$(\mathrm{CPT})^{-1}J^\mu(x)(\mathrm{CPT}) = -J^\mu(-x). \tag{3.46}$$

Hence, we can conclude

$$\begin{aligned}-\left(\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, J^\mu(0)\Psi_0\right) = \left(\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, -J^\mu(0)\Psi_0\right)\\ = \left(\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, (\mathrm{CPT})^{-1}J^\mu(0)(\mathrm{CPT})\Psi_0\right) = \left(\Psi_0, J^\mu(0)\Psi^{\mathrm{in}}_{\mathbf{p}',n^{\mathrm{c}};\,\bar{\mathbf{p}},n}\right),\end{aligned} \tag{3.47}$$

which, together with (3.35) and (3.37), leads to the condition

$$-\frac{q}{(2\pi)^3}\frac{1}{\sqrt{2p'^0\ 2\bar{p}^0}}(p'-\bar{p})^\mu Y(k^2) = \frac{q}{(2\pi)^3}\frac{1}{\sqrt{2p'^0\ 2\bar{p}^0}}(-p'+\bar{p})^\mu Z(k^2). \tag{3.48}$$

This implies that the functions Y and Z must be the same,

$$Y(k^2) = Z(k^2). \tag{3.49}$$

In fact, since Y (or Z) and F have been defined on non-overlapping domains we are free to denote them all by the same symbol F. Let us now explore this further so that we can better understand why it makes a good sense to consider Y, Z and F as a single function. As we briefly mentioned in Sec. 2.2, the matrix element (3.12),

$$\left(\Psi^{\mathrm{out}}_{\mathbf{p}',n}, J^\mu(0)\Psi^{\mathrm{in}}_{\mathbf{p},n}\right),$$

[10] See, again, [9]. One can show that the free photon gauge field $a^\mu(x)$ transforms as $(\mathrm{CPT})^{-1}a^\mu(x)(\mathrm{CPT}) = -a^\mu(-x)$. Since CPT commutes both with H_0 and H this result also extends to the same field expressed in the Heisenberg picture. Electromagnetic currents appear together with the photon gauge field in the gauge-invariant combination $J^\mu A_\mu$. (J^μ, A_μ here represent the full, interacting fields.) This, together with the required commutation properties of the interaction and the CPT inversion, $[\mathrm{CPT}, V] = 0$, allows us to infer the transformation property of any electromagnetic current J^μ under the CPT operator.

can be obtained from the momentum-space Green's function

$$\int d^4x\, d^4y\, \mathrm{e}^{ip'y}\, \mathrm{e}^{-ipx} \left(\Psi_0, T\left\{J^\mu(0)\Phi(y)\Phi^\dagger(x)\right\}\Psi_0\right) \tag{3.50}$$

by the application of the LSZ reduction formula. For this to work, the scalar field $\Phi(x)$ of (3.50) must be such that $\Phi^\dagger(x)\Psi_0$ has non-zero scalar products with single-particle states $\Psi_{\mathbf{p},n}$ of the charged particle. In practice, the application of the LSZ formula here entails bringing the external momenta p and p' of (3.50) on the mass shell (and both with positive zeroth components, because as we discussed earlier, in (3.50) p' exits the diagram and p enters it), amputating the external propagators for the charged particle (which are singular on the mass shell), supplying instead of them appropriate coefficient functions of on-shell momenta p and p' as required by usual Feynman rules for incoming and outgoing particles, and supplying an additional constant factor that depends on the renormalization scheme used for the field $\Phi(x)$. For our purposes these details are not very important, except for the fact that we may disregard the external propagator poles of (3.50) since those are stripped away in (3.12).

The reason why we are considering (3.50) is that while in the original matrix element (3.12) the four-momenta p, p' are necessarily on-shell, in the expression (3.50) we are allowed to take them off the mass shell. In particular, we can attempt to change the sign of the time component of either of the four-momenta — that is, their energy — from positive to negative. Then, if we return to the mass shell again, but with the energy being negative, the integral corresponds to the original amplitude but with the corresponding external line "crossed"; if before that line corresponded to the particle entering the graph now it corresponds to its antiparticle leaving the graph, and vice versa. For instance, we can try to change the four-momentum p into the four-momentum $-p$. If we denote this new value by $\bar{p}$, that is, if we put $\bar{p} = -p$, and apply the LSZ formula, we obtain the matrix element (3.29),

$$\left(\Psi^{\mathrm{out}}_{\mathbf{p}',n;\,\bar{\mathbf{p}},n^{\mathrm{c}}}, J^\mu(0)\Psi_0\right).$$

Similarly, if instead of "crossing" the line carrying the four-momentum p we "cross" the line carrying the four-momentum p', that is, we change the four-momentum p' to $\bar{p}' = -p'$, we obtain an expression that corresponds to the matrix element (3.36),

$$\left(\Psi_0, J^\mu(0)\Psi^{\mathrm{in}}_{\bar{\mathbf{p}}',n^{\mathrm{c}};\,\mathbf{p},n}\right).$$

We can now check that different definitions of k^2 which we chose in form factors F, Y, and Z follow this correspondence. The same can be said also about the four-momentum factors that multiply those form factors in the expressions for the corresponding matrix elements. (For example, in (3.21)

we defined $k = p' - p$, while in (3.35) we defined $k = p' + \bar{p}$, which is the same as the first definition if we put $\bar{p} = -p$. Similarly, the momentum factor multiplying the form factor in (3.21) is $(p' + p)$, while in (3.35) it is $(p' - \bar{p})$.)

This shows why it makes a good sense to denote all three form-factors by the same symbol F. However, there is a problem with this argument. If, for instance, we start with (3.50) and p, p' both on-shell and future-oriented, then $k^2 = (p' - p)^2 < 0$. Now, if we keep all the components of p and p' real and continuously change p and p' along some path that starts at the original values of p and p' and ends with p' back at its original value but p reversed to $-p$, then along this path the value of k^2 must also change continuously, and furthermore must always be real. But since the path starts at a negative value of k^2, $(k^2)_{\text{initial}} < 0$, and ends at a value that satisfies $(k^2)_{\text{final}} \geq 4m^2$, there must be a point on this path when $k^2 = 4m^2$. And this is a problem, because using the methods that we have briefly discussed in Sec. 2.5 it can be shown that at $k^2 = 4m^2$ the Green's function (3.50) has a normal threshold singularity. Thus, in order to connect the region of $k^2 < 0$ with the region $k^2 > 4m^2$, we are forced to venture into the complex plane of k^2. In fact, one can make strong claim that $F(k^2)$ is analytic in the upper half-plane of k^2. There is a branch cut, starting at the lowest threshold at $k^2 = 4m^2$ and going to $k^2 = +\infty$. As we move from the region $k^2 < 0$ towards $k^2 > 4m^2$ we can avoid all normal thresholds that stand in our way by giving k^2 a small positive imaginary part. It can be shown that the physical region, which corresponds to (3.29), (3.36), can indeed be reached by approaching the branch cut from above, that is, by taking the limit of $k^2 = \Re\{k^2\} + i\epsilon$ for $\epsilon \to 0^+$. Furthermore, as we have shown earlier, near the equation (3.19), for $k^2 < 0$ the form factor $F(k^2)$ must be real. Assuming, therefore, that for $\Im\{k^2\} > 0$ the form factor is indeed analytic, one can use the Schwarz's reflection principle to show that F is analytic everywhere on $\mathbb{C} - [0, \infty)$, and, furthermore, that $F\left([k^2]^*\right) = F(k^2)^*$. In fact, one can argue that F is real all the way to the lowest threshold, which then implies that F is analytic on $\mathbb{C} - [4m^2, \infty)$.

3.4 Spin 1/2 particle

Let us now discuss the case of a spin 1/2 particle. In this case the form factor formula for the amplitude (3.1) becomes more complex, because a spin 1/2 particle has more complicated Lorentz-transformation properties than a spin 0 particle. Transformation properties of a spin 1/2 particle are given by

$$U(\Lambda)\Psi_{\mathbf{p},\sigma} = \sqrt{\frac{(\Lambda p)^0}{p^0}} \sum_{\sigma'} D^{\left(\frac{1}{2}\right)}_{\sigma'\sigma}(W(\Lambda, p))\, \Psi_{\mathbf{p}_\Lambda,\sigma'}, \tag{3.51}$$

where $D^{(\frac{1}{2})}_{\sigma'\sigma}(R)$ is the matrix corresponding to the transformation of a spin 1/2 system under the rotation R, and the rotation $W(\Lambda, p)$ is the little group transformation corresponding to the Lorentz transformation Λ and the initial four-momentum p. (See, e.g., [9] for details.) We will not describe here the details[11] but one can show that a general Lorentz-covariant amplitude (3.1) of a spin-1/2 particle can be expressed in the form

$$\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) = i\frac{q}{(2\pi)^3}\bar{u}(\mathbf{p}',\sigma')\Gamma^\mu(p',p)u(\mathbf{p},\sigma), \tag{3.52}$$

where u is the coefficient function of the Dirac field (a precise convention will be specified below) and Γ^μ is a four-vector 4×4 matrix function of p', p and the gamma matrices. We can expand an arbitrary 4×4 matrix in the basis constructed from products of gamma matrices: $\mathbf{1}$, γ^μ, $[\gamma^\mu,\gamma^\nu]$, $\gamma^\mu\gamma_5$, and γ_5. To build a four-vector matrix Γ^μ, using the above matrices and the four-vectors p' and p, we can use a linear combination of the following terms:

- Using the unit matrix $\mathbf{1}$ we can construct two independent four-vector terms $p'^\mu\mathbf{1}$ and $p^\mu\mathbf{1}$.
- Using the matrices γ^μ we can construct the independent terms γ^μ, $p'^\mu p'_\nu\gamma^\nu$, $p^\mu p'_\nu\gamma^\nu$, $p'^\mu p_\nu\gamma^\nu$, and $p^\mu p_\nu\gamma^\nu$.
- From $[\gamma^\mu,\gamma^\nu]$ we can construct the independent four-vectors $[\gamma^\mu,\gamma^\nu]p'_\nu$, $[\gamma^\mu,\gamma^\nu]p_\nu$, $p'^\mu[\gamma^\rho,\gamma^\sigma]p'_\rho p_\sigma$, and $p^\mu[\gamma^\rho,\gamma^\sigma]p'_\rho p_\sigma$.
- Since $\gamma^\mu\gamma_5$ transforms as a pseudo-vector (or, in other words, as an axial vector), we must contract it with $\epsilon_{\mu\nu\rho\sigma}$ to produce a vector-like quantity that transforms as a true vector under space inversions. Hence, only one independent combination is possible: $\epsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5 p'_\rho p_\sigma$.
- The matrix γ_5 transforms as pseudo-scalar. Hence, we again need to create a product of it with $\epsilon_{\mu\nu\rho\sigma}$. However, since there are only two independent momenta p', p available, all such four-vector combinations vanish. (E.g., $\epsilon^{\mu\nu\rho\sigma}p_\nu p_\rho p'_\sigma$ is identically zero.)

All of these terms can be multiplied by any scalar functions of p' and p. As before, we will express those functions in terms of the invariant square of the momentum transfer, $k^2 = (p'-p)^2$.

Considering all these options, it may seem that a general form of Γ^μ is very complicated. Fortunately, one can greatly reduce the number of independent terms by using the fact that the coefficient functions u and $\bar{u}$

[11] Note, however, that if one is satisfied with an argument limited to the perturbation theory then the result follows directly from the Feynman rules and the methods described in Sec. 2.2.

satisfy the Dirac equations[12]

$$(-ip_\mu\gamma^\mu + m)u(\mathbf{p},\sigma) = 0, \tag{3.53}$$

and

$$\bar{u}(\mathbf{p}',\sigma')(-ip'_\mu\gamma^\mu + m) = 0. \tag{3.54}$$

For example, the term $p'^\mu p'_\nu\gamma^\nu$ effectively reduces to $-imp'^\mu$ when multiplied by $\bar{u}$ from the left. Similarly, one can see that all the terms proportional to a single gamma matrix except for γ^μ are equivalent to one of the terms $p'^\mu\mathbf{1}$ or $p^\mu\mathbf{1}$. Furthermore, we have

$$\begin{aligned}\bar{u}(\mathbf{p}')[\gamma^\mu,\gamma^\nu]p'_\nu u(\mathbf{p}) &= \bar{u}(\mathbf{p}')(\{\gamma^\mu,\gamma^\nu\}p'_\nu - 2\gamma^\nu\gamma^\mu p'_\nu)u(\mathbf{p}) \\ &= \bar{u}(\mathbf{p}')(-2p'^\mu + 2im\gamma^\mu)u(\mathbf{p}),\end{aligned}$$

which is a linear combination of $\bar{u}\mathbf{1}u$ and $\bar{u}\gamma^\mu u$. We can handle other terms containing the commutator of two gamma matrices similarly. And lastly, to express $\epsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5 p'_\rho p_\sigma$ as a linear combination of $\mathbf{1}$ and γ^μ we can write

$$\begin{aligned}&\epsilon^{\mu\nu\rho\sigma}\gamma_\nu\gamma_5 p'_\rho p_\sigma = \\ &\quad\frac{i}{6}\left[\gamma^\mu\gamma^\rho\gamma^\sigma + \gamma^\sigma\gamma^\mu\gamma^\rho + \gamma^\rho\gamma^\sigma\gamma^\mu - \gamma^\rho\gamma^\mu\gamma^\sigma - \gamma^\sigma\gamma^\rho\gamma^\mu - \gamma^\mu\gamma^\sigma\gamma^\rho\right]p'_\rho p_\sigma\end{aligned}$$

and move all the matrices γ^ρ to the left and the matrices γ^σ to the right, so that we can apply the Dirac equations. Altogether, we find that it is in effect possible to express Γ^μ as a linear combination of only three terms, p'^μ, p^μ, and γ^μ. Hence, a general Lorentz-covariant expression for the amplitude is

$$\begin{aligned}\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma',n}, J^\mu(0)\Psi^{\text{in}}_{\mathbf{p},\sigma,n}\right) = i\frac{q}{(2\pi)^3}\bar{u}(\mathbf{p}',\sigma')\Big[&\gamma^\mu F(k^2) \\ &- \frac{i}{2m}(p'+p)^\mu G(k^2) - \frac{1}{2m}(p'-p)^\mu H(k^2)\Big]u(\mathbf{p},\sigma).\end{aligned} \tag{3.55}$$

From the Hermiticity of J^μ we can obtain the condition

$$\Gamma^\mu(p,p') = -\beta\Gamma^{\mu\dagger}(p',p)\beta,$$

which means that the functions F, G and H, as defined above, are all real. The current-conservation condition (3.3) requires[13] that $H(k^2) = 0$. That

[12] Warning: This subsection is less detailed, and since we mainly follow the treatment of [9], I decided to use the definition of gamma matrices used in that textbook, to make it easier for the reader to refer to a more complete treatment. This leads to a somewhat unusual convention, when (unlike [9]) we use the metric $\eta = \text{diag}(1,-1,-1,-1)$, and gamma matrices that satisfy $\{\gamma_\mu,\gamma_\nu\} = -2\eta_{\mu\nu}$.

[13] We have

$$(p'-p)_\mu(p'+p)^\mu = p'^2 + p'\cdot p - p\cdot p' - p^2 = m^2 - m^2 = 0,$$

and

$$\bar{u}\gamma^\mu(p'-p)_\mu u = i\bar{u}\left[(-i\gamma^\mu p'_\mu + m) - (-i\gamma^\mu p_\mu + m)\right]u = 0.$$

is, we have obtained the formula

$$\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) = i\frac{q}{(2\pi)^3}\bar{u}(\mathbf{p}',\sigma')\left[\gamma^\mu F(k^2) - \frac{i}{2m}(p'+p)^\mu G(k^2)\right]u(\mathbf{p},\sigma). \quad (3.56)$$

Thus, as we can see, for a spin-1/2 particle there are two independent form factors. To derive the consequence of the normalization condition (3.11) we can use the identity $\{\gamma^\mu, -i\gamma^\nu p_\nu + m\} = 2ip^\mu + 2m\gamma^\mu$ to obtain

$$\bar{u}(\mathbf{p})(-2ip^\mu)u(\mathbf{p}) = \bar{u}(\mathbf{p})(2m\gamma^\mu)u(\mathbf{p}).$$

Also, since $\bar{u}(\mathbf{p},\sigma')u(\mathbf{p},\sigma) = \delta_{\sigma'\sigma}m/p^0$, we arrive at

$$\left(\Psi^{\rm out}_{\mathbf{p},\sigma',n}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) = \frac{q}{(2\pi)^3}\delta_{\sigma'\sigma}\left[\frac{p^\mu}{p^0}F(0) + \frac{p^\mu}{p^0}G(0)\right]. \quad (3.57)$$

Hence, the normalization condition (3.11) in this case reads

$$F(0) + G(0) = 1. \quad (3.58)$$

As we have seen, there are many ways to construct a four-vector matrix from gamma matrices and two four-momenta p' and p but when sandwiched between $\bar{u}$ and u only 3 of the terms are independent. This means that there is a substantial freedom of choice when deciding in what form to express the product $\bar{u}\Gamma^\mu u$. In the rest of this subsection we will describe a representation which is probably the most common. This representation uses the matrices γ^μ and $[\gamma^\mu,\gamma^\nu](p'-p)_\nu$ instead of the pair γ^μ and $(p'+p)^\mu$ employed in (3.56). More specifically, we write

$$\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^\mu(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) = i\frac{q}{(2\pi)^3}\bar{u}(\mathbf{p}',\sigma')\left[\gamma^\mu F_1(k^2) - \frac{i}{2}\left[\gamma^\mu,\gamma^\nu\right](p'-p)_\nu F_2(k^2)\right]u(\mathbf{p},\sigma), \quad (3.59)$$

where F_1 is called the *Dirac form factor* and F_2 the *Pauli form factor* of the particle. To express these form factors in terms of the form factors F and G we need to express the matrix $[\gamma^\mu,\gamma^\nu](p'-p)_\nu$ in terms of matrices γ^μ and $(p'+p)^\mu$ that figure in the original expression (3.56). We have

$$\begin{aligned}\bar{u}(\mathbf{p}',\sigma')&\left(\left[\gamma^\mu,\gamma^\nu\right](p'-p)_\nu\right)u(\mathbf{p},\sigma)\\ &= \bar{u}(\mathbf{p}',\sigma')\left(-2p'_\nu\gamma^\nu\gamma^\mu + \{\gamma^\mu,\gamma^\nu\}p'_\nu - \gamma^\mu\gamma^\nu 2p_\nu + \{\gamma^\mu,\gamma^\nu\}p_\nu\right)u(\mathbf{p},\sigma)\\ &= \bar{u}(\mathbf{p}',\sigma')\left(4im\gamma^\mu - 2(p'+p)^\mu\right)u(\mathbf{p},\sigma).\end{aligned} \quad (3.60)$$

Using this in (3.59) and comparing with (3.56) gives

$$\begin{aligned}F(k^2) =& F_1(k^2) + 2mF_2(k^2),\\ G(k^2) =& -2mF_2(k^2).\end{aligned} \quad (3.61)$$

The normalization condition now reads

$$F_1(0) = 1. \tag{3.62}$$

And lastly, let us remark that it is often useful to define the *electric form factor* G_E and the *magnetic form factor* G_M,

$$\begin{aligned} G_E(k^2) =& F_1(k^2) - \frac{k^2}{2m} F_2(k^2), \\ G_M(k^2) =& F_1(k^2) + 2m F_2(k^2). \end{aligned} \tag{3.63}$$

(Or, in terms of F and G, $G_E = F - \frac{k^2+4m^2}{4m^2} G$ and $G_M = F$.)

3.5 Magnetic dipole moment of a charged particle of spin 1/2

The interaction of a charged particle with an external magnetic field that is time-independent should be well-described by the interaction term of the form

$$V(0) = -\int d^3\mathbf{x}\, \mathbf{J}(\mathbf{x}, 0) \cdot \mathbf{A}(\mathbf{x}), \tag{3.64}$$

where $\mathbf{A}(\mathbf{x})$ is the static vector potential of the external field. Under these circumstances, it should be therefore possible to provide a complete description of that interaction in terms of only $\mathbf{A}(\mathbf{x})$ and form factors of the interacting particle. In case of a spin 1/2 particle, one of the measurable parameters of such an interaction is the magnetic anomaly of that particle, and in this subsection we are going to derive the relation between this quantity and the particle's electromagnetic form factors.

The g-factor g can be defined by the statement that in a weak, static, and slowly (in space) varying magnetic field a stationary particle of spin 1/2 has the interaction matrix elements

$$\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma'}, V\Psi^{\text{in}}_{\mathbf{0},\sigma}\right) = -\frac{gq}{2m} \left(\mathbf{J}^{(\frac{1}{2})}\right)_{\sigma'\sigma} \cdot \mathbf{B}\, \delta^3\left(\mathbf{p}'\right), \tag{3.65}$$

where $\mathbf{J}^{(\frac{1}{2})}$ are the angular momentum matrices for spin 1/2, and the interaction can be expressed as in (3.64). In order to obtain the relation between g and form factors, we will need to take an appropriate limit of (3.56) and compare it to (3.65).

In fact, it will be convenient to use yet another representation of (3.56), which is of the form

$$\begin{aligned} &\bar{u}(\mathbf{p}', \sigma')\Gamma^\mu(p', p)u(\mathbf{p}, \sigma) = \\ &\bar{u}(\mathbf{p}', \sigma') \left[-\frac{i}{2m}(p' + p)^\mu A(k^2) - \frac{i}{2}\left[\gamma^\mu, \gamma^\nu\right](p' - p)_\nu B(k^2)\right] u(\mathbf{p}, \sigma). \end{aligned} \tag{3.66}$$

Using (3.60) we can rewrite this as

$$\bar{u}(\mathbf{p}',\sigma')\Gamma^{\mu}(p',p)u(\mathbf{p},\sigma) = \bar{u}(\mathbf{p}',\sigma')\left[-\frac{i}{2m}(p'+p)^{\mu}\left(A(k^2)-2mB(k^2)\right)+2m\gamma^{\mu}B(k^2)\right]u(\mathbf{p},\sigma). \tag{3.67}$$

Comparing this with (3.56) gives $F = 2mB$ and $G = A - 2mB$, or

$$\begin{aligned} A &= G + F, \\ B &= \frac{1}{2m}F. \end{aligned} \tag{3.68}$$

To summarize, we have found that

$$\left(\Psi^{\text{out}}_{\mathbf{p}',\sigma',n}, J^{\mu}(0)\Psi^{\text{in}}_{\mathbf{p},\sigma,n}\right) = i\frac{q}{(2\pi)^3}\bar{u}(\mathbf{p}',\sigma')\Big[-\frac{i}{2m}(p'+p)^{\mu}\left(G(k^2)+F(k^2)\right) - \frac{i}{4m}\left[\gamma^{\mu},\gamma^{\nu}\right](p'-p)_{\nu}F(k^2)\Big]u(\mathbf{p},\sigma). \tag{3.69}$$

Let us now calculate the spatial components of this amplitude in the limit of very small momenta. In the representation[14]

$$\gamma^0 = -i\begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \qquad \boldsymbol{\gamma} = -i\begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{3.70}$$

we have[15]

$$\mathcal{J}^{ij} = -\frac{i}{4}\left[\gamma^i,\gamma^j\right] = \frac{1}{2}\epsilon^{ijk}\begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}, \tag{3.71}$$

and

$$\mathcal{J}^{i0} = -\frac{i}{4}\left[\gamma^i,\gamma^0\right] = \frac{i}{2}\begin{pmatrix} \sigma_i & 0 \\ 0 & -\sigma_i \end{pmatrix}. \tag{3.72}$$

Here $\boldsymbol{\sigma}$ are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.73}$$

Furthermore,

$$u(\mathbf{0},\frac{1}{2}) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad u(\mathbf{0},-\frac{1}{2}) = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}. \tag{3.74}$$

[14]Reminder: For an ease of comparison, we are using the representation of gamma matrices from [9], although we use a different metric convention. Hence, in terms of our metric, $\{\gamma^{\mu},\gamma^{\nu}\} = -2\eta^{\mu\nu}$.

[15]Latin indices such as i or j assume only the values corresponding to the spatial components, e.g., $i \in \{1,2,3\}$. Also, $\epsilon^{123} = +1$.

Hence, we can obtain

$$\bar{u}(\mathbf{0},\sigma')[\gamma^i,\gamma^j]u(\mathbf{0},\sigma) = 2i\epsilon^{ijk}\,(\sigma_k)_{\sigma'\sigma} = 4i\epsilon^{ijk}\left(J_k^{(\frac{1}{2})}\right)_{\sigma'\sigma}, \tag{3.75}$$

and

$$\bar{u}(\mathbf{0},\sigma')[\gamma^i,\gamma^0]u(\mathbf{0},\sigma) = -\,(\sigma_i)_{\sigma'\sigma} + (\sigma_i)_{\sigma'\sigma} = 0. \tag{3.76}$$

Now we are ready to consider the limit of the space components of (3.69) when $|\mathbf{p}'|, |\mathbf{p}| << m$. To the first order in the momenta we have

$$\begin{aligned}\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^k(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) \to i\frac{q}{(2\pi)^3}\Big[&-\frac{i}{2m}(p'^k + p^k)\delta_{\sigma'\sigma}\,(G(0)+F(0)) \\ &-\frac{1}{m}\epsilon^{klm}(p'-p)_l\left(J_m^{(\frac{1}{2})}\right)_{\sigma'\sigma}F(0)\Big].\end{aligned} \tag{3.77}$$

Recalling now the space-time dependence

$$\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^k(x)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) = \mathrm{e}^{i(p'-p)\cdot x}\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, J^k(0)\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right),$$

we see that in order to obtain the matrix element of the interaction

$$V(0) = -\int d^3\mathbf{x}\,\mathbf{A}(\mathbf{x})\cdot\mathbf{J}(\mathbf{x},0),$$

we just multiply (3.77) with $\exp\left(i(p'-p)\cdot x\right)$, form the scalar product with (minus) the vector potential of the external field, and integrate it over the whole three-space. The term proportional to $\delta_{\sigma'\sigma}\,(\mathbf{p}'+\mathbf{p})\cdot\mathbf{A}(\mathbf{x})$ corresponds to the interaction of the moving charge with the external field, and we are not interested in it. Instead, we will focus on the second term, which can be expressed as below,

$$\begin{aligned}&\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, V\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) \to i\frac{q}{(2\pi)^3}\times \\ &\int d^3\mathbf{x}\;\mathrm{e}^{-i(\mathbf{p}'-\mathbf{p})\cdot\mathbf{x}}\;\frac{1}{m}\mathbf{A}(\mathbf{x})\cdot\left[(\mathbf{p}'-\mathbf{p})\times\left(\mathbf{J}^{(\frac{1}{2})}\right)_{\sigma'\sigma}\right]F(0)+\ldots,\end{aligned} \tag{3.78}$$

where we do not explicitly write the contribution from the first term. Writing $(\mathbf{p}'-\mathbf{p})\;\mathrm{e}^{-i(\mathbf{p}'-\mathbf{p})\cdot\mathbf{x}}$ as $i\nabla\mathrm{e}^{-i(\mathbf{p}'-\mathbf{p})\cdot\mathbf{x}}$ and integrating per partes yields

$$\begin{aligned}&\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, V\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) \to -\frac{q}{(2\pi)^3}\times \\ &\int d^3\mathbf{x}\;\mathrm{e}^{-i(\mathbf{p}'-\mathbf{p})\cdot\mathbf{x}}\;\frac{1}{m}\left(\nabla\times\mathbf{A}(\mathbf{x})\right)\cdot\left(\mathbf{J}^{(\frac{1}{2})}\right)_{\sigma'\sigma}F(0)+\ldots.\end{aligned} \tag{3.79}$$

If we now assume that the magnetic field $\mathbf{B} = \nabla\times\mathbf{A}$ is approximately constant across the space, we can easily evaluate the integral and obtain

$$\left(\Psi^{\rm out}_{\mathbf{p}',\sigma',n}, V\Psi^{\rm in}_{\mathbf{p},\sigma,n}\right) \to -\frac{q}{m}\delta^3(\mathbf{p}'-\mathbf{p})\mathbf{B}\cdot\left(\mathbf{J}^{(\frac{1}{2})}\right)_{\sigma'\sigma}F(0)+\ldots. \tag{3.80}$$

Comparing this with (3.65) then leads to[16]

$$g = 2F(0). \tag{3.81}$$

In terms of the Dirac and Pauli form factors we have $F(0) = F_1(0)+2mF_2(0)$. With help of the normalization condition $F_1(0) = 1$ we can then write

$$g = 2 + 4mF_2(0). \tag{3.82}$$

If we calculate the amplitude (3.1) for a spin 1/2 particle considering only the tree level diagram we obtain $F_2 = 0$. This corresponds to the original Dirac's prediction $g = 2$ in his relativistic quantum mechanics. Also, using the definition $a = (g - 2)/2$ of the *magnetic anomaly* a we can express the result (3.82) more simply as

$$a = 2mF_2(0). \tag{3.83}$$

In other words, the magnetic anomaly corresponds to the value of the Pauli form factor F_2 at $k^2 = 0$. (Note, however, that there is no complete consensus on the exact definition of the form factors and the Pauli form factor is often defined as the whole product $2mF_2(k^2)$. One would then have directly $a = F_2(0)$.)

[16]Note that the terms represented by dots do not contribute if the particle does not move.

Chapter 4

Unitary and Analytic model

At this point we are finally ready to introduce the Unitary and Analytic model. (From now on, we will use the abbreviation U&A for "Unitary and Analytic".) This model provides an approximate description of electromagnetic form factors of hadrons. At its core the model is determined by the particular analytical structure that it imposes on the form factor. It describes the form factor as an analytic function that has exactly two branch points, both of them of the square-root type, and a number of particle poles corresponding to resonances which couple to the virtual photon. Since the number of resonances that enter the model is not determined in any fixed way, in a certain sense the U&A model is a framework rather than a model. Within that framework we can construct different models for the same form factor and — even more importantly — we can construct various models for different form factors. It is this universality that is perhaps the most attractive feature of this phenomenological framework.[1]

Before we move on to heuristically introduce the model, let us first summarize some of the key properties that this model aims to satisfy:

- We will take for granted the assumption that on the physical sheet the form factor is analytic everywhere except for its normal thresholds, which are located on the positive real axis of the Mandelstam variable s. It can be shown that the lowest threshold is always of the square-root type (see, e.g. [1]). The U&A model works under the approximation that there are only two normal thresholds, both of which are assumed to be of the square-root type. The lower of those two thresholds is called the *elastic* threshold and corresponds to the lowest threshold of the real form factor. The higher one, on the other

[1] It should be noted, however, than when the model is applied to hadrons that contain more than two valence quarks an additional care must be taken to ensure an acceptable asymptotic behavior at large energies. The established procedure (see [18]) for dealing with these cases becomes cumbersome rather quickly as the number of valence quarks grows. We will briefly return to this topic at the end of this chapter.

hand, does generally not correspond to any specific physical threshold. Instead, it plays the role of an "effective" threshold, meant to approximate the cumulative effect of all higher thresholds of the real form factor. We will call it the *inelastic* threshold. Since these two thresholds are both of the square-root type, they generate a four-sheeted Riemann surface, and in accordance with our assumption that there are no other singularities present in the form factor on the physical sheet, we need to make sure that our model is analytic on the first sheet of its Riemann surface.

- The interaction between the virtual photon and the hadron is envisaged as being mediated by vector meson resonances. We need to consider resonances that have the same (conserved or approximately conserved) quantum numbers as the photon. This means that the resonances must be spin 1 particles with the space inversion parity η and the charge conjugation parity ξ both equal to -1.[2] Since the electromagnetic current has both isospin 0 and 1 components, we may consider both isoscalar and isovector resonances. We will place a selected number of vector meson resonance poles on the Riemann surface on which the U&A model is defined. Since on the physical sheet the form factor should be analytic, we place those poles only on the unphysical sheets — that is, only on the second, the third or the fourth sheet of the Riemann surface, but never on its first sheet.

- The property of Hermitian analyticity requires that on the physical sheet the form factor is real below the lowest threshold. To make sure that this property holds, we provide with each resonance pole also its associated "shadow" pole located at the complex conjugate position. Note that the reality of the form factor below the lowest threshold is related to the condition of unitarity. (Recall our mention of how the Hermitian analyticity follows from the generalized unitarity in Sec. 2.5.) It is partially because it satisfies this property that the model has the word "unitary" in its name.

[2]Recall that values of internal parities, such as η, ξ or ζ are in general not fully determined by the theory. This is because we can combine their corresponding operators — P, C, or T — with any operators of internal symmetry to define new operators P, C, and T. For example, if P is conserved, then so is also the operator $e^{i(\alpha Q+\beta B+\gamma L)}$P, for any real values of α, β, and γ, where Q, B, L are the operators of the electric charge, baryon number, and lepton number, respectively. This new operator can serve as the operator of space inversion just as well as does the original operator P. Thus, we can always adjust the numbers α, β, and γ to fix the internal space-inversion parities of any three particles, provided that they have linearly independent values of Q, B, and L. In fact, people often use this freedom of choice of the operator P to set the internal parities η of the proton, the neutron, and the electron to be all equal to $+1$. However, internal parities of completely neutral particles that carry no conserved internal charges — such as the photon — are independent of such re-definitions of P or C, and are therefore meaningful on their own. (For further discussion of this topic, see, e.g., Chapter 3 of [9].)

- We also need to make sure that the model is correctly normalized and that it possesses a suitable asymptotic behavior at high energies, as predicted by the QCD. (Although we will ignore the logarithmic corrections.) Note that the first requirement, the normalization, is always simple to satisfy. The second requirement, on the other hand, is satisfied for mesons automatically, but requires special care if ones wishes to formulate the model for hadrons that consist of more than two valence quarks.

- Under these assumptions, on the physical sheet the form factor is fully determined by the location of its lowest branch point and the discontinuity across the branch cut. (This follows, for instance, from the dispersion relation (4.1) derived below.) We select resonances and adjust locations of their poles and values of their residues, as well as the position of the inelastic threshold, in such a way as to faithfully approximate the discontinuity of the real form factor. As a rule, this is achieved by fitting the model to the experimental data for the corresponding total cross section. Since it is possible to derive from unitarity a constraint on this discontinuity, this is another sense in which the model can be said to be unitary.

In the following sections we will more closely introduce the model. We will focus mainly on the general structure rather than on specific technical details. Let me also mention that some aspects of the presentation given here are my own and different from earlier treatments. The reader interested in a more traditional, and in some sense also more detailed treatment, is advised to consult the review [18]. If one is interested in the history of the model, a good first step could be to look at the work [19], which is the earliest English-language reference that I know of in which an early version of the U&A model was presented.

4.1 From dispersion relation for form factor to vector meson poles and branch cuts

In this section we are going to motivate the introduction of the U&A model. We will choose a path that is not the most direct one, but which will lead us to several other useful concepts along its way.

We will start by formulating what is commonly known as the dispersion relation for the form factor. It is a consequence of the following two properties. The first one of them is the assumption that on the physical sheet the form factor is analytic everywhere except for the interval $[s_0, \infty)$, with the positive number s_0 representing the location of the lowest threshold. The second assumption regards the asymptotic properties of the form factor for large values of its argument. In accordance with QCD we assume that for

large values of $|s|$ the modulus $|F(s)|$ of the form factor decreases at least as fast as $|s|^{-1}$. Then, if we integrate the expression $F(s')/(s'-s)$ — as a function of s' — over the contour depicted in Fig. 4.1, in the limit when the radius of the circle goes to infinity, we can easily evaluate the integral using the residue theorem. This yields the value $2\pi i F(s)$ for the whole integral, because the integration contour is traversed in the counterclockwise direction and the residue of $F(s')/(s'-s)$ at $s'=s$ is $F(s)$. Next, we observe that because of the asymptotic properties of F the part of the integral over the large circle vanishes. Only the parts of the contour that go right below or right above the branch cut contribute, and since $F(s^*) = F(s)^*$, their overall value equals

$$\int_{s_0}^{\infty} ds' \left(F(s'+i\epsilon) - F(s'-i\epsilon)\right) = \int_{s_0}^{\infty} ds' \left(F(s'+i\epsilon) - F(s'+i\epsilon)^*\right) = 2i \int_{s_0}^{\infty} ds' \Im\{F(s'+i\epsilon)\},$$

where ϵ is positive and infinitesimal. Combining these results, we find that

$$F(s) = \frac{1}{\pi} \int_{s_0}^{\infty} ds' \, \frac{\Im\left\{F(s'+i\epsilon)\right\}}{s'-s}. \tag{4.1}$$

This is a very useful result on its own. We can see that the discontinuity across the cut fully determines the form factor. There are various ways how one can attempt to determine or approximate this discontinuity. Here we will introduce a simple but useful approach, which basically corresponds to the vector meson dominance model.

Roughly speaking, in this approach one describes the interaction of the virtual photon and a hadron as being fully mediated by vector meson resonances which have the same conserved quantum numbers as the photon. Let us suppose that we are considering in total N such resonances and denote them by the label $r = 1, \ldots, N$. In the first approximation we pretend that these resonances appear as single-particle contributions in the sum over intermediate states of *definite* total four-momentum. Hence, in this approximation we consider them as if they were stable particles, which means that their associated poles should appear in the form factor on the physical sheet. These poles need to be at values $s = s_r = m_r^2$ that correspond to the squares of their approximate energies. Of course, since these particle are not really stable, their energies all lie above the lowest threshold: $s_r > s_0$ for all r. This means that these resonance poles all lie inside the interval $[s_0, \infty)$ and are therefore consistent with the analytic properties of the form factor that we assumed during the derivation of the dispersion relation (4.1). If we assume that no other intermediate states contribute to the discontinuity of F, then we can write

$$\Im\{F(s)\} = \pi \sum_{r=1}^{N} a_r m_r^2 \delta(s - m_r^2), \tag{4.2}$$

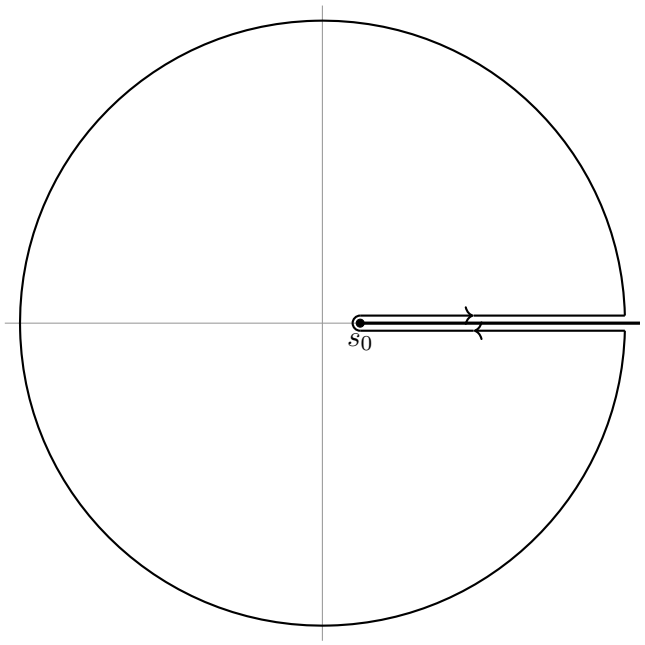


Figure 4.1: The integration contour that can be used to obtain the dispersion relation (4.1) for the form factor. We assume that the form factor is analytic everywhere except for the interval that starts at the lowest threshold (at $s_0 > 0$) and continues to infinity along the positive real axis. In the limit of infinite radius the part of the integral over the large circle vanishes as a consequence of the assumed asymptotic properties of the form factor. (Note that the integral over the small left semicircle centered at $s = s_0$ does not contribute anything beyond what has already been included in (4.1) as the discontinuity at $s = s_0$. However, for this to be true that discontinuity at $s = s_0$ must included in the integral, so strictly speaking the integration bounds in (4.1) should be from $s_0 - \delta$ to ∞, in the limit that $\delta \to 0^+$. Under normal circumstances, however, this does not make a difference in (4.1), so we will ignore it here.)

where a_r are some unknown coefficients. (Those coefficients are related to particular coupling constants, but we will not need to use that detail here.) If we use this approximation in the dispersion relation (4.1), we can easily calculate the corresponding form factor

$$F(s) = \sum_{r=1}^{N} a_r \frac{m_r^2}{m_r^2 - s}. \tag{4.3}$$

This is the Vector Meson Dominance (VMD) model of the form factor. (See [20] for a more complete discussion of this approach.)

There are several problems with the function (4.3). First of all, there should also be contributions of other intermediate states to the discontinuity, not just the contribution from resonances. Starting from the VMD model (4.3), however, it is not at all straightforward to include these other contributions, since resonances are not orthogonal to multi-particle states of definite momentum. The second problem, related to the first one, is that the electromagnetic form factor should have no poles on the physical sheet, and, furthermore, that resonances should not be treated as stable particles. The U&A model does partly address some of these issues. It retains the VMD model's fixation on the resonance poles, but introduces two normal thresholds, and this allows one to treat resonances in a more appropriate way by hiding them behind the branch cuts. This also means that the resonances no longer figure in the discontinuity as if they were stable particles.

As was already mentioned several times, in the U&A model there are two branch points. The branch point that is located at the higher value of s is called the inelastic threshold and its position is usually left as a free parameter that is determined by a comparison with data. On the other hand, the lower, elastic threshold should correspond to the physical lowest threshold, and its position is therefore fully determined. Let us now discuss where should this elastic threshold be located. Before we proceed we need to make it clear that form factors that we are trying to describe by the U&A model are meant to correspond to matrix elements (3.1) evaluated with electromagnetic and weak interactions switched off. These form factors represent only the effects of strong interactions. The lowest mass strongly-interacting particles that exist in scattering states are pions. Since in the current chapter we will work under the approximation that the isotopic spin symmetry is an exact symmetry of strong interactions, both the charged and the neutral pion will be described as having the same mass. Generally, one would expect that the elastic threshold is the threshold for the production of two pions, but that is not completely correct, because sometimes the two-pion intermediate states are not allowed. This a consequence of the fact that we often consider isospin zero and isospin one components of the electromagnetic current separately. (Sometimes this is forced upon us, but often it is just a convenient choice.) To see how this works, consider some

general state consisting of a pion and its anti-particle described in its center-of-mass frame,

$$\Psi = \sum_t \int d^3\mathbf{p}\, \chi(\mathbf{p}, t; -\mathbf{p}, -t)\, a^\dagger(\mathbf{p}, t) a^\dagger(-\mathbf{p}, -t)\, \Psi_0, \tag{4.4}$$

where Ψ_0 is the vacuum state, $a^\dagger(\mathbf{p}, t)$ is the operator that creates the pion of three-momentum $\mathbf{p}$ and isospin 3-component t and $\chi(\mathbf{p}_1, t_1; \mathbf{p}_2, t_2)$ is the wave-function for this two-particle state. Note that for the pion the variable t can have values -1, 0, and $+1$, corresponding to the particles π^-, π^0, and π^+, respectively, and that π^- is the anti-particle of π^+, and π^0 is its own anti-particle. This is the reason why the isospin 3-components in (4.4) are specified as t and $-t$. Now, pions are bosons, which means that the creation operators $a^\dagger(\mathbf{p}, t)$ commute with each other. As a consequence, we may replace $\chi(\mathbf{p}, t; -\mathbf{p}, -t)$ in (4.4) with its symmetrized version $\{\chi(\mathbf{p}, t; -\mathbf{p}, -t) + \chi(-\mathbf{p}, -t; \mathbf{p}, t)\}/2$ without changing the state Ψ. In other words, we can assume without the loss of generality that under the exchange of both the momentum and the isospin variables the wave-function χ is symmetric. However, if the state Ψ is to be produced from a virtual photon it must have the same total angular momentum as the photon: its total angular momentum must be 1. But because pions are spin zero particles this implies that the orbital angular momentum of the two-pion state Ψ must be 1, and χ is therefore anti-symmetric under the exchange of the momentum variables, $\chi(\mathbf{p}, t; -\mathbf{p}, -t) = -\chi(-\mathbf{p}, t; \mathbf{p}, -t)$. It must therefore be also anti-symmetric when only the isospin variables are exchanged. But a direct product of two $j = 1$ representations of the group SU(2) decomposes into a direct sum of one $j = 2$, one $j = 1$ and one $j = 0$ irreducible representations. When we exchange the order of the two $j = 1$ systems that form the direct product, the components of the irreducible representations with $j = 2$ and $j = 0$ remain unchanged and the components of the $j = 1$ representation are multiplied by -1. In other words, only the part with the total isospin 1 is anti-symmetric under the exchange of the isospin variables of the two pions, the other two irreducible representations are symmetric. To summarize, we found that if the state Ψ of (4.4) has the orbital angular momentum $l = 1$, as it does if it is produced by a virtual photon, then its value of the total isospin must be 1. This means that if a form factor corresponds to the isovector part of the electromagnetic current — the part with isospin 1 — then its lowest threshold is indeed at $s = 4m_\pi^2$, where m_π stands for the pion's mass. However, if a form factor corresponds instead to the isoscalar part of the electromagnetic current — which is the isospin 0 component — then two-pion intermediate states are not allowed. In such cases there is no singularity at $s = 4m_\pi^2$ and the lowest threshold is located at $s = 9m_\pi^2$, which corresponds to the opening of the three-pion channel.

Note that the two-pion state (4.4) with χ that is anti-symmetric both under $t \to -t$ and under $\mathbf{p} \to -\mathbf{p}$ has the same value of the charge conjugation

parity as the photon. (If that was not the case, then two-pion intermediate states would be forbidden even for isovector components of form factors.) We can verify this by directly applying the charge conjugation operator C on Ψ of (4.4)

$$\begin{aligned}
\mathrm{C}\Psi &= \sum_t \int d^3\mathbf{p}\, \chi(\mathbf{p},t;-\mathbf{p},-t)\, \left[\mathrm{C}a^\dagger(\mathbf{p},t)\mathrm{C}^{-1}\right]\left[\mathrm{C}a^\dagger(-\mathbf{p},-t)\mathrm{C}^{-1}\right]\mathrm{C}\Psi_0 \\
&= \sum_t \int d^3\mathbf{p}\, \chi(\mathbf{p},t;-\mathbf{p},-t)\, \left[\xi_t\, a^\dagger(\mathbf{p},-t)\right]\left[\xi_{-t}\, a^\dagger(-\mathbf{p},t)\right]\Psi_0 \\
&= \sum_t \int d^3\mathbf{p}\, \chi(\mathbf{p},t;-\mathbf{p},-t)\, a^\dagger(\mathbf{p},-t)a^\dagger(-\mathbf{p},t)\Psi_0 \\
&= \sum_t \int d^3\mathbf{p}\, \chi(\mathbf{p},-t;-\mathbf{p},t)\, a^\dagger(\mathbf{p},t)a^\dagger(-\mathbf{p},-t)\Psi_0 = -\Psi, \quad (4.5)
\end{aligned}$$

where we denoted by ξ_t the charge conjugation parity of the pion with isotopic spin 3-component t. To obtain the third line from the second line we used the fact that the product of the charge conjugation parities of a particle and its anti-particle is always $+1$. This implies that $\xi_t\xi_{-t} = 1$. In the last step we used the anti-symmetry of χ under the interchange of isospin variables. The result $\mathrm{C}\Psi = -\Psi$ means that the charge conjugation parity of Ψ is -1, the same as that of the photon.

Both the elastic and the inelastic threshold of the U&A model are of the square-root type. Together they therefore generate a four-sheeted Riemann surface. In the core of the U&A model lies a particular transformation that maps this four-sheeted Riemann surface in a one-to-one fashion onto a single complex plane, in which the form factor is then constructed. We will introduce this mapping in the next section.

4.2 Unfolding the Riemann surface

There are many ways to map a four-sheeted Riemann surface onto a single complex plane. The U&A model employs a convenient choice of such a map, but since that choice is not unique in any sense that I am aware of, I will not attempt here to derive that mapping. Instead I will simply introduce that mapping and describe its properties.

However, before we start discussing the transformation employed in the U&A model, let us first briefly consider a simpler case, in which we assume that the form factor $F(s)$ is an analytic function with only a single branch point. This will allow us to introduce in a more accessible setting some of the elements that will be employed later in the U&A model. We assume here that the (only) branch point of $F(s)$ is of the square-root type, and that it is located at $s = s_0 > 0$. We choose the branch cut to go along the positive real axis from s_0 to infinity. In accordance with our earlier discussion we

assume that for $s > s_0$ the physical region can be reached by approaching the branch cut on the physical sheet from above. (Note that the physical sheet is the first sheet of the Riemann surface.) If we then cross the branch cut from above, we appear on the second sheet of the Riemann surface, and if we continue further along a circle around s_0 in the clockwise direction — that is, if we encircle the branch point on this second, unphysical sheet — and then cross the branch cut from above once again, we return to the first sheet and appear just below the branch cut. This is because the branch point is assumed to be of the square-root type, which means that the Riemann surface has only two sheets. The most straightforward way of mapping this Riemann surface onto a single complex plane is to define the variable q as

$$q = \sqrt{s - s_0}, \tag{4.6}$$

where $\sqrt{\cdot}$ is the branch of the square root with the branch cut on the positive real axis, defined in such a way that on its first sheet it maps positive real numbers onto positive real numbers. This means that we have

$$\sqrt{r\mathrm{e}^{i\phi}}\big|_{\mathrm{I.}} = \sqrt{r}\mathrm{e}^{i\phi/2}, \text{ where } 0 \leq \phi < 2\pi, \tag{4.7}$$

where we used the subscript "I." to indicate that this holds on the first sheet of $s = r\mathrm{e}^{i\phi}$. (And $\sqrt{r}$ is just the usual positive square root of a positive real number.) On its second sheet, this square root is instead

$$\sqrt{r\mathrm{e}^{i\phi}}\big|_{\mathrm{II.}} = \sqrt{r}\mathrm{e}^{i(\pi+\phi/2)} = -\sqrt{r\mathrm{e}^{i\phi}}\big|_{\mathrm{I.}}, \text{ where } 0 \leq \phi < 2\pi. \tag{4.8}$$

Clearly, the function (4.6) lives on a two-sheeted Riemann surface of s with the branch point at $s = s_0$, and with the branch cut going from s_0 further along the positive real axis. It maps the branch point to $q = 0$, and the physical sheet of the Riemann surface — which corresponds to the first sheet of $\sqrt{\cdot}$ — onto the upper half-plane of q. It maps the second sheet onto the lower half-plane.

Note that at the branch point the map $q(s)$ is not analytic and, in particular, it does not preserve angles there. This becomes very obvious if we consider the real line on the first sheet of the Riemann surface, which clearly is quite an important line for the application of the form factor $F(s)$. If $s < s_0$ it is unambiguous whether we are on the first or the second sheet, but if $s > s_0$ we are right on the branch cut, so a greater care needs to be taken. In accordance with the definition (4.7) the physical region corresponds to the line that can be reached by approaching the branch cut on the physical sheet from above. (For this reason it might be better to just say that we are considering the line $\mathbb{R} + i\epsilon$, $\epsilon \to 0^+$, on the physical sheet of s.) Now, if we inspect (4.6) we can easily see that q maps this line onto the curve that starts at "$+i\infty$" on the q-plane, goes along the imaginary axis straight down to $q = 0$ — which corresponds to the branch point $s = s_0$ — and from

there continues along the positive real semi-axis to infinity. The fact that at $q = 0$ this straight line suddenly breaks at a right angle illustrates that $q = q(s)$ is singular[3] at $s = s_0$.

Now, suppose that we wish to construct a model of $F(s)$ on this two-sheeted Riemann surface. One way we could do that is by constructing a suitable function $\hat{F}(q)$ defined on the complex plane of q, and then defining $F(s) = \hat{F}(q(s))$. An advantage of this is that all the complications related to the branch structure of the Riemann surface are automatically taken care of by the transformation $s \to q(s)$ of (4.6). For example, if $\hat{F}$ is analytic everywhere on the q-plane, then F is an analytic function everywhere on the Riemann surface of s, except for the single branch point at s_0. According to our previous discussion, however, this would not be the most appropriate choice of $\hat{F}$. Indeed, $F(s)$ should be analytic — but only on the physical sheet! Hence, we would like to choose $\hat{F}(q)$ that is analytic on the upper half-plane of q, but on the second sheet we expect $F(s)$ to have some singularities. In the spirit of the present chapter we want those singularities to be resonance poles. Suppose that there is only a single resonance available, of mass $m > s_0$ and decay rate Γ. As we mentioned in Sec. 2.6, the "Breit–Wigner" pole at $s = (m - i\Gamma/2)^2$ is most directly accessed from the physical region by moving downwards across the branch cut. It is therefore located on the second sheet of s. Just as we expected, on the q-plane this pole is located in the lower half-plane. We can calculate its position directly from (4.6), but we need to be careful to choose the formula for the second sheet of s — that is, we must use the square root (4.8). Denoting the position of this pole by q_1, we have

$$q_1 = \sqrt{m^2 - \Gamma^2/4 - s_0 - im\Gamma}\big|_{\text{II.}},$$

which is located in the lower right quadrant of the q-plane. As we also discussed in Sec. 2.6, this pole should have an associated "shadow" pole located at the complex conjugate position. That is, the "shadow" pole is at $s = (m + i\Gamma/2)^2$ on the sheet that is accessed from the physical sheet by crossing the branch cut from below. In our case, this is again the second sheet of s. Denoting the "shadow" pole's position on the q-plane by q_2, we have

$$q_2 = \sqrt{m^2 - \Gamma^2/4 - s_0 + im\Gamma}\big|_{\text{II.}}.$$

This point lies in the lower left quadrant of the q-plane. In fact, we can easily see that q_1 and q_2 are located symmetrically at mirror positions with respect to the imaginary axis: we have $\Re\{q_1\} = -\Re\{q_2\}$ and $\Im\{q_1\} = \Im\{q_2\}$. This can be expressed more economically as $q_2 = -q_1^*$. Now, if we choose their respective residues such that one is minus the complex conjugate of the

[3] As is well-known, analytic functions preserve angles. This is the reason why another term that can be used to say that a function is analytic is to say that it is *conformal*.

other, then resulting form factor will be real on the imaginary axis.[4] But the positive imaginary axis of q corresponds to the values of $s < s_0$ on the physical sheet, so we see that this choice of $\hat{F}(q)$ leads to $F(s)$ that satisfies the condition of Hermitian analyticity.

Let us now without further delay proceed to the discussion of the mapping used in the U&A model. The Riemann surface now has four sheets, which are generated by two branch points of the square-root type. The elastic threshold is located at $s = s_0$ and the inelastic threshold is located at $s = s_{\rm in}$, and we assume that $0 < s_0 < s_{\rm in}$. We take care of the first branch point in the exactly same way as we did in (4.6), that is, we transform from s to q by

$$q = \sqrt{s - s_0}, \tag{4.9}$$

where the square root is the one of (4.6), that is, it satisfies (4.7) and (4.8). However, this time we also need to incorporate the second branch point at $s = s_{\rm in}$. Let us denote the value of q that corresponds to $s = s_{\rm in}$ by $q_{\rm in}$,

$$q_{\rm in} = q(s_{\rm in})\big|_{\rm I.} = \sqrt{s_{\rm in} - s_0}\big|_{\rm I.} > 0. \tag{4.10}$$

We will introduce the second branch cut by applying one more transformation, from q to W, so that together we are going to have a transformation $s \to q(s) \to W(q(s))$. Now, if we just defined

$$W(q) = \sqrt{q - q_{\rm in}},$$

in a direct analogy with (4.9) the new branch cut would be present on the physical sheet of s if we approached the $[s_0, \infty)$-cut from above (and, of course, towards some $s > s_{\rm in}$) — because that corresponds to values

[4] We assume that F vanishes at least as fast as $|s|^{-1}$ for large arguments. This means that $\hat{F}$ vanishes at least as $|q|^{-2}$ as $|q| \to \infty$. Thus, if we integrate $\hat{F}(q')/(q' - q)$ over a large circle, we find

$$\hat{F}(q) = \frac{c_1}{q - q_1} + \frac{c_2}{q - q_2},$$

where c_1 and c_2 are the residues of $\hat{F}$ at q_1 and q_2, respectively. Now, if $c_1 = -c_2^*$ and if we denote $c_1 = c$ then (recall that $q_2 = -q_1^*$)

$$\hat{F}(q) = \frac{c}{q - q_1} - \frac{c^*}{q + q_1^*},$$

and so for q imaginary

$$\hat{F}(q)^* = \frac{c^*}{-q - q_1^*} - \frac{c}{-q + q_1} = \hat{F}(q), \text{ if } \Re\{q\} = 0.$$

That is, on the imaginary axis the function $\hat{F}$ is real. (This is just a minor modification of an argument related to the Schwarz's reflection principle.) Note that this example also shows that it can be reasonable to wonder if physical quantities such as F might be fully or almost fully determined by their branch structure and locations of their poles and values of their residues. An early exploration of this and related questions can be found in the excellent article [21].

of q that are in the upper half-plane and have a positive real component $\Re\{q\} > q_{\text{in}}$ — but the new branch cut would be absent if we approached the cut from below, because that corresponds to values of q with negative real part (satisfying $\Re\{q\} < -q_{\text{in}}$), and the square root (4.7),(4.8) does not have any branch cut in the left half-plane. We can address this issue by considering mappings that contain terms dependent not only on $\sqrt{q - q_{\text{in}}}$, but also on $\sqrt{-q - q_{\text{in}}}$. (Because in $\sqrt{-q - q_{\text{in}}}$ the branch cut which starts at $s = s_{\text{in}}$ is visible on the physical sheet if we approach the corresponding location in the s-plane from below, though not if we approach it from above.)

The transformation employed in the U&A model is in a sense such a mapping. It contains terms that are related to $\sqrt{q - q_{\text{in}}}$ and $\sqrt{-q - q_{\text{in}}}$, but it combines them in such a way that the first two sheets of s are mapped onto the unit disk. The transformation reads

$$W(q) = i\frac{(q_{\text{in}} + q)^{\frac{1}{2}} - (q_{\text{in}} - q)^{\frac{1}{2}}}{(q_{\text{in}} + q)^{\frac{1}{2}} + (q_{\text{in}} - q)^{\frac{1}{2}}}. \tag{4.11}$$

We will now proceed to discuss this rather complicated formula. First of all, the form presented here is the one usually encountered in the literature. Literature, however, does never seem to properly discuss the branch structure of this expression.[5] Now, if in the physical region we move along the real axis of s in the positive direction then at the point $s = s_0$ we encounter the lowest threshold. According to our convention we want the associated branch cut to continue from $s = s_0$ further along the positive real axis. Next,

[5]Let us take a step back to make sure we understand what does that mean. (Since thinking about branches of rather complicated expressions such as (4.11) can become very confusing.) The Riemann surface does not know anything about branch cuts or branches, only about branch points. Branch cuts are only a tool — similar to coordinates — that can be helpful when we want to speak about that Riemann surface but they do not have any deeper physical or mathematical meaning. Hence, if we consider the function $W(q(s))$ given by (4.9) and (4.11) strictly as a map from the Riemann surface of s onto the W-plane, then definitions of branch cuts and labeling of branches of the square roots $\sqrt{\cdot}$ and $(\cdot)^{\frac{1}{2}}$ are not substantial. The only thing that is affected by those conventions is where on the W-plane are located the branch cuts of s. If we want the branch cuts to be located as in Fig. 4.2, then we need to use definitions of square roots as specified in the main text. Note, however, that just by themselves these branch cuts have absolutely no effect on functions $\hat{F}(W)$ that we may define on the W-plane. To reiterate: the branch cuts are nothing more than a part of the system of coordinates (composed of the variable s and a discrete label that determines what sheet we are on) that we use to speak about the points on the Riemann surface. Hence, if we make sure that the physical region is located on the W-plane correctly [the location of the physical region is depicted in Fig. 4.2; note that when, for some real s, we calculate the physical value of some form factor $F(s)$ that is defined using a function $\hat{F}(W)$ we need to know to what W does that s correspond, and the answer is: among the four solutions $W(q(s))$ it is the one that lies in the physical region of W] and that the function $\hat{F}(W)$ is implemented correctly [which in our case means that the poles must be at the correct spots, as depicted in Fig. 4.6] then we may ignore the question of branch cuts altogether.

as we move (still along the real line) to even higher values of s we eventually encounter also the second, inelastic threshold, which is at $s = s_{\rm in}$. In our convention, its associated branch cut then also continues further in the positive direction along the real axis. As we discussed above, the transformation $q(s)$ of (4.9) already reproduces the first branch cut correctly. We also mentioned that the terms $\sqrt{q - q_{\rm in}}$ and $\sqrt{-q - q_{\rm in}}$, if employed together in a suitable fashion, could then correctly reproduce the second branch cut. But in (4.11) the square root has arguments $q_{\rm in} - q$ and $q_{\rm in} + q$, rather than $q - q_{\rm in}$ and $-q - q_{\rm in}$. Since these two sets are related by the reversal of the overall sign, we would expect that a different branch of the square root is needed. In fact, it turns out that the convention that works as intended is the following. In (4.9) we use the square root $\sqrt{\cdot}$ defined in (4.7),(4.8), but in (4.11) we use the square root $(\cdot)^{\frac{1}{2}}$, that has the branch cut on the negative real axis. More specifically, we have

$$\left(r\mathrm{e}^{i\phi}\right)^{\frac{1}{2}}\big|_{\rm I.} = \sqrt{r}\mathrm{e}^{i\phi/2}, \text{ where } -\pi \le \phi < +\pi, \tag{4.12}$$

and

$$\left(r\mathrm{e}^{i\phi}\right)^{\frac{1}{2}}\big|_{\rm II.} = -\left(r\mathrm{e}^{i\phi}\right)^{\frac{1}{2}}\big|_{\rm I.}, \text{ where } -\pi \le \phi < +\pi. \tag{4.13}$$

We adhere to this distinction between the two definitions of the square root throughout this whole chapter: $\sqrt{\cdot}$ refers to (4.7),(4.8), and $(\cdot)^{\frac{1}{2}}$ refers to (4.12), (4.13). I caution anyone who decides to implement this model to check that the branch structure is correct. (However, see also the footnote 5.) In particular, it is a good idea to verify that the physical region of s (that is, the line $\mathbb{R} + i\epsilon$ on the first Riemann sheet) is mapped onto the W-plane as described below.

The complete transformation $s \to W(q(s))$ defined by (4.9) and (4.11) maps the four-sheeted Riemann surface of s onto the single complex plane of W. Its branch structure is depicted in Fig. 4.2. As one can verify,[6] it maps

[6]One way to do this is to implement the transformation in a computer and inspect its plots. But, of course, one can also proceed analytically. A possible direction from which one can attack this problem is to rewrite (4.11) as

$$W(q) = i\frac{(q_{\rm in}+q)^{\frac{1}{2}} - (q_{\rm in}-q)^{\frac{1}{2}}}{(q_{\rm in}+q)^{\frac{1}{2}} + (q_{\rm in}-q)^{\frac{1}{2}}} = i\frac{\frac{(q_{\rm in}+q)^{\frac{1}{2}}}{(q_{\rm in}-q)^{\frac{1}{2}}} - 1}{\frac{(q_{\rm in}+q)^{\frac{1}{2}}}{(q_{\rm in}-q)^{\frac{1}{2}}} + 1}.$$

The complicated step is then to show that in this expression we can write $\sqrt{\frac{q_{\rm in}+q}{q_{\rm in}-q}}$ instead of $\frac{(q_{\rm in}+q)^{\frac{1}{2}}}{(q_{\rm in}-q)^{\frac{1}{2}}}$. In this presentation I skip this technical step. But if we take it for granted, then we can analyze

$$W(q(s)) = i\frac{\sqrt{\frac{q_{\rm in}+q(s)}{q_{\rm in}-q(s)}} - 1}{\sqrt{\frac{q_{\rm in}+q(s)}{q_{\rm in}-q(s)}} + 1},$$

the physical sheet onto the left half of the unit disk. The second sheet, which is accessed from the physical sheet by crossing the branch cut between s_0 and $s_{\rm in}$ from above, is mapped onto the right half of the unit disk. If we instead cross the branch cut above the inelastic threshold we appear on the third sheet, and this sheet is mapped onto left half-plane minus the unit disk. If, from the third sheet we cross the branch cut between s_0 and $s_{\rm in}$ we appear on the fourth sheet, and this sheet is mapped onto the right half-plane minus the unit disk. The point $s = s_0$ is mapped onto the points $W = 0$ and $W = \infty$ and the point $s = s_{\rm in}$ onto the points $W = +i$ and $W = -i$. If, in the physical region of s, we move along the real axis from minus to plus infinity, the corresponding path in the W-plane starts at $W = -1$, which corresponds to $s = -\infty$, continues along the real axis to $W = 0$, which corresponds to $s = s_0$, then continues upwards along the imaginary axis to $W = +i$, which corresponds to $s = s_{\rm in}$, and finally goes counterclockwise along the unit circle from $+i$ back to -1.

It will be useful to also know the inverse transformation to $W(q(s))$. Note that from (4.11) we find

$$W - \frac{1}{W} = i\frac{(q_{\rm in}+q)^{\frac{1}{2}} - (q_{\rm in}-q)^{\frac{1}{2}}}{(q_{\rm in}+q)^{\frac{1}{2}} + (q_{\rm in}-q)^{\frac{1}{2}}} - (-i)\frac{(q_{\rm in}+q)^{\frac{1}{2}} + (q_{\rm in}-q)^{\frac{1}{2}}}{(q_{\rm in}+q)^{\frac{1}{2}} - (q_{\rm in}-q)^{\frac{1}{2}}} = i\frac{4q_{\rm in}}{2q}, \quad (4.14)$$

which (using (4.9) and (4.10)) means that

$$\left[W - \frac{1}{W}\right]^2 = -4\frac{s_{\rm in} - s_0}{s - s_0}. \quad (4.15)$$

The inverse transformation, $W \to s$, is therefore

$$s = s_0 - \frac{4(s_{\rm in} - s_0)}{[W - 1/W]^2}. \quad (4.16)$$

Note that from this equation it is very easy to see what is the relation between values of W that correspond to the same value of s but on different sheets. Since the variable W enters the right-hand side of (4.16) only in

as a successive application of the simple square root $s \to q(s)$, the spin transformation $q \to u_1(q) = \frac{q+q_{\rm in}}{-q+q_{\rm in}}$, another square root $u_1 \to u_2(u_1) = \sqrt{u_1}$, another spin transformation $u_2 \to u_3(u_2) = \frac{u_2-1}{u_2+1}$, and the multiplication by i, $u_3 \to W = iu_3$. The point of all of this is that each of those steps is easy to analyze. This goes without saying for the square root and the multiplication by i. But spin transformations (also known as Möbius transformations) are also simple to consider. As is well-known, they transform lines or circles onto lines or circles (or, equivalently, circles onto circles on the associated Riemann sphere). For details about this topic see, for instance, the textbook [22]. For an interesting discussion of spin transformations in relativistic physics see the first chapter of [23].

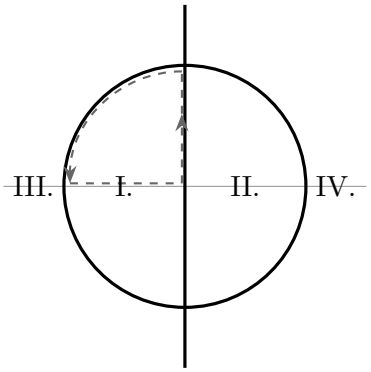


Figure 4.2: *The branch structure projected onto the W-plane.* This scheme depicts the basic structure of the mapping $s \to W(q(s))$, as given by (4.9) and (4.11). Branch cuts are depicted by thick lines. As indicated in the figure, the first sheet of the Riemann surface of s is mapped onto the left unit half-disk, the second sheet onto the right half-disk, and the third and the fourth sheet onto the left and right half-planes minus the unit disk, respectively. The dashed line sketches the path taken by W as s goes from $-\infty$ to $+\infty$ in the physical region. The path starts at $W = -1$, goes straight to $W = 0$, then upwards to $W = i$ and then counterclockwise along the circle back to $W = -1$.

the combination $[W - 1/W]^2$, the value of s on the left-hand side is unchanged if we substitute either $W \to -W$ or $W \to 1/W$. That is, for each solution $W(s)$ of (4.16) there are three more solutions $-W(s)$, $1/W(s)$, and $-1/W(s)$, each of them corresponding to a different sheet of the Riemann surface of s. We could also recognize this directly from (4.11). Suppose we start at some point of the Riemann surface on which the value of s is s_1, the value of q is q_1 and the value of W is W_1. Then, if we cross the branch cut between s_0 and $s_{\rm in}$, and return to the original s but on the other sheet, the value of q changes from q_1 to $-q_1$. The equation (4.11) then implies that W changes from W_1 to $-W_1$. Suppose, next, that instead of crossing the branch cut between s_0 and $s_{\rm in}$ we cross it above the threshold $s_{\rm in}$. This corresponds to crossing the branch cut associated with the elastic threshold but also of the branch cut associated with the inelastic threshold. (Because we decided to draw them one over another.) Crossing the branch cut associated with the elastic threshold just reverses the sign of q. But crossing the branch cut associated with the inelastic threshold has also an effect. Either the sign of $(q_{\rm in} - q)^{\frac{1}{2}}$ changes, if the real part of q is positive, or the sign of $(q_{\rm in} + q)^{\frac{1}{2}}$ changes, if the real part of q is negative. Either one of those two cases leads to the same effect on W: $W \to -1/W$. If we combine this with the effect $W \to -W$ from crossing the "elastic" branch cut, the overall effect of crossing the branch cut above $s = s_{\rm in}$ is to change the value of W

from W_1 to $1/W_1$.

4.3 Constructing the model on the W-plane

We are now ready to construct the U&A model of the form factor. We will do this by first constructing a suitable function $\hat{F}(W)$ defined on the W-plane, and then defining the form factor $F(s)$ — a meromorphic function on the four-sheeted Riemann surface of s — by the relation

$$F(s) = \hat{F}\left(W(q(s))\right). \tag{4.17}$$

Here the mapping $W(q(s))$, which we have introduced in the previous section, introduces into F two branch points of the square-root type. This mapping is defined by the equations (4.9), (4.10), (4.11), in which the specific branches (4.7), (4.8) and (4.12), (4.13) of the square root are used.

In our approach, we will start from the VMD model (4.3). Our strategy will be to first rewrite the VMD model in the form (4.17), which entails finding a suitable function $\hat{F}$ defined on the W-plane which — when used in (4.17) — reproduces on the physical sheet the VMD model. Afterwards, we will introduce non-zero decay rates for the resonances. As we will see, this cannot be done directly with the VMD-model $\hat{F}$-function, because the result of that would not satisfy the condition of Hermitian analyticity, and would have resonance poles present on the physical sheet, just like the VMD model (4.3) does if we use complex masses m_r in its denominators. As we will see, we are going to need to adjust the function $\hat{F}$ to avoid these issues.

Let us consider a single term of the VMD model (4.3),

$$F^{(r)}_{\mathrm{VMD}}(s) = \frac{m_r^2}{m_r^2 - s}, \tag{4.18}$$

where we have removed the coefficient a_r because at this point it is not going to concern us. The subscript r corresponds to a single resonant particle which is associated with this term of the VMD model. Of course, in general we will want to consider the effect of several resonances, and in the U&A model we deal with this in the same way as is done in the VMD model — we will add together several terms, each corresponding to a single resonance. Since our goal now is to express $F^{(r)}_{\mathrm{VMD}}$ in terms of the variable W, we will need to define some constants that correspond to special points on the W-plane related to the function (4.18). First, the resonance pole is at $s = m_r^2$ and there are four points associated with it on the W-plane. These are the values of $W(q(m_r^2))$ on the four branches of (4.9), (4.11), or — equivalently — the four solutions of (4.16) for $s = m_r^2$. We choose *any* one of these four points and denote it by $W_r^{(0)}$. The superscript (0) is meant to indicate that at this stage we are still working in the approximation that the resonance is

stable and its pole is at a real value of s. With this definition of $W_r^{(0)}$ the equation (4.16) gives

$$m_r^2 = s_0 - 4\frac{s_{\rm in} - s_0}{\left[W_r^{(0)} - 1/W_r^{(0)}\right]^2}. \tag{4.19}$$

It will be very convenient to also define the point W_N that corresponds to the value $s = 0$, that is, $W_N = W(q(0))$. Of course, there is again one such a value of W for each sheet of s, and we will define W_N to be that value of $W(q(0))$ which corresponds to the physical sheet. This means that W_N lies on the interval $(-1, 0)$ on the W-plane. We have

$$0 = s_0 - 4\frac{s_{\rm in} - s_0}{[W_N - 1/W_N]^2}. \tag{4.20}$$

Using these two results together with the inverse map (4.16), we can write

$$\begin{aligned}
F_{\rm VMD}^{(r)} = \frac{m_r^2 - 0}{m_r^2 - s} &= \left(\frac{1}{\left[W_r^{(0)} - 1/W_r^{(0)}\right]^2} - \frac{1}{[W_N - 1/W_N]^2}\right) \\
&\quad \Big/ \left(\frac{1}{\left[W_r^{(0)} - 1/W_r^{(0)}\right]^2} - \frac{1}{[W - 1/W]^2}\right) \\
&= \frac{[W - 1/W]^2}{[W_N - 1/W_N]^2} \times \frac{W_N^2 + 1/W_N^2 - {W_r^{(0)}}^2 - 1/{W_r^{(0)}}^2}{W^2 + 1/W^2 - {W_r^{(0)}}^2 - 1/{W_r^{(0)}}^2} \\
&= \frac{[1 - W^2]^2}{[1 - W_N^2]^2} \times \frac{W_N^4 + 1 - W_N^2\left({W_r^{(0)}}^2 + 1/{W_r^{(0)}}^2\right)}{W^4 + 1 - W^2\left({W_r^{(0)}}^2 - 1/{W_r^{(0)}}^2\right)} \\
&= \frac{[1 - W^2]^2}{[1 - W_N^2]^2} \times \frac{\left(W_N^2 - {W_r^{(0)}}^2\right)\left(W_N^2 - 1/{W_r^{(0)}}^2\right)}{\left(W^2 - {W_r^{(0)}}^2\right)\left(W^2 - 1/{W_r^{(0)}}^2\right)}.
\end{aligned} \tag{4.21}$$

Thus, we have found that

$$\begin{aligned}
\frac{m_r^2}{m_r^2 - s} &= \left(\frac{1 - W^2}{1 - W_N^2}\right)^2 \\
&\times \frac{\left(W_N - W_r^{(0)}\right)\left(W_N + W_r^{(0)}\right)\left(W_N - 1/W_r^{(0)}\right)\left(W_N + 1/W_r^{(0)}\right)}{\left(W - W_r^{(0)}\right)\left(W + W_r^{(0)}\right)\left(W - 1/W_r^{(0)}\right)\left(W + 1/W_r^{(0)}\right)}.
\end{aligned} \tag{4.22}$$

To summarize, what we have found is that if the constants $W_r^{(0)}$ and W_N are defined as is specified near the equations (4.19) and (4.20), and if W and s satisfy the equation (4.16), then the VMD-model term $m_r^2/(m_r^2 - s)$ equals the right-hand side of (4.22). Note that equality is satisfied for W anywhere on the W-plane, provided that s has the value dictated by (4.16). This means that if we put

$$\hat{F}^{(r)}_{\rm VMD}(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^2 \times \frac{\left(W_N - W_r^{(0)}\right)\left(W_N + W_r^{(0)}\right)\left(W_N - 1/W_r^{(0)}\right)\left(W_N + 1/W_r^{(0)}\right)}{\left(W - W_r^{(0)}\right)\left(W + W_r^{(0)}\right)\left(W - 1/W_r^{(0)}\right)\left(W + 1/W_r^{(0)}\right)}, \quad (4.23)$$

then $F(s) = \hat{F}^{(r)}_{\rm VMD}(W(q(s)))$ is a function defined on the four-sheeted Riemann surface with branch points at $s = s_0$ and $s = s_{\rm in}$ that reproduces on *each* sheet of the Riemann surface the function $F^{(r)}_{\rm VMD}(s)$ of (4.18). Of course, unlike $F^{(r)}_{\rm VMD}(s)$, the function $\hat{F}^{(r)}_{\rm VMD}(W(q(s)))$ is not defined at the points $s = s_0$ and $s = s_{\rm in}$, where $W(q(s))$ is singular, but everywhere else the two functions appear identical. In particular the function $\hat{F}^{(r)}_{\rm VMD}(W(q(s)))$ has no discontinuity across the branch cut except for the point $s = m_r^2$, where the resonance pole is located. Nevertheless, even if the branch cuts of $\hat{F}^{(r)}_{\rm VMD}(W(q(s)))$ are in a sense "invisible", they do exist and we will able to make good use of them when we introduce a non-zero width of the resonance in our next step. But before we do that, let us first better investigate the function $\hat{F}^{(r)}_{\rm VMD}(W)$.

It is convenient to consider the function $\hat{F}^{(r)}_{\rm VMD}$ of (4.23) as consisting of two parts. The factor $\left[(1-W^2)/(1-W_N^2)\right]^2$ is called the *asymptotic* factor, and we will discuss it later in this chapter. The remaining factor,

$$\frac{\left(W_N - W_r^{(0)}\right)\left(W_N + W_r^{(0)}\right)\left(W_N - 1/W_r^{(0)}\right)\left(W_N + 1/W_r^{(0)}\right)}{\left(W - W_r^{(0)}\right)\left(W + W_r^{(0)}\right)\left(W - 1/W_r^{(0)}\right)\left(W + 1/W_r^{(0)}\right)},$$

is called the *finite-energy* factor. It is of the form

$$\prod_{p\in P} \frac{W_N - p}{W - p}, \quad P = \{W_r^{(0)}, -W_r^{(0)}, 1/W_r^{(0)}, -1/W_r^{(0)}\}. \quad (4.24)$$

As we can see, apart from the normalization terms in the numerators the finite-energy factor is just the product of pole factors for the four solutions of (4.16) for $s = m_r^2$. This is the reason why it did not matter which of those four solutions we singled out when we defined the constant $W_r^{(0)}$ near the equation (4.19): in the function $\hat{F}^{(r)}_{\rm VMD}$ all the four solutions appear,

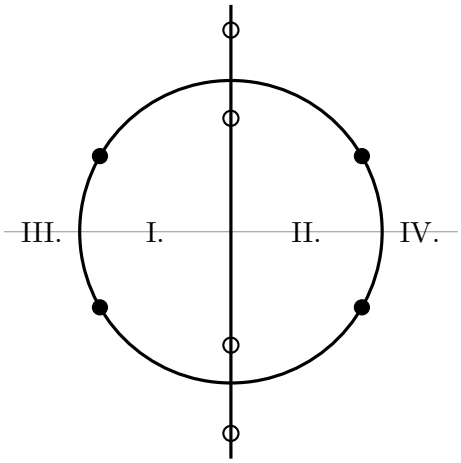


Figure 4.3: *Resonance poles of the function* $\hat{F}^{(r)}_{\rm VMD}$. Resonance poles lie on the branch cuts, because resonances are treated here as if they were stable. We illustrate one example of the case when the resonance lies below the inelastic threshold — we depicted positions of the corresponding poles by empty circles — and one example of the case when the resonance lies above the inelastic threshold — those poles' positions are depicted by full circles.

independent on the original choice of $W_r^{(0)}$. Since we are still ignoring the fact that the resonance decays, these resonance poles are located on the branch cuts. As is depicted in Fig. 4.3, if $s_0 < m_r^2 < s_{\rm in}$ then on the W-plane the poles are located on the imaginary axis, and if $s_{\rm in} < m_r^2$ then the poles are located on the unit circle. [This is because, as we mentioned, the physical region corresponding to the interval $s_0 < s < s_{\rm in}$ is mapped onto the interval $(0, i)$ of the W-plane. If $W_r^{(0)}$ lies in this interval, then $-W_r^{(0)}$ lies in $(-i, 0)$, $1/W_r^{(0)}$ lies in $(-i\infty, -i)$, etc. The physical region interval $s_{\rm in} < s$ is mapped onto the upper left quadrant of the unit circle, and if $W_r^{(0)}$ lies there, then $1/W_r^{(0)} = W_r^{(0)^*}$, and the remaining poles also lie on the circle.]

All that we have achieved so far is that we put the VMD model on our four-sheeted Riemann surface. What we would like to do now is to introduce the instability of the resonance, while making sure that the form factor remains real below the lowest threshold and that there are no resonance poles on the physical sheet. As we discussed in Sec. 2.6, the resonance pole should be located at the value of $s = s_r$ given by

$$s_r = (m_r - i\Gamma_r/2)^2, \tag{4.25}$$

where Γ is the decay rate of the resonance. However, if we simply substituted $W_r^{(0)} \to W_r$ in (4.23), where W_r would be the value

$$W_r = W(q(s_r))$$

on any of the sheets, the resonance poles would be located as illustrated in Fig. 4.4. The four solutions of (4.16) for $s = s_r$ are still related by the operations $W_r \to -W_r$ and $W_r \to 1/W_r$ and there is exactly one solution present for each sheet of the associated Riemann surface. It is therefore clear that this does not lead to a satisfactory form factor. First, there are resonance poles on the physical sheet. (On the W-plane these are located in the lower half of the left unit disk, because this is the part that corresponds to the lower half-plane of s.) And furthermore, the form factor is no longer real below the lowest threshold. This is because the physical region below the lowest threshold corresponds to the interval $(-1, 0)$ on the W-plane. Hence, if $\hat{F}(W)$ is to be real on this interval the poles must located at complex conjugate positions[7], which — as we can see from Fig. 4.4 — they are clearly not. If we take a step back we realize that none of these results should be a surprise. When the decay rate was zero, the function $\hat{F}^{(r)}_{\rm VMD}$ corresponded to the VMD model (4.18). But $\hat{F}^{(r)}_{\rm VMD}$ is analytic in the parameter $W_r^{(0)}$ (except for the pole), just as the VMD model (4.18) is in the parameter m_r^2. Hence, if we set the decay rate to a positive value the function $\hat{F}^{(r)}_{\rm VMD}$ must correspond to the VMD model (4.18) but with $s_r = (m_r - i\Gamma_r/2)^2$ substituted for m_r^2. And we already knew that this has a pole on the physical sheet and is not real below the lowest threshold — after all, that is exactly the reason why we are trying to formulate the U&A model! What we need to do now is to take some inspiration from our discussion in Sec. 2.6 and make small adjustments to positions of the poles in Fig. 4.4 in order to make the model more satisfactory.

Before we introduce the needed adjustments let us first state what properties that we want our model to satisfy:

1. First, as we recognized already in Sec. 2.6, the property of Hermitian analyticity requires that to each resonance pole there must exist a complex conjugate "shadow" pole. We cannot therefore limit ourselves to considering only poles at $s = s_r = (m_r - i\Gamma_r/2)^2$. For each such a pole we must include also its "shadow" pole, which can be accessed from the physical region by taking the complex conjugate path to any path by which we can access the resonance pole at $s = s_r$. That is, the "shadow" pole is located at $s = s_r^* = (m_r + i\Gamma_r/2)^2$, on the sheet that from the physical region can be accessed by the complex conjugate path to a path that can be taken to access the sheet where its associated resonance pole at $s = s_r$ is located. Note that on the W-plane the resonance pole and its associated "shadow" pole are located at complex conjugate positions. (This follows from the fact that on the Riemann surface the two poles can be reached by complex conjugate paths and that complex conjugate paths on the Riemann surface are

[7] This follows from the Schwarz's reflection principle.

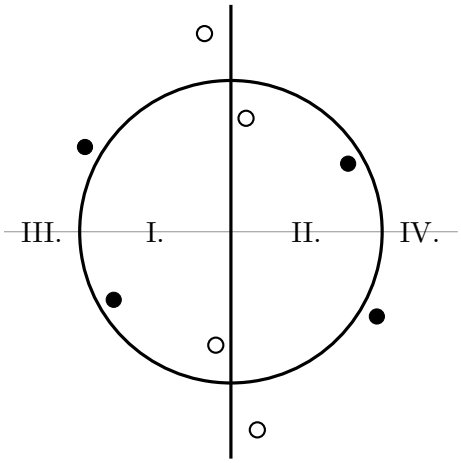


Figure 4.4: *Resonance poles of the function $\hat{F}^{(r)}_{\mathrm{VMD}}$ after the introduction of a positive decay width.* This figure illustrates possible locations to which the resonance poles depicted in Fig. 4.3 move after we introduce a non-zero decay rate.

mapped onto complex conjugate paths on the W-plane.)

2. Second, we do not want any poles to be present on the physical sheet.

3. We consider the VMD model to be a good approximation in the limit $\Gamma \to 0^+$. More specifically, we require that for $\Gamma = 0$ our model and the VMD model given by $\hat{F}^{(r)}_{\mathrm{VMD}}$ coincide, and that the two models approach each other in a continuous manner as $\Gamma \to 0^+$.

We will try to achieve this by adjusting the finite-energy factor (4.24). Note that as we adjust the value of $W_r^{(0)}$ (or change the value of Γ) in (4.24) the total number of poles remains unchanged (except for some special cases, such as $W_r^{(0)} = 0$, when some of the poles coincide or are pushed away to the infinity). We expect the same to be true for the U&A finite-energy factor. That means that as we vary Γ, the poles generally do not appear or disappear, they only move on the W-plane. However, the property 2 implies that for $\Gamma > 0$ some of the poles that are present in the VMD finite-energy factor (4.24) must not be present in the U&A finite-energy factor, and the property 1 implies that the U&A finite energy factor must contain some poles that are not present in the VMD finite-energy factor. But for $\Gamma = 0$ both the factors should have the same poles. The simplest way to achieve this is to select a suitable subset of VMD poles and make them — in the U&A finite energy factor — to move on some different, carefully chosen paths as Γ is gradually increased to positive values. For a fixed small value of $\Gamma > 0$ this just means that to obtain the U&A model we slightly shift

a subset of VMD poles to suitable locations on the W-plane. We will now proceed to present how one does do that in the U&A model.

Let us start by inspecting the positions of the poles of $\hat{F}^{(r)}_{\mathrm{VMD}}$ when $\Gamma > 0$. These are depicted in Fig. 4.4. Empty circles indicate the positions of the poles corresponding to $s = s_r$ for some particular value of m_r^2 that lies between s_0 and s_{in}. The pole that lies in the right half of the unit disk is the correct Breit-Wigner pole. On the Riemann surface of s this pole lies just below the physical region, on the second sheet. However, its corresponding "shadow" pole is missing. At the same time, there is an extra, unwanted pole present in the left half of the unit disk. We need to remove this pole from that half-disk, because we must not have any poles on the physical sheet. We can do this by moving this pole to the right half-plane, symmetrically with respect to the imaginary axis of W. This way the pole disappears from the physical sheet and reappears on the second sheet at the position $s = s_r^*$ — as the "shadow" of the resonance pole! We still need to discuss what to do with the poles on the third and the fourth sheet. The problem with them is that they are not at complex conjugate positions. To solve this, we move one of them symmetrically across the imaginary axis of W to the position on the other sheet of s that corresponds to $s = s_r^*$. This way we obtain another pair of complex conjugate poles. The only question is whether we should move the pole that is on the fourth sheet (to the third sheet) or the pole that is on the third sheet (to the fourth sheet). From what has been said in this chapter both options are reasonable. However, in the U&A model we move the pole that is on the third sheet, and place it on the fourth sheet.[8] Altogether, we move the poles as indicated in Fig. 4.5. We use a similar approach also when $m_r^2 > s_{\mathrm{in}}$. Positions of the poles of $\hat{F}^{(r)}_{\mathrm{VMD}}$ for such a case and $\Gamma > 0$ are depicted in Fig. 4.4 by full circles. Just as before, we move the pole that lies in the left half-disk across the branch cut to the position corresponding to $s = s_r^*$, but this time we must move the pole to the third sheet. (Recall our requirement that for small $\Gamma > 0$ the adjustments in the poles' positions should also be small.) This way that pole becomes the "shadow" pole of the pole that has been already present on the third sheet at $s = s_r$. And similarly as before, we move the pole that is on the second sheet to the fourth sheet to provide a "shadow" pole for the pole that has already been there. Again, these adjustments are indicated by

[8] I think that a reasonable argument in favor of this choice is that if we gradually adjust the coupling constant to make the resonance stable, as we were discussing in Sec. 2.6, we want the stable particle poles to be present on the first and the third sheet, which means that the resonance poles should be on the second and the fourth sheets. Historically, however, this choice in the U&A model is a consequence of a heuristic argument used in the model's original derivation. There one adjusts the VMD function $\hat{F}^{(r)}_{\mathrm{VMD}}$ in such a way as to obtain a function that is manifestly Hermitian analytic, but which is equivalent to the original VMD function if $\Gamma = 0$. Then one sets $\Gamma > 0$ and if the care is taken to avoid the appearance of poles on the physical sheet, one acquires a configuration that is presented here.

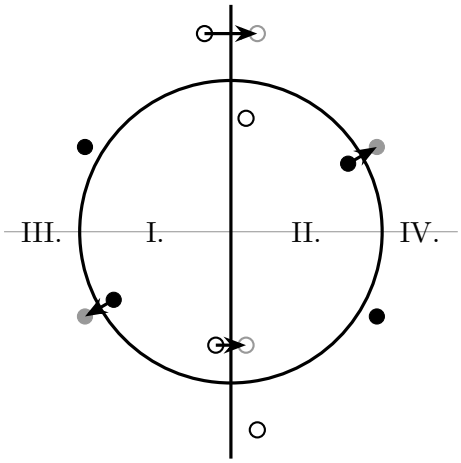


Figure 4.5: *How to pass from a $\Gamma > 0$ VMD model to the U&A model.* When decay rates are non-zero the VMD model starts manifesting certain unphysical features. We can fix some of those by shifting some of the resonance poles in the W-plane, as indicated by the arrows. The resulting form factor is analytic on the physical sheet and real below the lowest threshold.

arrows in Fig. 4.5.

This concludes the introduction of the core aspects of U&A model. Just as we indicated in (4.17) we construct the form factor $F(s)$ as a function on the four-sheeted Riemann surface in terms of a meromorphic function $\hat{F}_{\mathrm{U\&A}}(W)$ defined on the W-plane as

$$F(s) = \hat{F}_{\mathrm{U\&A}}\left(W(q(s))\right).$$

In analogy with the VMD model (4.3), the function $\hat{F}_{\mathrm{U\&A}}(W)$ is constructed as a sum of terms, with each of the terms corresponding to a single resonance,

$$\hat{F}_{\mathrm{U\&A}}(W) = \sum_{r=1}^{N} a_r \hat{F}^{(r)}_{\mathrm{U\&A}}(W). \tag{4.26}$$

And, as we just described, the functions $\hat{F}^{(r)}_{\mathrm{U\&A}}$ are of the form

$$\hat{F}^{(r)}_{\mathrm{U\&A}}(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^2 \prod_{p\in P} \frac{W_N - p}{W - p}, \tag{4.27}$$

where P is the set of positions of the poles of $\hat{F}^{(r)}_{\mathrm{U\&A}}$. This set contains locations of two resonance poles — these are at the values of W corresponding to $s = s_r = (m_r - i\Gamma/2)^2$ — and their two associated "shadow" poles, which are at the values of W corresponding to $s = s_r^* = (m_r + i\Gamma/2)^2$. As we discussed above, if $s_0 < m_r^2 < s_{\mathrm{in}}$, then these poles are from the second and

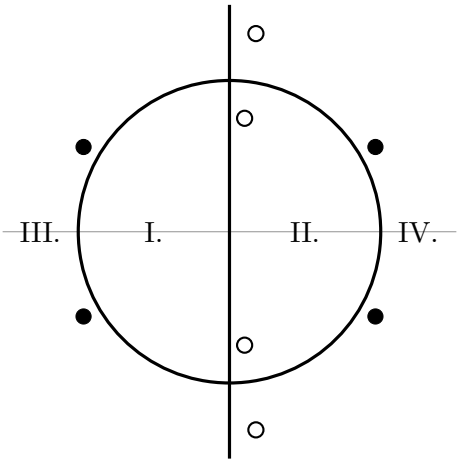


Figure 4.6: *Positions of poles of the function* $\hat{F}^{(r)}_{\mathrm{U\&A}}$. Empty circles indicate the pole locations for a case when $s_0 < m_r^2 < s_{\mathrm{in}}$, and the full circles indicate their locations for a case when $s_{\mathrm{in}} < m_r^2$.

the fourth sheet, and if $s_{\mathrm{in}} < m_r^2$, then the poles are from the third and the fourth sheet. Positions of these poles are illustrated in Fig. 4.6.

This ensures that the function $\hat{F}_{\mathrm{U\&A}}(W)$ yields a form factor $F(s) = \hat{F}_{\mathrm{U\&A}}(W(q(s)))$ that satisfies the following properties:

1. The form factor is analytic everywhere except for two branch points of the square root type and a number of resonance poles (and their "shadow" poles).

2. There are no poles on the physical sheet.

3. In the physical region the function is real below the lowest threshold.

There are two remaining properties that we wish to discuss in this chapter, the normalization condition and the asymptotic behavior of the form factor.

4.3.1 Normalization condition in the U&A model

As we saw in the chapter 3, the value of a form factor at $s = 0$ is usually fixed by a normalization condition of the form,

$$F(0) = C. \tag{4.28}$$

Since the function $\hat{F}^{(r)}_{\mathrm{U\&A}}$ of (4.27) has been explicitly constructed to satisfy $\hat{F}^{(r)}_{\mathrm{U\&A}}(W_N) = 1$, and since in the physical region we have $W(q(0)) = W_N$, the U&A model yields

$$F(0) = \hat{F}_{\mathrm{U\&A}}(W_N) = \sum_{r=1}^{N} a_r \hat{F}^{(r)}_{\mathrm{U\&A}}(W_N) = \sum_{r=1}^{N} a_r. \tag{4.29}$$

Thus, to satisfy the normalization condition (4.28) we just need to make sure that

$$\sum_{r=1}^{N} a_r = C. \tag{4.30}$$

4.3.2 Asymptotic behavior of the U&A model

There is one more general property that we want the U&A model to satisfy. We want the model to follow the correct asymptotic behavior for large values of $|s|$, although we will ignore the logarithmic corrections. Using quantum field theoretical arguments, one can show [24, 25, 26] that a form factor $F(s)$ of a hadron that consists of n_q valence quarks is expected to behave as

$$|F(s)| \propto |s|^{1-n_q} \text{ for } |s| \to \infty. \tag{4.31}$$

This means that electromagnetic form factors of mesons, such as the form factor of the pion, are expected to vanish as $|s|^{-1}$ for large values of $|s|$, while form factors of baryons, such as the form factors of the proton, are expected to vanish as fast as $|s|^{-2}$.

The standard approach to ensure that the U&A model satisfies the asymptotic property (4.31) starts from the observation that when $|s|$ is large then both the elastic and inelastic thresholds are far away. Their effects can therefore be neglected and the effect of the resonance poles is asymptotically the same as the effect of the corresponding pole in the VMD model. Hence, for large values of $|s|$ the U&A model (4.17), (4.26) and the corresponding VMD model (4.3) follow the same behavior.

In the usual approach [27, 28] one therefore studies how to make the VMD model to follow the correct asymptotics, and then directly applies the same method to the U&A model. A detailed description of this procedure can be found in [18, 28, 27]. Here I will address only the main ideas.

Note that the VMD model (4.3)

$$F(s) = \sum_{r=1}^{N} a_r \frac{m_r^2}{m_r^2 - s}$$

contains a sum of terms that decrease each as s^{-1} for large values of $|s|$. Hence, unless some special condition is satisfied, the VMD model is expected to decrease asymptotically as $|s|^{-1}$. This is the expected asymptotic behavior of form factors of mesons, which means that if we are constructing a form factor of a meson, no special care is needed — the VMD model, as well as the U&A model, satisfy the correct asymptotic property automatically. But, of course, that does not mean that the VMD model cannot decrease faster than $|s|^{-1}$. A function of the form (4.3) can in principle vanish as fast as $|s|^{-N}$, where N is the number of resonance terms in the model. For

example, if $N = 2$, the model (4.3) reads

$$F(s) = a_1 \frac{m_1^2}{m_1^2 - s} + a_2 \frac{m_2^2}{m_2^2 - s} = \frac{-s\left(a_1 m_1^2 + a_2 m_2^2\right) + \left(a_1 + a_2\right) m_1^2 m_2^2}{\left(m_1^2 - s\right)\left(m_2^2 - s\right)},$$

and although for general values of the parameters a_1, a_2, m_1^2, m_2^2 this function decreases as $|s|^{-1}$ when $|s|$ becomes large, in the special case when $a_1 m_1^2 + a_2 m_2^2 = 0$ the function vanishes as fast as $|s|^{-2}$.

More generally, for a VMD model (4.3) with N resonance terms we can write

$$F(s) = \sum_{r=1}^{N} a_r \frac{m_r^2}{m_r^2 - s} = \frac{A_{N-1} s^{N-1} + \cdots + A_1 s + A_0}{\prod_{r=1}^{N} \left(m_r^2 - s\right)}, \tag{4.32}$$

where the coefficients A_{N-1}, ..., A_0 are somewhat complicated expressions that depend on the coefficients a_1, ..., a_N and the masses m_1^2, ..., m_N^2. We get the asymptotic behavior $|F(s)| \propto |s|^{-K}$, for K an integer $1 \leq K \leq N$, if the $K-1$ coefficients A_{N-1}, A_{N-2}, ..., A_{N+1-K} all vanish. This yields $K-1$ equations for N coefficients a_1, ..., a_N. Note that these equations come in addition to the normalization equation (4.30). In particular, this means that if we require a model that contains N resonance terms to have the asymptotic dependence $|s|^{-N}$, then the resulting $N-1$ equations, together with the single normalization condition, do in principle determine the coefficients a_1, ...,a_N uniquely, provided that the masses are known.

Although the procedure described above is straightforward, the expressions for A_1, ..., A_{N-1} are quite complicated and their manipulation can be cumbersome. Luckily, we can easily derive an equivalent but simpler set of algebraic constraints [27]. Suppose that $|F(s)| \propto |s|^{-K}$ for large $|s|$, and consider the integration contour depicted in Fig. 4.1, which we used to derive the dispersion relation for the form factor. We can integrate over this contour any of the functions $F(s)$, $s\,F(s)$, ..., $s^{K-2}F(s)$, and because all of those functions vanish at least as fast as $|s|^{-2}$ for large $|s|$, the part of the integral over the large circle does not contribute (in the limit of the infinitely large circle). From the assumption that the form factor F is analytic everywhere on the physical sheet, except for the positive real axis, we then obtain the conditions

$$\int_0^\infty ds\ \Im\{F(s)\} = 0, \tag{4.33}$$

$$\int_0^\infty ds\ s\,\Im\{F(s)\} = 0, \tag{4.34}$$

$$\ldots \tag{4.35}$$

$$\int_0^\infty ds\ s^{K-2}\,\Im\{F(s)\} = 0. \tag{4.36}$$

Using now the expression (4.2) for the imaginary part of the VMD model (4.3), we obtain the equations

$$\sum_{r=1}^{N} a_r m_r^2 = 0, \tag{4.37}$$

$$\sum_{r=1}^{N} a_r \left(m_r^2\right)^2 = 0, \tag{4.38}$$

$$... \tag{4.39}$$

$$\sum_{r=1}^{N} a_r \left(m_r^2\right)^{K-1} = 0. \tag{4.40}$$

Under usual conditions, these equations form a set of $K-1$ independent equations. They were derived from the assumption that the VMD model form factor follows the behavior $|F(s)| \propto |s|^{-K}$ for large $|s|$. In other words, this set of $K-1$ equations represents a necessary condition for the VMD model to follow that asymptotic behavior. However, as we have argued above, the VMD model follows this asymptotic behavior if and only if the $K-1$ coefficients A_{N-1}, A_{N-2}, ..., A_{N+1-K} are all zero. But since this is also just a set of $K-1$ independent equations, it is clear that the two sets must be equivalent.[9]

The approach presented above may be quite natural, especially if one starts from the assumption that the form factor should be determined mainly by the character of its resonance poles. On the other hand, it also turns out to be rather cumbersome. For this reason I will suggest here also another approach, based on a modification of the formula (4.27) for the function $\hat{F}^{(r)}_{\mathrm{U\&A}}(W)$, which is used in the definition of the U&A model. As far as I know this approach has not been suggested before, but it also does not constitute a central point of this thesis. Beyond the short discussion provided here I will not return to it in any other place of this work. As we will see, the advantage of this approach over the established one presented above lies in its great simplicity. However, I do not propose here any physical argument in its favor.

From the equation (4.16)

$$s = s_0 - \frac{4(s_{\mathrm{in}} - s_0)}{[W - 1/W]^2}$$

for the inverse map to $W(q(s))$, we can easily see that there are only two points on the W-plane that are mapped to the infinity; these are $W = 1$ and $W = -1$. Among them, only $W = -1$ can be approached from the physical

[9] I assume here that in both of the sets all $K-1$ equations are independent. This condition is normally satisfied.

region. Furthermore, because

$$\left[W - \frac{1}{W}\right]^2 = \frac{[W^2-1]^2}{W^2} = [W+1]^2 \frac{[W-1]^2}{W^2}, \tag{4.41}$$

we see from (4.16) that as $W \to -1$

$$|s|^{-1} \propto |1+W|^2, \quad (\text{for } W \to -1). \tag{4.42}$$

If we now inspect the formula (4.27) for a single resonance term of the U&A model,

$$\hat{F}^{(r)}_{\text{U\&A}}(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^2 \prod_{p\in P} \frac{W_N - p}{W - p},$$

we see that as W approaches the value -1, both the finite-energy factor $\prod_{p\in P} \frac{W_N - p}{W - p}$ and the part $\left([1-W]/[1-W_N^2]\right)^2$ of the asymptotic factor both approach a finite non-zero value. The only factor that determines the asymptotic behavior in the physical region is therefore the remaining part $(1+W)^2$ of the asymptotic factor. Comparing this with (4.42) we see indeed that as $|s| \to \infty$

$$\left|\hat{F}^{(r)}_{\text{U\&A}}(W(q(s)))\right| \propto |s|^{-1}.$$

We also see that if we change the power of the asymptotic factor from 2 to $2K$,[10]

$$\left(\frac{1-W^2}{1-W_N^2}\right)^2 \to \left(\frac{1-W^2}{1-W_N^2}\right)^{2K},$$

then the asymptotic behavior of $\hat{F}^{(r)}_{\text{U\&A}}$ becomes $|s|^{-K}$. Hence, recalling that the condition (4.31) says that for a form factor of a hadron that contains n_q valence quarks we have $K = n_q - 1$, we can consider the modification

$$\hat{F}^{(r)}_{\text{U\&A,modified}}(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^{2n_q-2} \prod_{p\in P} \frac{W_N - p}{W - p}, \tag{4.43}$$

of the standard U&A resonance term (4.27). This breaks down the correspondence between the VMD model and the U&A model, but we achieve the desired asymptotic behavior without imposing any further restrictions on the coefficients a_1, ..., a_N. Note also that this modification does not violate any of the properties that we wanted the form factor to satisfy. The resulting form factor is still analytic on the physical sheet, has two branch points, is real below the lowest threshold, and the normalization at $s = 0$ remains unchanged.

[10] There are two main reasons why it seems reasonable to change the power of the whole asymptotic factor, rather than only of $(1+W)^2$. First, it keeps the normalization unchanged. Second, it changes the asymptotic behavior on all the four sheets in the same way, rather than changing it to $|s|^{-K}$ on the first and the third sheet, and keeping it at $|s|^{-1}$ on the second and the fourth sheet.

Part III

Applications

In Part III of this thesis I will present some of the applications of the U&A model that I have worked on during my doctoral study. Compared to the first two parts of the thesis, the third part is slightly more concise, because many of the details can be readily found in the published papers.

We begin by discussing the running of the fine structure constant in the chapter 5. After a general introduction of that concept in the section 5.1, we will discuss its significance for the evaluation of the hadronic contribution to the magnetic anomaly of the muon in the section 5.2. Here we will also mention the possibility of evaluating this quantity from space-like data, and briefly introduce the article [29], on which I collaborated during my doctoral study, and which considered that method for the evaluation of the anomaly. The discussion of the section 5.2 will also help us to appreciate some of the recent experimental endeavors to measure the imaginary and real parts of the running fine structure constant, which will be mentioned in the section 5.3. Among the original goals of the thesis was a review of some theoretical aspects of the running of the fine structure constant, namely of the relation between its imaginary part and the total cross section of the annihilation of the electron–positron pair in the time-like region, and the argument for the reality of the running fine structure constant in the space-like region. We will therefore continue by discussing these two topics in sections 5.4 and 5.5. And lastly, in the section 5.6 I will mention some other related results that I have worked on but which have not yet been published.

The last chapter of the thesis then consists of two much shorter sections. The section 6.1 discusses the topic of the damped oscillatory structures that were found in the data for the effective form factor of the proton by A. Bianconi and E. Tomasi-Gustafsson [30]. I collaborated on two papers [31, 32] that studied this topic, and which will also be introduced in that section. The section 6.2 then discusses another topic that I collaborated on during my doctoral work: the value of the ratio $R = \phi \to K^+K^-/K_L^0K_S^0$ of the decay rates of the ϕ-resonance into the charged and neutral kaons. In that section I will introduce the paper [33].

Chapter 5

The running of the fine structure constant

5.1 The running of the fine structure constant

5.1.1 Running coupling constants

One of the major puzzles of relativistic quantum field theory is the appearance of ultraviolet divergences in higher orders of perturbation theory in physically relevant theories such as the quantum electrodynamics or the standard model in four space-time dimensions. Those divergences arise from certain regions of integration over the loop momenta, and can always be associated with particular one-particle-irreducible sub-graphs. The study of this problem and associated methods is usually called the renormalization theory.[1] The standard way[2] of dealing with the ultraviolet divergences is to decompose the Lagrangian density of the theory into three parts

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\text{int}} + \mathcal{L}_{\text{ct}}, \tag{5.1}$$

instead of decomposing it into the free and interacting parts only, as would be suggested by the usual perturbation theory. The part $\mathcal{L}_{\text{ct}}$ is called the counterterm Lagrangian density. The counterterms present in $\mathcal{L}_{\text{ct}}$ are meant to cancel ultraviolet divergences that appear in the perturbation theory.

[1] The renormalization theory is addressed — although to a varying degree of clarity and completeness — in most textbooks of the quantum field theory. A modern and readable treatment entirely dedicated to this topic is presented in [34]. Some aspects of this topic are also nicely introduced in [9, 35], which also present the modern "effective field theory" perspective on these issues, including the interpretation of the standard model as a low-energy approximation to a more fundamental theory.

[2] There is also another very common approach that was pioneered by K. Wilson, and which is based in a fundamental way on the presence of some kind of an ultraviolet cutoff. In this approach the coupling constants of the Lagrangian are cutoff dependent, with their dependence assumed to be such as to make physically observable quantities independent of the cutoff.

A detailed formulation of this procedure is given by the Zimmermann's forest formula. Of course, in order for all of this to work, there must be a counterterm of an appropriate structure present in the Lagrangian for each one-particle-irreducible subgraph that is ultraviolet-divergent. In general, it is not trivial to show that that is the case — mainly because the structure of possible counterterms is directly restricted by assumed symmetries of the Lagrangian in a way in which the structure of ultraviolet divergences might not be.[3] Nonetheless, we will not discuss these issues here, and for a more proper explanation of this and related topics the reader is recommended to consult [34].

What is important for our purposes is that the decomposition of the Lagrangian density $\mathcal{L}$ into two parts, one consisting of its free and interaction parts $\mathcal{L}_0 + \mathcal{L}_{\text{int}}$ together and the other of its counterterms $\mathcal{L}_{\text{ct}}$, is not unique. The only requirement that we need to satisfy is that the counterterms have infinite parts[4] that cancel the infinities from ultraviolet divergent subgraphs. We are, however, quite free to move finite quantities between $\mathcal{L}_0 + \mathcal{L}_{\text{int}}$ and $\mathcal{L}_{\text{ct}}$ as we wish. As a consequence, there is a freedom in what values of renormalized couplings, masses or fields we use — and we can change those without affecting any observable quantities, provided that we also make appropriate adjustments of counterterms in order to keep the overall Lagrangian intact. For example, suppose that the Lagrangian density $\mathcal{L}$ contains the term $-g_0\Phi^4/4!$ where Φ is a real scalar field, and that we divided this term between the interaction $\mathcal{L}_{\text{int}}$ and the counterterms $\mathcal{L}_{\text{ct}}$ by assigning to the interaction the quantity $-g\Phi^4/4!$ and keeping what remains as a counterterm. Hence, the counterterm is $-g_{\text{ct}}\Phi^4/4!$, where

$$g_{\text{ct}} = g_0 - g.$$

(Here g is the renormalized coupling. We also assume for simplicity that Φ represents the already renormalized field.) Now, if δg is some finite real number, we can move the quantity $-\delta g\Phi^4/4!$ from $\mathcal{L}_{\text{int}}$ to $\mathcal{L}_{\text{ct}}$. This amounts to the re-definitions

$$\begin{aligned} g &\to g' = g - \delta g, \\ g_{\text{ct}} &\to g'_{\text{ct}} = g_{\text{ct}} + \delta g. \end{aligned}$$

Thus, although we now have a new value g' of the coupling, observable quantities remain unchanged because the overall Lagrangian has not been

[3] The problems appear if infinitesimal versions of the symmetry transformations are not linear in fields. (If they all *are* linear, then Slavnov–Taylor identities imply that ultraviolet divergences obey the symmetries of the Lagrangian.) An important case when they are not linear is that of the BRST symmetry. It is therefore not simple to show that gauge theories can be renormalized.

[4] The term "infinite part" must first be given a proper meaning by the use of some form of regularization, such as for instance by using the dimensional regularization or by formulating the theory on a lattice. At the final step, after the renormalization has been appropriately applied, the regularizator should be removed.

affected. Different methods of assigning finite parts between $\mathcal{L}_0 + \mathcal{L}_{\rm int}$ and $\mathcal{L}_{\rm ct}$ correspond to different *renormalization prescriptions*.

Now, any kind of renormalization prescription corresponds to some energy scale. Sometimes, that can be quite obvious, such as when we define the renormalized couplings based on values of relevant Green's functions calculated at that energy. Other times, it can be more subtle, as in the case of the physical renormalization prescription, when the scale is related to the physical masses of the particles in the theory. In perturbation theory this has important consequences, because if we use a renormalization prescription that corresponds to some energy scale μ and calculate a quantity whose characteristic energy scale is E, then individual terms in the perturbation series will each contain one or more factors of $\ln(E/\mu)$. When $E >> \mu$ (or $E << \mu$), those logarithms become very large, slowing down the convergence of the perturbation series, or even invalidating it altogether. For this reason, when calculating a quantity at an energy scale E it is important to use couplings defined at some renormalization scale μ similar to the energy E. That will tame the logarithms, and if the coupling g_μ at that energy scale is small enough, we are free to use the perturbation expansion. (If $|g_\mu| > 1$ then there is no known way of carrying out perturbative calculations at that energy scale.) One way of looking at this is that as we change the renormalization prescription some contributions move between different terms in the perturbation series. Ideally, we want to use a prescription for which the largest possible part of the overall value of the quantity is contained in the first few terms of the series. Another perspective that can be helpful when thinking about the renormalization energy scale is to notice that when we choose a renormalization scale μ we effectively impose an approximate cutoff on transverse momenta larger than μ. Or, expressed from yet another perspective, by choosing the renormalization scale μ we are effectively integrating out the degrees of freedom that are accessible only at higher energies.

This is one of the reasons why effective couplings that correspond to renormalization at some given energy scale μ play such an important role in high energy physics. Suppose now, for simplicity, that there is only a single coupling g_0 present in the original Lagrangian, and denote its effective value at the renormalization scale μ by g_μ. If μ is inside the region where perturbation theory works — that is, if $|g_\mu| < 1$ — then from the knowledge of g_μ at the scale μ we should be able to perturbatively calculate the effective coupling $g_{\mu'}$ at some different energy scale μ', provided that μ' is not too distant from μ (so that the logarithm $\ln(\mu'/\mu)$ remains harmless). This implies that there should exist a function F of μ', defined in some neighborhood of μ, that depends parametrically on μ, g_μ and the various masses m present in the theory, and which yields the value of $g_{\mu'}$,

$$g_{\mu'} = F(\mu'; \mu, g_\mu, m). \tag{5.2}$$

The exact form of the function F depends both on the Lagrangian and the precise definition of how does the renormalization scale μ determine g_μ.

We are now going to sketch an important argument in which a central role is played by dimensional analysis. For that reason it will be very convenient to work with dimensionless couplings.[5] That is, we assume that in the units of energy the coupling g_μ has dimension zero. This simply means that if the coupling g_0 appearing in the original Lagrangian has dimension d, then the corresponding effective coupling is defined as $g_\mu \mu^d$. For example, if g_0 appeared in the Lagrangian density in the term $g_0 \Phi^3/3!$ (in which case we would normally have $d = 1$), then the corresponding term in $\mathcal{L}_{\text{int}}$ would be $g_\mu \mu^d \, \Phi^3/3!$. Now, if both g_μ and $g_{\mu'}$ are dimensionless, dimensional analysis implies that instead of (5.2) we can also write

$$g_{\mu'} = G\left(g_\mu, \frac{\mu'}{\mu}, \frac{m}{\mu}\right), \tag{5.3}$$

where G is a dimensionless function of dimensionless variables. Differentiating with respect to μ' yields

$$\frac{d}{d\mu'} g_{\mu'} = \frac{1}{\mu} \frac{\partial}{\partial y} G\left(g_\mu, y, \frac{m}{\mu}\right)\Big|_{y=\frac{\mu'}{\mu}}. \tag{5.4}$$

A deeper analysis would reveal that there should be no zero-mass singularities present in G. This means that for $\mu >> m$ we can use the approximation

$$\frac{d}{d\mu'} g_{\mu'} = \frac{1}{\mu} \frac{\partial}{\partial y} G\left(g_\mu, y, 0\right)\Big|_{y=\frac{\mu'}{\mu}}, \quad \text{for } \mu >> m. \tag{5.5}$$

Now we can set $\mu' = \mu$ to obtain a differential equation for g_μ,

$$\mu \frac{d}{d\mu} g_\mu = \frac{\partial}{\partial y} G\left(g_\mu, y, 0\right)\Big|_{y=1}, \quad \text{for } \mu >> m. \tag{5.6}$$

If we also use the conventional definition of the beta function

$$\beta(g_\mu) = \frac{\partial}{\partial y} G\left(g_\mu, y, 0\right)\Big|_{y=1}, \tag{5.7}$$

we can re-write that equation in the following form

$$\mu \frac{d}{d\mu} g_\mu = \beta\left(g_\mu\right), \quad \text{for } \mu >> m. \tag{5.8}$$

[5] Note that in some of our earlier statements, such as when we compared the size of the coupling with 1, we have been already implicitly assuming that the coupling was dimensionless.

This is a simple example of the *Callan-Symanzik equation*, or the renormalization group equation. Beta functions can often be calculated in perturbation theory, but the equation (5.8) leads to valuable insights even if that is not the case.

We often refer to effective, renormalization-scale dependent coupling constants as *running coupling constants*.

5.1.2 Renormalization of the electric charge

In this section we are interested in one particular case of a running constant, the running fine structure constant. Recall how in the chapter 3 we found that the operator of the electric charge Q commutes with generators of the Poincaré group. This allowed us to show that the result of acting by the operator Q on a single-particle asymptotic state $\Psi_{\mathbf{p},\sigma,n}$ is given by the equation (3.9),

$$Q\Psi_{\mathbf{p},\sigma,n} = q_n^{(\mathrm{phys})}\Psi_{\mathbf{p},\sigma,n},$$

where $q_n^{(\mathrm{phys})}$ is the physical electric charge of the particle n. Let us now use the index l to label all elementary fields in the Lagrangian as $\Psi^l(x)$. The electromagnetic current $J^\mu(x)$ is associated with a particular global symmetry, to which corresponds the infinitesimal transformation of the form

$$\Psi^l(x) \to \Psi^l(x) + i\epsilon\, q_l \Psi^l(x). \tag{5.9}$$

We do not sum over the index l on the right-hand side. In the formula (5.9) the symbol ϵ represents a real infinitesimal constant and q_l is the electric charge associated with the field $\Psi^l(x)$. Since this is an internal symmetry, which leaves the Lagrangian density $\mathcal{L}$ intact, we can use the transformation law (5.9) to construct an explicit formula for the electromagnetic current density $J^\mu(x)$

$$J^\mu(x) = -i\sum_l \frac{\partial\mathcal{L}(x)}{\partial\left(\partial_\mu\Psi^l(x)\right)} q_l\Psi^l(x). \tag{5.10}$$

The time component of this formula reads

$$J^0(x) = -i\sum_n P_n q_n \Psi^n(x), \tag{5.11}$$

where $P_n(x)$ is the canonical conjugate of the canonical coordinate $\Psi^n(x)$. The index n is summed only over those values of l for which $\Psi^l(x)$ is indeed a canonical coordinate.[6] We can now use the canonical commutation or

[6] Ignoring some subtleties the statement that $\Psi^n(x)$ is a canonical coordinate does simply mean that its time derivative $\partial_0\Psi^l$ appears in the Lagrangian. Then $\frac{\partial\mathcal{L}}{\partial\left(\partial_0\Psi^l\right)}$ corresponds to the conjugate momentum P_l. If $\partial_0\Psi^l$ does not appear in the Lagrangian then $\frac{\partial\mathcal{L}}{\partial\left(\partial_0\Psi^l\right)} = 0$ and the field Ψ^l drops out from the summation over l in the zeroth component of (5.10). That leads to the sum over n in (5.11).

anti-commutation relations to show that for a canonical coordinate Ψ^n,

$$[J^0(\mathbf{y},t),\Psi^n(\mathbf{x},t)]_- = -q_n\Psi^n(\mathbf{y},t)\delta^3(\mathbf{y}-\mathbf{x}). \tag{5.12}$$

If we integrate $J^0(\mathbf{y},t)$ over the whole space of $\mathbf{y}$, and use the fact that $dQ/dt = 0$, we obtain the commutation relation

$$[Q,\Psi^n(x)]_- = -q_n\Psi^n(x). \tag{5.13}$$

This can be interpreted as saying that when we apply the operator Ψ^n to some state the value of its overall charge Q decreases by q_n.

Consider now the matrix element of the commutation relation (5.13) between the vacuum state Ψ_0 and a single-particle state $\Psi_{\mathbf{p},\sigma,n}$. We can write

$$\begin{aligned} -q_n\left(\Psi_0,\Psi^n(x)\Psi_{\mathbf{p},\sigma,n}\right) &= \left(\Psi_0,[Q,\Psi^n(x)]_-\Psi_{\mathbf{p},\sigma,n}\right) \\ &= \left(0-q_n^{(\mathrm{phys})}\right)\left(\Psi_0,\Psi^n(x)\Psi_{\mathbf{p},\sigma,n}\right), \end{aligned} \tag{5.14}$$

where in the last line we used the properties $Q\Psi_0 = 0$ and $Q\Psi_{\mathbf{p},\sigma,n} = q_n^{(\mathrm{phys})}\Psi_{\mathbf{p},\sigma,n}$, which were both derived near (3.9). If the field $\Psi^n(x)$ annihilates the particle n, then $(\Psi_0,\Psi^n(x)\Psi_{\mathbf{p},\sigma,n})$ is not zero and we obtain the condition

$$q_n = q_n^{(\mathrm{phys})}. \tag{5.15}$$

At a first sight this might appear to be saying that the renormalized physical charge $q_n^{(\mathrm{phys})}$ is the same as the charge q_n of the corresponding field in the Lagrangian, but we need to be more careful when interpreting this result. The problem is that the global symmetry (5.9) does not define the overall scale of electric charges q_l, but only their relative values for different fields: if we multiplied each q_l by the same non-zero constant the transformation (5.9) would still be a symmetry of the Lagrangian. This has a counterpart in the fact that as far as only the global symmetry transformation (5.9) is concerned, the overall scale of the electromagnetic current (5.10) is not determined. To obtain the equation (5.15) we first had to fix some, from this point of view a completely arbitrary scale of the current J^μ, and the equation therefore only says that the relative physical charges of stable particles are the same as the relative charges — determined by the transformation (5.9) — of their corresponding fields. In fact, a completely analogous statement holds also for composite particles.[7] Nonetheless, if the charges q_l in the

[7] Suppose that the particle $\Psi_{\mathbf{p},\sigma,n}$ does not correspond to any elementary field in the Lagrangian, but is instead annihilated by a product Φ of a number of elementary fields and/or their adjoints. We can repeat the argument given above for an elementary particle with only a few minor modifications. The commutation relation (5.13) would now instead read

$$[Q,\Phi] = -q\Phi,$$

where q is the sum of q_ls for the elementary fields present in the field operator Φ minus the sum of q_ls for the elementary field adjoints in the product. The condition (5.15) then simply becomes $q_n = q$.

transformation law (5.9) stand for the bare, unrenormalized charges, then the transformation should definitely be a symmetry of the Lagrangian. Our result (5.15) can therefore be interpreted as stating that the charges of all particles are renormalized by the same factor relative to their bare values, independently of all other details regarding the particles. For example, the charge of the electron, which corresponds to an elementary field and interacts only electromagnetically and weakly, is renormalized by the same factor as the charge of the proton, which is a composite particle that interacts also with the strong force.

In order to say anything about the overall scale of the charges q_l we must actually introduce the electromagnetic interaction. We must therefore consider the local version of the symmetry transformation (5.9), which is

$$\Psi^l(x) \to \Psi^l(x) + i\epsilon(x)\, q_l \Psi^l(x). \tag{5.16}$$

This is a gauge transformation because $\epsilon(x)$ is no longer a constant but depends on the space-time coordinate x. As is well-known, in order to make the action invariant under such local transformations we need to introduce into the Lagrangian the photon gauge field A^μ, which transforms under the gauge transformation as

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \epsilon(x), \tag{5.17}$$

and replace all derivatives $\partial_\mu \Psi^l$ of the matter fields with their gauge-covariant counterparts

$$D_\mu \Psi^l(x) = \left[\partial_\mu - i q_l A_\mu(x)\right] \Psi^l(x). \tag{5.18}$$

Here, the symbols q_l and A_μ are not meant to indicate any particular normalization yet.

Now, as far as only the matter fields are concerned, we still cannot say anything about the overall scale of the electric charge. This is because their transformations (5.16) under the local symmetry depend on $\epsilon(x)$ only in the combination $q_l\epsilon(x)$. This means that any redefinition of the overall scale of electric charges can be absorbed into the corresponding redefinition of the overall scale of $\epsilon(x)$. The gauge field, however, transforms under the local symmetry as (5.17) — in a way that depends only on the parameter $\epsilon(x)$, and not on the value of the electric charge. Hence, the scale of the gauge field does in fact fix the overall scale of ϵ and consequently also the scale of the electric charge.

To see how this works, suppose now that the transformation law (5.17) holds for the bare field $A^{(0)}_\mu$,

$$A^{(0)}_\mu(x) \to A^{(0)}_\mu(x) + \partial_\mu \epsilon(x),$$

and that the derivative D_μ of (5.18) is gauge covariant when the gauge field that appears in the derivative is bare and the charges q_l there are the bare

electric charges, which we denote by $q_l^{(0)}$,

$$D_\mu \Psi^l(x) = \left[\partial_\mu - i q_l^{(0)} A_\mu^{(0)}(x)\right] \Psi^l(x). \tag{5.19}$$

If we write the renormalized photon field $A_\mu(x)$ in terms of the bare field $A_\mu^{(0)}$ as

$$A_\mu(x) = Z_3^{-\frac{1}{2}} A_\mu^{(0)}(x), \tag{5.20}$$

then its gauge transformation is $A_\mu \to A_\mu + Z_3^{-1/2}\partial_\mu \epsilon$. This shows that to construct a gauge-covariant derivative in terms of the renormalized gauge field we have no other choice than to substitute $A_\mu^{(0)}(x) = \sqrt{Z_3} A_\mu(x)$ into the expression (5.19) for D_μ,

$$D_\mu \Psi^l(x) = \left(\partial_\mu - i q_l^{(0)} \sqrt{Z_3} A_\mu(x)\right) \Psi^l(x).$$

However, the coefficient of "$-i$ times the gauge field" in the gauge-covariant derivative corresponds to the charge, which implies that the renormalized charge q_l is equal to the quantity $q_l^{(0)}\sqrt{Z_3}$,

$$q_l = \sqrt{Z_3}\, q_l^{(0)}. \tag{5.21}$$

This shows that the electric charge is renormalized only by those effects that renormalize the photon gauge field, and does not depend on any other radiative correction.[8] In other words, the running of the effective electric charge depends only on the polarization of vacuum, or the full photon propagator. For this reason, we are now going to briefly review some properties of that propagator.

5.1.3 The photon propagator and the running fine structure constant

Now that we have seen that the renormalization of the electric charge is fully determined by radiative corrections to the photon propagator we need to review some of its properties. We will denote the free propagator of the photon by $i\Delta_{\mu\nu}(x,y)$

$$i\Delta_{\mu\nu}(x,y) = \left(\Phi_0, \mathrm{T}\left\{a_\mu(x) a_\nu(y)\right\}\Phi_0\right), \tag{5.22}$$

where Φ_0 is the free-particle vacuum and $a_\mu(x)$ is the gauge field expressed in the interaction picture. The momentum-space free propagator $i\Delta_{\mu\nu}(q)$ is

[8] This can also be understood as a consequence of the Ward–Takahashi identities, which imply that all the other radiative corrections to the electromagnetic vertex cancel out in the limit when the external fermion lines are on the mass-shell and the four-momentum transfer is zero. These identities follow from the current-conservation condition $\partial_\mu J^\mu = 0$, which means that this and the approach presented above are not really independent.

Figure 5.1: This figures illustrates the decomposition of the full photon propagator into an infinite series of terms containing one-particle-irreducible insertions connected by free propagators. The symbol on the left-hand side represents the full propagator. In terms of Feynman diagrams it corresponds to the sum of all diagrams that have two external photon lines, and no other external lines. (This sum also includes the free photon propagator.) On the right-hand side we have an infinite series. The circle containing the label "1PI" represents the sum of all one-particle-irreducible diagrams that have two external photon lines and no other external lines, but with the external photon propagators amputated. (Note that this "1PI" sum does not contain the free propagator.) Hence, the n-th term in the series contains $n-1$ such sums of one-particle-irreducible subdiagrams, all of them connected by free propagators and with an extra propagator at each end of the diagram.

given by the definition

$$\int d^4x\, d^4y\, \mathrm{e}^{ipx}\mathrm{e}^{-iqy}\, (\Phi_0, \mathrm{T}\,\{a_\mu(x)a_\nu(y)\}\,\Phi_0) \\ = (2\pi)^4\delta^4\,(p-q)\; i\Delta_{\mu\nu}(q). \quad (5.23)$$

Here, q is the four-momentum entering the amplitude through the vertex at y and p the four-momentum exiting at the vertex at x. Combining the last two equations, we can write

$$i\Delta_{\mu\nu}(x,y) = i\int \frac{d^4q}{(2\pi)^4}\Delta_{\mu\nu}(q)\mathrm{e}^{-iq(x-y)}. \quad (5.24)$$

We will use the apostrophe to denote the full propagator variants of $\Delta_{\mu\nu}(x,y)$ and $\Delta_{\mu\nu}(q)$, that is, we will write

$$i\Delta'_{\mu\nu}(x,y) = (\Psi_0, \mathrm{T}\,\{A_\mu(x)A_\nu(y)\}\,\Psi_0)\,, \quad (5.25)$$

$$i\Delta'_{\mu\nu}(x,y) = i\int \frac{d^4q}{(2\pi)^4}\Delta'_{\mu\nu}(q)\mathrm{e}^{-iq(x-y)}, \quad (5.26)$$

where Ψ_0 is the true vacuum and $A_\mu(x)$ is the gauge field expressed in the Heisenberg picture.

When we study the full propagator from the point of view of perturbation theory it is very useful to express it in terms of appropriate one-particle-irreducible diagrams.[9] Recall that we say that a connected diagram

[9]There are several reasons for this, one of them being that ultraviolet divergences and thus also counterterms correspond directly to appropriate one-particle-irreducible diagrams.

is one-particle irreducible if it is not possible to divide it into two disconnected parts by cutting through any single of its internal lines. In the context of the photon propagator the relevant one-particle-irreducible diagrams are those that contain exactly two external lines, both of them corresponding to the photon. Let us denote by $-i(2\pi)^4\Pi_{\mu\nu}(q)$ the sum of all such one-particle-irreducible diagrams, expressed in the momentum space, but with both external photon lines amputated. Here, q is the four-momentum flowing through the external lines. We have extracted the factor of $-i(2\pi)^4$ for later convenience. Note, that $(2\pi)^4$ would appear together with the overall four-momentum conservation delta-function that is associated with any connected Feynman diagram, but which we suppose has already been extracted and does not figure in the expression $-i(2\pi)^4\Pi_{\mu\nu}(q)$. Then, from the usual rules for Feynman diagrams, we see that $\left[i/(2\pi)^4\right]\Delta'(s)$ corresponds to the sum of diagrams that can be written in a familiar way as the free propagator, plus the term that contains the one-particle-irreducible factor $-i(2\pi)^4\Pi_{\mu\nu}(q)$ with two free propagators connected at both ends, plus the term containing two one-particle-irreducible factors connected by the free propagator and with two more propagators each connected to one end of the diagram, etc. (See Fig. 5.1.) This corresponds to the equation

$$\begin{aligned}\frac{i}{(2\pi)^4}\Delta'_{\mu\nu}(q) = \frac{i}{(2\pi)^4}\Delta_{\mu\nu}(q) \\ + \left[\frac{i}{(2\pi)^4}\Delta_{\mu\alpha_1}(q)\right]\left[-i(2\pi)^4\Pi^{\alpha_1\alpha_2}(q)\right]\left[\frac{i}{(2\pi)^4}\Delta_{\alpha_2\nu}(q)\right] \\ + \left[\frac{i}{(2\pi)^4}\Delta_{\mu\alpha_1}(q)\right]\left[-i(2\pi)^4\Pi^{\alpha_1\alpha_2}(q)\right]\left[\frac{i}{(2\pi)^4}\Delta_{\alpha_2\alpha_3}(q)\right] \\ \times\left[-i(2\pi)^4\Pi^{\alpha_3\alpha_4}(q)\right]\left[\frac{i}{(2\pi)^4}\Delta_{\alpha_4\nu}(q)\right] + \ldots . \quad (5.27)\end{aligned}$$

We can slightly simplify this series by canceling the "$i/(2\pi)^4$"-terms associated with propagators with the "$-i(2\pi)^4$"-terms that stand in front of one-particle-irreducible insertions. We obtain

$$\begin{aligned}\Delta'_{\mu\nu}(q) = \Delta_{\mu\nu}(q) + \Delta_{\mu\alpha_1}(q)\,\Pi^{\alpha_1\alpha_2}(q)\,\Delta_{\alpha_2\nu}(q) \\ + \Delta_{\mu\alpha_1}(q)\,\Pi^{\alpha_1\alpha_2}(q)\,\Delta_{\alpha_2\alpha_3}(q)\,\Pi^{\alpha_3\alpha_4}(q)\,\Delta_{\alpha_4\nu}(q) + \ldots . \quad (5.28)\end{aligned}$$

For our purposes, it will be more convenient to rewrite this series as an implicit equation for the full propagator,

$$\Delta'_{\mu\nu}(q) = \Delta_{\mu\nu}(q) + \Delta_{\mu\alpha_1}(q)\,\Pi^{\alpha_1\alpha_2}(q)\,\Delta'_{\alpha_2\nu}(q). \quad (5.29)$$

This equation might become more easy to understand if we temporarily stop displaying the indices of Δ', Δ, and Π, and treat those objects as matrices. For now, we also cease displaying the four-momentum argument q. The

equation (5.29) then simply reads $\Delta' = \Delta + \Delta\Pi\Delta'$. We can also write it in the form

$$\Delta' - \Delta\Pi\Delta' = \Delta. \tag{5.30}$$

Assuming that $\mathbf{1} - \Delta\Pi$ is invertible we can now obtain an explicit formula for Δ',

$$\Delta' = [\mathbf{1} - \Delta\Pi]^{-1}\,\Delta. \tag{5.31}$$

We also assume that Δ and Δ' themselves are invertible, which allows us to re-write this result in a form that is even more convenient for explicit calculations,

$$\Delta'^{-1} = \Delta^{-1}\,[\mathbf{1} - \Delta\Pi] = \Delta^{-1} - \Pi. \tag{5.32}$$

To move forward, we will need to employ a property that is sometimes called the gauge-invariance of the S-matrix and which states that for any matrix element $M_{\beta\alpha}^{\mu_1\mu_2\cdots}(q_1, q_2, \dots)$ defined by

$$\begin{aligned}&(2\pi)^4\delta^4\left(p_\beta + q_1 + q_2 + \cdots - p_\alpha\right) M_{\beta\alpha}^{\mu_1\mu_2\cdots}(q_1, q_2, \dots)\\ &= \int d^4x_1 d^4x_2 \dots \mathrm{e}^{iq_1x_1}\mathrm{e}^{iq_2x_2}\dots\left(\Psi_\beta^{\mathrm{out}}, \mathrm{T}\left\{J_1^\mu(x_1)J_2^\mu(x_2)\dots\right\}\Psi_\alpha^{\mathrm{in}}\right)\end{aligned} \tag{5.33}$$

we have

$$q_{1\mu_1} M_{\beta\alpha}^{\mu_1\mu_2\cdots}(q_1, q_2, \dots) = q_{2\mu_2} M_{\beta\alpha}^{\mu_1\mu_2\cdots}(q_1, q_2, \dots) = \cdots = 0. \tag{5.34}$$

This follows from the current-conservation condition $\partial_\mu J^\mu = 0$ and from the fact that the electromagnetic current is itself an electrically neutral operator.[10] Note that this property holds also for more general amplitudes: the only requirement is that all external charged particles must be on the mass-shell, but neutral external particles are all allowed to be off the mass-shell.

Here we are interested only in the simplest case of this formula, when α and β both stand for the vacuum and there are only two current operators in the product. We can derive the relation between the corresponding matrix element $M_{00}^{\mu\nu}$ and the full propagator $i\Delta'_{\mu\nu}(q)$ if we consider both quantities in perturbation theory. Since the gauge field is coupled only to the electromagnetic current, we have

$$\begin{aligned}i\Delta'_{\mu\nu}(q) = i\Delta_{\mu\nu}(q) &- \left[i\Delta_{\mu\alpha_1}(q)\right] M_{00}^{\alpha_1\alpha_2}(q, -q)\left[i\Delta_{\alpha_2\nu}(q)\right]\\ &+ \left[i\Delta_{\mu\alpha_1}(q)\right]\left(-i(Z_3 - 1)\right)\left[q^2\eta^{\alpha_1\alpha_2} - q^{\alpha_1}q^{\alpha_2}\right]\left[i\Delta_{\alpha_2\nu}(q)\right],\end{aligned} \tag{5.35}$$

where the minus sign in front of the second term has its origin in the two factors of i, each of which is associated with one of the two electromagnetic vertices that connect to the external photon propagators. The last term corresponds to the single insertion of the gauge field renormalization

[10] For a derivation see, e.g., the chapter 10 of [9].

counterterm, which needs now to be displayed explicitly.[11] Comparing this equation with (5.29) yields (in the matrix notation)

$$\Pi\Delta' = -iM_{00}\Delta + (Z_3 - 1)\left[q^2\mathbf{1} - qq^{\mathrm{T}}\right]\Delta. \tag{5.36}$$

Multiplying this with q from the left (or, if we return the indices, contracting with q_{α_1}) and using the gauge invariance condition (5.34) (together with the fact that $q_{\alpha_1}\left[q^2\eta^{\alpha_1\alpha_2} - q^{\alpha_1}q^{\alpha_2}\right] = 0$) on the right-hand side yields the condition

$$q \cdot \Pi\Delta' = 0.$$

Assuming that Δ' is invertible, this implies

$$q_\mu \Pi^{\mu\nu}(q) = 0. \tag{5.37}$$

The most general tensor $\Pi^{\mu\nu}(q)$ that can be constructed from q^μ and the metric tensor and which satisfies this condition can be written in the form[12]

$$\Pi^{\mu\nu}(q) = -\left(q^2\eta^{\mu\nu} - q^\mu q^\nu\right)\pi(q^2), \tag{5.38}$$

where $\pi(q^2)$ is a scalar function. This equation together with (5.32) imply that the full photon propagator can be expressed in terms of the function $\pi(q^2)$. However, in order to obtain such a formula we first need to specify of what form is the free propagator Δ.

The well-known complication related to the form of the free photon propagator is that the function $\Delta_{\mu\nu}(x-y)$ given by (5.22), as well as its momentum-space variant, do not transform under Lorentz transformations as tensors. This is because under Lorentz transformations the gauge field $a_\mu(x)$ transforms as a vector only up to a gauge transformation $\partial_\mu\omega(x)$, where $\omega(x)$ is a field constructed from the creation and annihilation operators of the photon, and whose precise form depends on the Lorentz transformation. For this reason we can consider only theories that have a gauge-invariant action. As a consequence many quantities are determined only up to a choice of gauge. Let us therefore work with a rather general formula for the free photon propagator,

$$\Delta_{\mu\nu}(q) = -\frac{\eta_{\mu\nu} - \xi(q^2)\frac{q_\mu q_\nu}{q^2+i\epsilon}}{q^2 + i\epsilon}, \tag{5.39}$$

[11] Earlier, when we were considering the sum of all one-particle-irreducible diagrams, we did not need to display that single insertion of the counterterm explicitly, and it was implicitly understood that its contribution has been included. That was quite consistent with the definition of that quantity. It would not be consistent, however, to include it into the second term of (5.35), since the diagrams represented by that term must contain two electromagnetic current vertices.

[12] To see this, one can start with the general ansatz $\Pi^{\mu\nu}(q) = A(q^2)\eta^{\mu\nu} + B(q^2)q^\mu q^\nu$ and apply the gauge-covariance condition.

where the function $\xi(q^2)$ depends on a particular choice of gauge that we choose to use.[13] To use the formula (5.32) we first need to calculate Δ^{-1}. A straightforward way to do that is to use a general ansatz

$$\left(\Delta^{-1}\right)_{\mu\nu}(q) = A(q^2)\eta_{\mu\nu} + B(q^2)q_\mu q_\nu,$$

and solve the equation $\Delta\Delta^{-1} = \mathbf{1}$ for A and B. We find

$$\left(\Delta^{-1}\right)_{\mu\nu}(q) = -q^2\eta_{\mu\nu} - \frac{\xi(q^2)}{1-\xi(q^2)}q_\mu q_\nu. \tag{5.40}$$

Note that in the Landau gauge, when we have $\xi(q^2) = 1$, this expression diverges. This is not surprising, because in that gauge $q \cdot \Delta(q) = 0$, which means that Δ is not invertible. We are going to ignore this issue and simply assume that $\xi(q^2) < 1$. We can now insert this result together with the expression (5.38) for Π into the equation (5.32) to obtain

$$\left(\Delta'^{-1}\right)_{\mu\nu}(q) = -\left(1-\pi(q^2)\right)q^2\eta_{\mu\nu} - \left(\frac{\xi(q^2)}{1-\xi(q^2)} + \pi(q^2)\right)q_\mu q_\nu. \tag{5.41}$$

Inverting this matrix then finally yields the explicit form of the full propagator expressed in terms of the function $\pi(q^2)$,

$$\Delta'_{\mu\nu}(q) = -\frac{\eta_{\mu\nu} - \left[\xi(q^2)\left(1-\pi(q^2)\right) + \pi(q^2)\right]\frac{q_\mu q_\nu}{q^2+i\epsilon}}{\left[1-\pi(q^2)\right]\left[q^2+i\epsilon\right]}. \tag{5.42}$$

The gauge-dependent term now contains the factor $\left[\xi(q^2)\left(1-\pi(q^2)\right) + \pi(q^2)\right]$ instead of $\xi(q^2)$. This can be important in practical calculations but in a gauge-invariant theory it should not have any physical consequences. Of a greater importance is the factor $1-\pi(q^2)$ in the denominator, which has measurable effects.

At this moment we might already try to answer some questions regarding the properties of the function $\pi(q^2)$. Perhaps the most pressing one of those is the question of the position of the pole of the full propagator (5.42). The factor $q^2 + i\epsilon$ in the denominator vanishes at $q^2 = 0$, so unless some other factor in the propagator cancels this zero the full propagator should still have a pole at $q^2 = 0$. The only factor in (5.42) that could possibly cancel the zero of $q^2 + i\epsilon$ is the factor $1-\pi(q^2)$ in the denominator. For this work the function $\pi(q^2)$ would need to have a simple pole at $q^2 = 0$. Now, the tensor $\Pi_{\mu\nu}(q)$ corresponds to the sum of one-particle-irreducible diagrams, which means that we expect it to have poles only at values of q^2 that correspond to masses of composite particles. Under usual circumstances there should be no such a composite particle of mass zero, so we do not expect $\Pi_{\mu\nu}(q)$ to

[13] The most common choices are the Feynman gauge for which $\xi = 0$ and the Landau (or Lorentz) gauge for which $\xi = 1$.

have a pole at $q^2 = 0$. Furthermore, $\Pi(q)_{\mu\nu}$ contains the term $q_\mu q_\nu \pi(q^2)$, so if $\pi(q^2)$ had a simple pole at $q^2 = 0$, the same would hold for $\Pi_{\mu\nu}(q)$. This suggests that $\pi(q^2)$ does not have a pole there. In other words, we do not expect radiative corrections to give photon a mass. Another topic that we can speak about right away is the normalization of $\pi(q^2)$ at $q^2 = 0$. This is because the renormalization prescription that we choose for the gauge field A_μ determines the value of the propagator's residue at its pole. Under the physical renormalization prescription, which is also known as the mass-shell prescription, we require that this residue be the same as that of the free propagator. This corresponds to the normalization condition

$$\pi(0) = 0. \tag{5.43}$$

Because the normalization condition (5.43) corresponds to the mass-shell renormalization prescription for the photon gauge field, it also fixes the renormalization prescription used for the electric charge. Let us label the charge renormalized according to this prescription by e.

The matter fields couple to the gauge field only through terms that originate from gauge-covariant derivatives of matter fields. The form (5.18) of those derivatives then ensures that in such interactions there is a single factor of the electric charge present at each end of the photon propagator. In other words, in electromagnetic interactions there is a single factor of e^2 associated with each photon propagator. This, combined with the fact that the running of the effective charge depends only on radiative corrections to the photon's propagator, means that it is often more convenient to speak about the running of the effective charge in terms of the energy dependence of the factor e^2 rather than of the charge e itself. There is an established dimensionless constant that can be used to describe the strength of the electromagnetic interaction and which contains the factor e^2: the *fine structure constant* α, defined as[14]

$$\alpha = \frac{e^2}{4\pi}. \tag{5.44}$$

The numerical value of this constant is roughly 1/137.036.

Instead of the physical renormalization prescription for the gauge field we could also consider some prescription characterized by a variable mass scale μ. The formula (5.42) shows that the gauge invariant part of the full propagator of the gauge field renormalized under the physical prescription is the same as that of the free propagator except for the extra factor of

[14]This holds when we use the Heaviside–Lorentz units for the electric charge (which effectively corresponds to us putting the vacuum permittivity ϵ_0 and the vacuum permeability μ_0 both equal to 1) and also put $\hbar = c = 1$. More generally, we have

$$\alpha = \frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c}.$$

$\left(1-\pi(q^2)\right)^{-1}$. It is therefore natural to define the field $A_\mu^{(\mu)}(x)$ renormalized at the scale μ by the relation

$$A_\mu^{(\mu)}(x) = \sqrt{1-\pi(\mu^2)}A_\mu(x), \tag{5.45}$$

where $A_\mu(x)$ is renormalized according the physical prescription. (I indicate the renormalization scale by the superscript "(μ)". The symbol "μ" appearing as a subscript denotes the component of the four-vector $A^{(\mu)}$.) The full propagator of the field $A_\mu^{(\mu)}(x)$ at the point $q^2=\mu^2$ then has the same value as does the free propagator. This prescription corresponds to the charge $e^{(\mu)}$ given by

$$e^{(\mu)} = \frac{e}{\sqrt{1-\pi(\mu^2)}}. \tag{5.46}$$

Accordingly, the running fine structure constant $\alpha(\mu^2)$ is defined as

$$\alpha(\mu^2) = \frac{\left[e^{(\mu)}\right]^2}{4\pi} = \frac{\alpha}{1-\pi(\mu^2)} = \frac{\alpha}{1-\Delta\alpha(\mu^2)}, \tag{5.47}$$

where it is customary in this context to denote $\pi(q^2)$ as $\Delta\alpha(q^2)$. We can think about this definition as associating the factor $\left(1-\pi(q^2)\right)^{-1}$ that appears in the full propagator (5.42) not with the propagator itself but instead with the two factors of the electric charge e that belong to the interaction vertices to which the propagator is connected. Note also that $\Delta\alpha(\mu^2)$ represents the contribution of the one-particle-irreducible insertion into the running of the fine structure constant.

In the usual canonical formalism it is required that the action be real, for otherwise the principle of stationary action would yield too many field equations.[15] For this reason we normally require the bare coupling constants that appear in the Lagrangian to be real. When speaking about renormalized coupling constants this restriction becomes less strict, and in principle we should be able to do calculations with complex renormalized couplings just fine, as long as we make sure that the counterterms make the overall action real. Nevertheless, such an approach leads to a number of complications and inconveniences, and for this reason it is usually better to apply

[15] In quantum field theory we often work with complex fields, but any such a complex field can always be described in terms of two real fields corresponding to its real and imaginary parts. If the action is real then the variational principle provides one equation for each such a real field. One more clarification is at place here: the requirement that the action be real does not always make it necessary for the Lagrangian to be real. The familiar example is provided by the Lagrangian density $\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi$ that is commonly used for the Dirac field, and which is not Hermitian. The difference between that Lagrangian density and its Hermitian conjugate, however, is a total derivative, and the resulting action is therefore real. [Note that this Lagrangian density represents a massless field and is expressed in the usual "PDG" notation, and not the notation which we used in the first part of this thesis.]

the formulas such as (5.46) or (5.47) only in those regions of μ^2 where the effective constants are real. (Which, for those two formulas would be for values of μ^2 that correspond to space-like four-momenta.) Another, related option is to employ such formulas everywhere but to discard any possible imaginary parts. There are, however, also many situation when it is useful to speak about complex running constants and in this section we will apply the definition (5.47) as it is also in the region of μ^2 corresponding to time-like values of the momentum transfer, where $\alpha(\mu^2)$ is complex.

5.2 The fine structure constant and the muon's magnetic anomaly

One of the contexts in which the knowledge of the running of the fine structure constant can be of great utility is the calculation of the magnetic anomaly of the muon. We already saw in the subsection 3.5 how the magnetic anomaly of a spin 1/2 particle can be calculated from the value of the Pauli form factor at zero momentum transfer. In the present subsection we are going to discuss the special case of the magnetic anomaly of the muon. We will focus on leading hadronic contributions to that quantity and their relation to the running of the fine structure constant.

Let us, however, first take a few steps back and discuss the magnetic moment of the muon in a greater generality. A massive particle of mass m and electric charge q has a magnetic dipole moment $\boldsymbol{\mu}$ that is proportional to its spin $\mathbf{s}$,

$$\boldsymbol{\mu} = g\frac{q}{2m}\mathbf{s}. \tag{5.48}$$

The dimensionless factor g that figures in this equation is called the *g-factor*[16] and in the subsection 3.5 we have already found the relation (3.82) between g and the Pauli form factor that holds if the particle has spin 1/2. Dirac, in his famous article [36] where he presented his relativistic quantum mechanics of the spin 1/2 particle predicted the value $g = 2$ for the electron. However, it became soon apparent that the true value of the electron's g-factor is not exactly equal to 2. One of the great successes of the new theory of quantum electrodynamics in 1940s was the calculation of the lowest order correction to the electron's g-factor by Schwinger [37]. Since then the magnetic moment of the electron has been calculated to several orders in perturbation theory and also measured to a very high precision and both quantities are in excellent agreement.[17]

[16] Sometimes g is also referred to as the *gyromagnetic ratio*. However, the same term is elsewhere used to describe instead the whole ratio between the magnetic moment $\boldsymbol{\mu}$ and the spin $\mathbf{s}$, that is, the full factor $gq/2m$. The latter terminology seems to me more correct, and accordingly, I will try to use the name *g-factor* rather than *gyromagnetic ratio* for g throughout this text.

[17] Considering the fact that the topic of the magnetic dipole moment has been so impor-

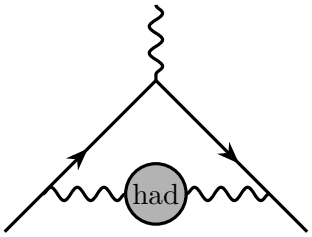


Figure 5.2: *The leading hadronic contribution to the magnetic moment.* The grey circle represents the strong-force contribution to the polarization of vacuum. If we interpret it as corresponding to the matrix element $(\Psi_0, \mathrm{T}\{J_\mu(x)J_\nu(y)\}\Psi_0)$ evaluated with only the strong interaction switched on, then this contribution is of the order α^2 relative to the tree diagram.

In the case of the magnetic moment of the muon the situation is quite different. The muon is about 207 times heavier than the electron and as a consequence its magnetic moment is much more sensitive to the effects of the weak and strong nuclear forces. In fact, at the precision that we have already achieved in laboratory the measurements of the magnetic moment of the muon are sensitive to all three forces of the standard model. This, coupled with the fact that there is a long-standing tension between experimentally measured values and theoretically calculated predictions, is the reason why the topic of the muon's magnetic moment attracts so much attention.

The fact that the measurement of the muon's magnetic moment is sensitive to all three forces of the standard model is a great advantage when it comes to testing the standard model but it also leads to some difficulties on the side of theoretical predictions. In particular, we must include into our calculations the effects of the strong interaction, and are therefore forced to confront the problem that in quantum chromodynamics we are unable to apply perturbation theory when calculating low energy processes. We must therefore look for some alternative approaches to estimate the strong-force contribution to the magnetic anomaly.

---

tant in the development of relativistic quantum field theory, one may wonder why other dipole moments, such as the electric or gravitational dipole moments, have played comparatively only minor roles in high energy physics. The reason is that for spin-1/2 particles those two moments are minuscule, which is related to the fact that the operation of the time reversal is very nearly a precise symmetry of physics. This is because in a relativistic quantum field theory, if the operator T of time inversion is conserved then energy eigenstates of any fermionic system must be at least twice degenerate. This phenomenon is called *Kramer's degeneracy*. A system containing a particle of a half-odd-integer spin in an external static electric or gravitational (but not magnetic) field is time-reversal invariant, but if that particle had any electric or gravitational dipole moment then the degeneracy among its different spin states would all be lost. Such moments are therefore forbidden in time-reversal invariant theories.

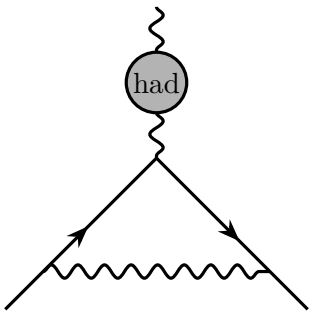


Figure 5.3: *Another insertion of the hadronic vacuum polarization into the diagram for the electromagnetic vertex.* Even though this diagram is of the same order in α as the diagram in Fig. 5.2 we can ignore it when calculating the magnetic moment. This is because its contribution vanishes when the four-momentum flowing through the external photon line is zero. (And for the same reason we do not consider the diagram that is just as this one but with the internal photon line which connects the two fermion lines removed. That diagram is only $\mathcal{O}(\alpha)$ relative to the tree diagram.)

The leading strong-force contribution to the magnetic anomaly is due to the lowest order hadronic contribution to the polarization of vacuum. The relevant diagram is depicted in Fig. 5.2. This diagram contains two extra factors of the fine structure constant relative to the tree diagram. Any contributions to the magnetic moment that contain hadronic vacuum polarization insertions are directly related to the hadronic contribution to the running of the fine structure constant. It is for this reason that $\alpha(s)$ is important in the evaluation of the muon's magnetic anomaly.

It might appear that there is another diagram that is of the same order in α and that has the hadronic vacuum polarization insertion on the external photon line, as is depicted in Fig. 5.3. This diagram would indeed contribute if we were considering the Pauli form factor at an arbitrary value of the four-momentum transfer q, but the magnetic moment is associated with the value at $q^2 = 0$, and because as a result of the condition (5.43) we have $\pi(q^2)\big|_{q^2=0} = 0$, this diagram has no influence on the magnetic moment.

It should be emphasized that the study of the contribution depicted in Fig. 5.2 to the magnetic moment of the muon is of crucial importance because it currently represents the largest source of uncertainty in the overall theoretical prediction. For a long time the only available approach to evaluate this contribution was to use analytic properties of the corresponding form factor and the optical theorem to express this contribution in terms of the total cross section for the annihilation of the electron–positron pair into hadrons [38, 39, 40, 41]. In this approach there are therefore two independent places where experimental data enter into consideration. First, it is necessary to measure the muon's magnetic moment directly, so that

this measurement can be compared with the theory. However, in order to produce a theoretical prediction yet another set of data, consisting of cross section measurements for the process $e^+e^- \to$ hadrons, is required so that we can calculate the strong-force contribution to the magnetic moment.

Let us now briefly review some technical aspects of this approach to the evaluation of the lowest order hadronic contribution to the muon's magnetic anomaly a_μ, depicted in Fig. 5.2. For details, the reader is encouraged to consult either the original sources [38, 39, 40, 41] or the book [42] by F. Jegerlehner, which is dedicated to the topic of the muon's magnetic anomaly. The detailed calculation shows that this contribution can be expressed as

$$a_\mu^{\mathrm{HVP,LO}} = \frac{\alpha^2}{3\pi^2} \int_{4m_\pi^2}^{\infty} ds \frac{K(s)}{s} R(s). \tag{5.49}$$

Here, R is the ratio of the "undressed"[18] total cross section of the annihilation of electron-positron pair into hadrons to the Born cross section of the process $e^+e^- \to \mu^+\mu^-$ calculated in the approximation of massless muons,

$$R(s) = \frac{\sigma_{\mathrm{tot}}^{\mathrm{undressed}}(e^+e^- \to \mathrm{hadrons})}{\frac{4\pi\alpha^2}{3s}}. \tag{5.50}$$

The symbol K represents a kernel function, which is given by the integral

$$K(s) = \int_0^1 dy \frac{y^2\,(1-y)}{y^2 + \frac{s}{m_\mu^2}\,(1-y)}. \tag{5.51}$$

This integral can be evaluated explicitly, and sometimes it is useful to consider that solution in its explicit form, which is

$$K(s) = \frac{x^2(2-x^2)}{2} + \frac{(1+x^2)(1+x)^2}{x^2}\left(\log(1+x) - x + \frac{x^2}{2}\right) + \frac{1+x}{1-x}x^2\log(x), \tag{5.52}$$

[18] The word "undressed" here means that the effects of the vacuum polarization on the photon propagator should be removed. That is, we need to multiply the measured cross section by the factor $\alpha^2/\alpha(s)^2$. (This discussion, as well as the derivation of (5.49) are precise only if the cross section is actually for the process $e^+e^- \to \gamma^* \to$ hadrons, that is, for the reaction mediated by a single intermediate virtual photon. It is usually a good approximation to simply ignore diagrams that do not conform to this picture, because their contribution to the cross section is of order α^2 or higher relative to the contribution of the process with only a single intermediate photon. The leading contribution is $\mathcal{O}(\alpha^2)$ relative to the tree diagram, so their contribution is $\mathcal{O}(\alpha^4)$, which is one order of α higher than the highest order contributions that are currently being significant relative to the uncertainty in the evaluation of the highest order contribution (5.49).) Alternatively, we can also use the original, "dressed" cross section in (5.50) but substitute $\alpha(s)^2$ for α^2 in the denominator of that equation.

where $x = \sqrt{1 - \frac{4m_\mu^2}{s}}$ is the velocity of the muon in the center-of-mass frame of a system of two muons whose total four-momentum squared is s.

Using currently available cross section data the uncertainty in the evaluation of the contribution (5.49) turns out to be rather large, which makes considerations of higher order corrections less critical from the practical point of view. Nevertheless, the total value of the next order corrections is not completely dominated by the uncertainty of (5.49), so it is still meaningful to consider those corrections even from a purely pragmatic perspective. Recall that the lowest order correction (5.49) is $\mathcal{O}(\alpha^2)$ relative to the tree diagram. There are five distinct classes of $\mathcal{O}(\alpha^3)$ corrections that contain hadronic contributions, depicted in figures 5.4–5.8. They were first calculated in [43, 44] and they are also discussed in [42]. The first class consists of diagrams with the single vacuum polarization insertion as in Fig. 5.2 but with one additional internal photon line. An example of such a diagram is depicted in Fig. 5.4. Another class consists of diagrams with a single hadronic vacuum polarization insertion and a single leptonic vacuum polarization loop, both on the same internal photon line, as depicted in Fig. 5.5. To account for contributions of these two classes of diagrams one can still use an equation of the form (5.49) but with an appropriately adjusted kernel function K. (See [43, 44] or [42].) The third class contains only one diagram, which has two hadronic vacuum polarization insertions and is depicted in Fig. 5.6. Its contribution can also be calculated from experimental cross section data, although the corresponding equation contains two factors of R and an adjusted kernel function,

$$a_\mu^{\text{HVP, class 3}} = \frac{1}{9}\left(\frac{\alpha}{\pi}\right)^3 \int_{4m_\pi^2}^{\infty} \frac{ds}{s}\frac{ds'}{s'} R(s)R(s') \times \int_0^1 dy \frac{y^4\,(1-y)}{\left[y^2 + \frac{s}{m_\mu^2}\,(1-y)\right]\left[y^2 + \frac{s'}{m_\mu^2}\,(1-y)\right]}. \quad (5.53)$$

The fourth class consists of all diagrams that correspond to Fig. 5.2 but with an additional internal photon line inside the hadronic vacuum polarization insertion. We can account for those contributions by calculating $R(s)$ from cross sections that are inclusive with respect to hard photons.[19] The last class of hadronic $\mathcal{O}(\alpha^3)$ diagrams contains the insertion of the subdiagram that represents the hadronic light by light scattering instead of the vacuum polarization. It is depicted in Fig. 5.8 and among all hadronic $\mathcal{O}(\alpha^3)$ diagrams it is the most problematic one. We will not discuss it here but will briefly return to it at the end of the present section.

For decades the dispersive approach had been the only viable option to calculate the leading order hadronic contributions. This started to change

[19] A discussion of this issue can be found in [42], or also [45, 46].

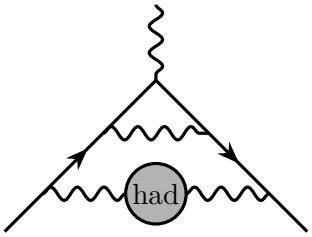


Figure 5.4: *An example of a "class 1" diagram for hadronic contributions to the magnetic moment.* This is one of several $\mathcal{O}(\alpha^3)$ diagrams that contain a single hadronic vacuum polarization insertion and a single additional internal photon line.

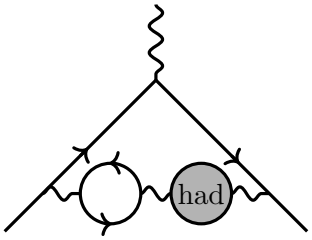


Figure 5.5: *A "class 2" diagram for hadronic contributions to the magnetic moment.* This is one of several $\mathcal{O}(\alpha^3)$ diagrams that contain a single hadronic vacuum polarization insertion and a single vacuum polarization lepton loop.

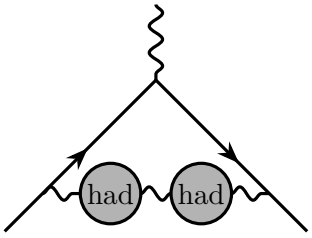


Figure 5.6: *The "class 3" diagram for hadronic contributions to the magnetic moment.* This $\mathcal{O}(\alpha^3)$ diagram contains two hadronic vacuum polarization insertions.

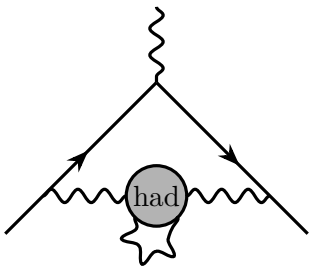


Figure 5.7: *A representation of a "class 4" diagram for hadronic contributions to the magnetic moment.* This is one of several $\mathcal{O}(\alpha^3)$ diagrams that contain a single hadronic vacuum polarization insertion with a single photon line inside.

in recent years, when the purely computational methods of lattice QCD began to approach levels of precision comparable to the dispersive approach. Particularly interesting is the recent work of Borsanyi *et al.* [47], which claims to have reached a precision similar to the recent works which employed the dispersive approach. The value of the hadronic contribution that they calculated shifted the overall prediction for the muon's magnetic moment closer towards its observed value, significantly reducing the tension between the theory and experiment. The discrepancy between the leading order hadronic contributions calculated by the dispersive approach and that computed in [47] is in the range of 2.0σ to 2.5σ. A further investigation will be needed to explain this discrepancy. (The author's suspicion is that the root cause of the discrepancy lies in the bad quality of the available cross section measurements of the electron–positron annihilation into pions and into kaons, where the available dataset contains tensions between individual experiments, suggesting that the authors of those experiments may have underestimated their systematic errors.) In any case, it is very likely that lattice QCD will play an increasingly important role in the evaluation of hadronic corrections.[20] These advances of lattice QCD, although very interesting and highly promising, do unfortunately lie outside of the scope of the present work and we will therefore not discuss them here any further.

There is, however, another recent development that is quite central to our topic. One can motivate it by noting that one of the reasons for the relatively small precision of the cross section data for the processes $e^+e^- \to$ hadrons, which are needed in the already mentioned dispersive approach, is that in the relevant energy region the total cross section exhibits abrupt fluctuations due to the presence of resonances and normal thresholds. This is related to the fact that the virtual photon created in the annihilation of an e^-e^+-pair carries a time-like four-momentum and the problem would therefore disappear if we could somehow estimate the hadronic contribu-

[20] For more recent results see also [48] and references cited therein.

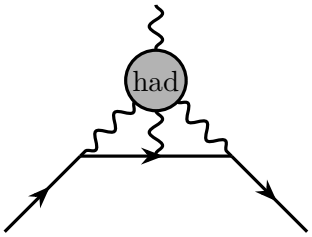


Figure 5.8: *A diagram representing the hadronic light-by-light contribution to the magnetic moment.* This contribution is of the order α^3 relative to the tree diagram. Together with Fig. 5.2 and the $\mathcal{O}(\alpha^3)$ hadronic vacuum polarization contributions, which are represented by diagrams in the figures 5.4–5.7, these are the most important diagrams in which the strong force contributes.

tion from measurements in the space-like rather than time-like region. In a sense, this is exactly what the new experiment MUonE suggested in [49] aims to accomplish. In this approach, using the dispersion relation for the polarization of vacuum, the hadronic contribution to the muon's magnetic moment is expressed in terms of the running fine structure constant $\alpha(s)$ in the space-like region. The proposed experiment then aims to deduce this running from very precise measurements of the cross section for elastic scattering of high energy muons on atomic electrons. This proposal is an iteration of an earlier suggestion by Calame *et al.* [50], where the Bhabha scattering $e^-e^+ \to e^-e^+$ was considered instead of the process $e\mu \to e\mu$. Proposals to measure the running of $\alpha(s)$ in space-like regions by the study of Bhabha scattering are even older, and date at least to the year 2004 when one such a proposal was given in [51]. Once the running of $\alpha(s)$ has been measured, one needs to subtract from $\Delta\alpha(s)$ its non-hadronic part in order to obtain the hadronic contribution $\Delta\alpha^{\text{had}}(s)$. The non-hadronic part can be calculated very precisely in perturbation theory.

Let us now explain little bit better the main idea behind this approach. We start from the formula (5.49) for the lowest order hadronic contribution to a_μ,

$$a_\mu^{\text{HVP,LO}} = \frac{\alpha^2}{3\pi^2}\int_{4m_\pi^2}^{\infty} ds \frac{K(s)}{s} R(s).$$

As we will discuss in greater detail in the section 5.4, the ratio $R(s)$ is directly related to the imaginary part of the hadronic contribution $\pi^{\text{had}}(s)$ to the vacuum polarization function,

$$\Im\left\{\pi^{\text{had}}(q^2 + i\epsilon)\right\} = -\frac{\alpha}{3} R(q^2). \tag{5.54}$$

We can therefore rewrite (5.49) also as

$$a_\mu^{\mathrm{HVP,LO}} = -\frac{\alpha}{\pi^2}\int_{4m_\pi^2}^{\infty} ds \frac{K(s)}{s} \Im\left\{\pi^{\mathrm{had}}(s+i\epsilon)\right\}. \tag{5.55}$$

Notice that although the integration is still restricted to time-like values of s, unlike $R(s)$ the hadronic vacuum polarization function $\pi(s)^{\mathrm{had}}$ is defined also for space-like value of s. In the section 5.5 we will see that $\pi^{\mathrm{had}}(s)$ should satisfy the once-subtracted dispersion relation

$$\pi^{\mathrm{had}}(s) = \frac{s}{\pi}\int_{4m_\pi^2}^{\infty} \frac{ds'}{s'} \frac{\Im\left\{\pi^{\mathrm{had}}(s'+i\epsilon)\right\}}{s'-s}. \tag{5.56}$$

If we now write in the integral (5.55) the kernel function (5.51) explicitly,

$$a_\mu^{\mathrm{HVP,LO}} = -\frac{\alpha}{\pi^2}\int_{4m_\pi^2}^{\infty} \frac{ds}{s}\int_0^1 dy\, \Im\left\{\pi^{\mathrm{had}}(s+i\epsilon)\right\} \frac{y^2(1-y)}{y^2+(1-y)\frac{s}{m_\mu^2}},$$

we can see that the integral over s resembles the dispersion integral for π^{had}. If we exchange the order of integration and slightly reorder the integrand the correspondence becomes more clear,

$$a_\mu^{\mathrm{HVP,LO}} = \frac{\alpha}{\pi}\int_0^1 dy\,(1-y)\frac{-\frac{y^2m_\mu^2}{1-y}}{\pi}\int_{4m_\pi^2}^{\infty}\frac{ds}{s}\frac{\Im\left\{\pi^{\mathrm{had}}(s+i\epsilon)\right\}}{s+\frac{y^2m_\mu^2}{1-y}}. \tag{5.57}$$

We can now see that although the original integral (5.55) was over time-like values of s, the integration over s in this double integral is actually in the form of the dispersion relation (5.56) and gives the value of π^{had} at the space-like value of $s = \tilde{s}(y) = -\frac{y^2m_\mu^2}{1-y} \leq 0$,

$$a_\mu^{\mathrm{HVP,LO}} = \frac{\alpha}{\pi}\int_0^1 dy\,(1-y)\pi^{\mathrm{had}}(\tilde{s}(y)). \tag{5.58}$$

As we mentioned earlier, the vacuum polarization function $\pi(s)$ is directly related to one-particle-irreducible effect on the running of the fine structure constant, $\Delta\alpha(s) = \pi(s)$. Hence, we can also write

$$a_\mu^{\mathrm{HVP,LO}} = \frac{\alpha}{\pi}\int_0^1 dx\,(1-x)\Delta\alpha^{\mathrm{had}}(\tilde{s}(x)). \tag{5.59}$$

This formula expresses $a_\mu^{\mathrm{HVP,LO}}$ directly in terms of the hadronic contribution to the running of the fine structure constant in the space-like region. Note that when we vary the variable x between 0 and 1, the values of $\tilde{s}(x) = -\frac{x^2m_\mu^2}{1-x}$ cover the whole negative semi-axis of s. This could pose a problem for practical applications of the formula (5.59) because we cannot

measure $\Delta\alpha(s)$ at arbitrarily large values of space-like s. (Not even mentioning that we cannot use QED and perturbation theory at arbitrarily large energy scales.) Luckily, this is where the factor $1 - x$, which is present in (5.59) comes into rescue: combined with the assumed asymptotic properties of $\pi^{\mathrm{had}}(s)$ this factor ensures that by ignoring the part of the integral (5.59) that corresponds to x between $1-\delta$ and 1 for some small positive δ we incur an error that can be bounded from above and made sufficiently small in a practically feasible experiment.

In relation to this topic, during my doctoral study I collaborated on the article [29], which considered these issues. In that work we used the U&A model to describe measured cross section data for relevant processes $e^+e^- \to$ hadrons, from which $R(s)$ and thus also $\Im\left\{\pi^{\mathrm{had}}(s)\right\}$ could be calculated. These results were then used in the dispersion relation (5.56) to predict the values of $\Delta\alpha^{\mathrm{had}}$ in the space-like region, which in turn were used in (5.59) to calculate $a_\mu^{\mathrm{HVP,LO}}$. We obtained the value

$$a_\mu^{\mathrm{HVP,LO}} = (707.23 \pm 4.81) \times 10^{-10}$$

for the lowest order hadronic contribution to the muon's magnetic anomaly. We then used predictions of the remaining contributions by other authors to obtain the overall prediction of the standard model. Namely, we used the value [52]

$$a_\mu^{\mathrm{QED}} = (11658471.8951 \pm 0.0080) \times 10^{-10}$$

for the QED contribution; the value [53]

$$a_\mu^{\mathrm{HVP,NLO}} = (-9.87 \pm 0.09) \times 10^{-10}$$

for hadronic vacuum polarization contributions of order $\mathcal{O}(\alpha^3)$, which are represented by diagrams in the figures 5.4–5.7, and the value [53]

$$a_\mu^{\mathrm{had,NNLO}} = (1.24 \pm 0.01) \times 10^{-10}$$

for $\mathcal{O}(\alpha^4)$ hadronic contributions; the value [54]

$$a_\mu^{\mathrm{HLbL}} = (10.5 \pm 2.6) \times 10^{-10}$$

for the $\mathcal{O}(\alpha^3)$ hadronic light-by-light contribution, which is represented by the diagram in Fig. 5.8; and the value [55]

$$a_\mu^{EW} = (15.36 \pm 0.1) \times 10^{-10}$$

for the weak force contribution. Thus, we have obtained the value [29]

$$a_\mu^{\mathrm{SM}} = (11659196.4 \pm 5.5) \times 10^{-10} \tag{5.60}$$

for the standard model prediction of the muon's magnetic anomaly. This deviated only by some 1.5σ from the 2021 world average experimental value [56]

$$a_\mu^{\text{EXP, 2021 avg}} = (11659206.1 \pm 4.1) \times 10^{-10}. \tag{5.61}$$

Let us briefly comment on this result. First, it may appear quite striking that the agreement with the average experimental value is so good, far better than in the case of the usually considered theoretical prediction. But that has nothing to do with us using the space-like approach (5.59). Instead, it is related to the fact that when the U&A model was in past used (in the dispersive approach (5.49)) to describe the cross section for the annihilation of the electron–positron pair into hadrons then it tended to yield results that were in better agreement with the experimentally measured value of a_μ than the consensual theoretical prediction. (See, for instance, [57].) It is definitely possible that this is because the physical requirements of analyticity, Hermitian reality, normalization, etc. that are "hardwired" into the model do partially correct the (low-quality) experimental cross section data. On the other hand, it is also possible that this improved agreement with experiment is just a coincidence or a result of some systematic error in the application of the model (considering that the model has always been applied by the same relatively small group of people). Nevertheless, even with all these qualifications the result (5.60) is interesting on its own. Second, the way in which the space-like approach (5.59) was employed in this article can only be regarded as an exercise that tested this approach and not a genuinely new result because the method was somewhat circular: we calculated space-like values of $\Delta\alpha^{\text{had}}$ from time-like values of $\Im\left\{\pi^{\text{had}}(s)\right\}$ using the dispersion relation (5.56) and then employed the formula (5.59) which calculates the anomaly from those space-like values but only because it relies on the same dispersion relation (5.56) to relate them back to time-like values of $\Im\left\{\pi^{\text{had}}(s)\right\}$ that figure in the original formula (5.49). Hence, we used the same dispersion relation twice, first in the direction "time-like region $\rightarrow$ space-like region" explicitly, and then in the opposite direction implicitly when we used the formula (5.59). The result should be therefore identical as if we had just calculated the anomaly directly from the time-like formula (5.49).

5.3 Measuring the running of the fine structure constant

As we have seen in previous subsections, the running of the fine structure constant plays an important role in high energy physics. As is also true with many other such quantities, our interest in studying and measuring this quantity is twofold. On the one hand, a physical theory should be

able to predict the value of $\alpha(s)$, and comparing such a prediction with measurements is useful in testing the validity and limits of that theory and our methods of calculation. In principle, it could even provide hints pointing in the direction of a better, more fundamental theory. On the other hand, $\alpha(s)$ itself enters into the calculation of many other measurable quantities, and once it is known — either from a calculation, a measurement, or a combination of both — it can be used to predict those quantities. We encountered an example of this in the previous section when we discussed the magnetic anomaly of the muon.

It is therefore worthwhile to attempt to measure the running of $\alpha(s)$. In the present section I will very briefly mention two such attempts. One of them is the proposed MUonE experiment [49], which we already mentioned in the previous section, and which has the ambition to measure the running of $\alpha(s)$ in the space-like region. The other one is an experiment that has been already carried out by the KLOE-2 Collaboration at DAΦNE and which measured both the real and imaginary parts of $\alpha(s)$ in the time-like region [58].

Let us start with the KLOE experiment [58]. This experiment used the initial state radiation technique to measure the process $e^+e^- \to \mu^+\mu^-\gamma(\gamma)$, in which the virtual photon created in the annihilation of the e^+e^--pair carried a time-like momentum in the energy range between 0.6 GeV and 0.975 GeV. The final state $\mu^+\mu^-\gamma(\gamma)$ contains one or more photons. One of those photons is the initial state radiation photon used to determine the center-of-mass energy that the e^+e^--pair had at the moment of its annihilation. At the precision achieved by this experiment one needs to take into account both the initial and final state radiation, and in [58] these effects were considered at the next-to-leading order using the PHOKHARA Monte Carlo event generator (see, e.g., [59, 58] and the references cited therein). At the achieved precision it was still possible to ignore the contribution of diagrams that contain two or more intermediate virtual photons, and under this approximation the ratio of the observed cross section to the Born cross section provides direct information on the magnitude of $\alpha(s)$,

$$\frac{d\sigma_{\rm exp}\left(e^+e^- \to \mu^+\mu^-(\gamma)\right)/d\sqrt{s}}{d\sigma_{\rm MC}^{Born}\left(e^+e^- \to \mu^+\mu^-(\gamma)\right)/d\sqrt{s}} = \left|\frac{\alpha(s)}{\alpha}\right|^2 . \tag{5.62}$$

The justification of this formula follows from the fact that while the Born cross section (ignoring now the final state radiation photons) is calculated from the diagram depicted in Fig. 5.9 that we will study in the next section, the experimental cross section – at least in the applied approximation — is given by the same diagram but with the intermediate free photon propagator $\Delta_{\mu\nu}$ replaced by its full variant $\Delta'_{\mu\nu}$.

In the next section we will also see that the imaginary part of $\alpha(s)$ can be determined by measuring the total cross section for the process $e^+e^- \to$

$\gamma^* \to$ "anything". In particular, the hadronic contribution to $\Delta\alpha(s)$, which cannot be calculated perturbatively at this energy scale, can be determined by the measurement of the process $e^+e^- \to \gamma^* \to$ "hadrons". In the next section we will review the derivation of this relation between the imaginary part of $\Delta\alpha^{\rm had}(s)$ and the total cross section for the process $e^+e^- \to \gamma^* \to$ "hadrons". This relation can be expressed in the form (5.100)

$$\Im\left\{\Delta\alpha^{\rm had}(s)\right\} = -\frac{\alpha}{3}R(s), \tag{5.63}$$

where the function R has been defined in (5.50) as the ratio of the total cross section for the process $e^+e^- \to \gamma^* \to$ "hadrons" to the total Born cross section for the process $e^+e^- \to \mu^+\mu^-$. The remaining leptonic contributions to $\Im\{\Delta\alpha(s)\}$ can be calculated in perturbation theory. Once we have measured both the modulus $|\alpha(s)|$ and determined the imaginary part of $\Delta\alpha(s)$ it is easy to extract also the real part of $\Delta\alpha(s)$. From the definition (5.47) we find

$$\frac{\alpha}{\alpha(s)} = 1 - \Delta\alpha(s), \tag{5.64}$$

which implies that

$$\left|\frac{\alpha}{\alpha(s)}\right|^2 = 1 + |\Delta\alpha(s)|^2 - 2\Re\{\Delta\alpha(s)\}. \tag{5.65}$$

If we now express $|\Delta\alpha(s)|^2$ on the right-hand side as the sum of the squares of the real and imaginary parts of $\Delta\alpha(s)$ we find

$$\left|\frac{\alpha}{\alpha(s)}\right|^2 - (\Im\{\Delta\alpha(s)\})^2 = (1 - \Re\{\Delta\alpha(s)\})^2, \tag{5.66}$$

where both quantities that appear on the left-hand side are presumed to be known. Hence, since we know that $1 - \Re\{\Delta\alpha(s)\} > 0$, we can write

$$\Re\{\Delta\alpha(s)\} = 1 - \sqrt{\left|\frac{\alpha}{\alpha(s)}\right|^2 - (\Im\{\Delta\alpha(s)\})^2}. \tag{5.67}$$

Let us now turn our attention to the MUonE experiment. This experiment was proposed in 2017 [49], although the main idea has been presented already two years earlier in [50]. A more detailed description of the technical aspects of the experiment can be found in the letter of intent [60]. In 2018 a test run of the experiment was carried out in CERN, the discussion of which can be found in [61, 62]. And finally, the most recent update regarding the main experiment appears to be [63].

As was explained in the previous section, the MUonE experiment aims to measure the running fine structure constant $\alpha(s)$ in the space-like region.

The quantity $\Delta\alpha(s)$ contains contributions from the QED and from the weak and strong nuclear forces,

$$\Delta\alpha(s) = \Delta\alpha^{\text{QED}}(s) + \Delta\alpha^{\text{weak}}(s) + \Delta\alpha^{\text{had}}(s). \tag{5.68}$$

Contributions from the QED and weak force can be calculated perturbatively with very high precision. One can therefore obtain $\Delta\alpha^{\text{had}}(s)$ by subtracting from the measured quantity $\Delta\alpha(s)$ the theoretically calculated values of $\Delta\alpha^{\text{QED}}(s)$ and $\Delta\alpha^{\text{weak}}(s)$. In this way we can obtain an independent evaluation of this problematic quantity.

The main promise of this new experiment is that it will provide a new way to directly measure the leading order hadronic contribution to the magnetic moment of the muon, using the equation (5.59),

$$a_\mu^{\text{HVP,LO}} = \frac{\alpha}{\pi}\int_0^1 dx\ (1-x)\Delta\alpha^{\text{had}}(\tilde{s}(x)),$$

with $\tilde{s}(x)$ defined as

$$\tilde{s}(x) = -\frac{x^2 m_\mu^2}{1-x}. \tag{5.69}$$

The contribution $a_\mu^{\text{HVP,LO}}$ is the dominant source of error in the overall prediction of a_μ. Furthermore, as we mentioned in the previous subsection, there also seems to be a discrepancy between the calculation of the lowest order hadronic contribution based on the old dispersive approach and the results coming from lattice QCD. In fact, the lattice QCD prediction of [47] is in acceptable agreement with the experiment. This suggests the possibility that the source of the long-lasting discrepancy between the theoretical predictions of a_μ and its experimental value might be hidden somewhere in the dispersive calculation. If nothing else, it is not hard to imagine the possibility that the measured cross section data for the annihilation of e^+e^- into pions and into kaons, which enter into that calculation, are not reliable enough and suffer from underestimated systematic errors. It would be therefore very desirable to have another, independent experimental way to estimate the leading order hadronic contribution to a_μ.

It should be noted that in order for the MUonE experiment to yield any useful results it must achieve a very high precision. To estimate the required precision, consider the ratio $\tilde{R}$ of the cross section for the process $e^+e^- \to$ "hadrons" to the same cross section but with only the electromagnetic and weak contributions to the running of $\alpha(s)$ included. Considering only the contribution of the diagrams containing a single intermediate full photon line, these two cross sections are the same except for the factor of $|\alpha(s)|^2$, which for one of them is evaluated with strong interactions switched on and

for the other one with strong interactions switched off. This leads to

$$\begin{aligned}\tilde{R}^{\rm LO}(s) &= \frac{d\sigma^{\rm LO}(s;\Delta\alpha^{\rm had}\neq 0)/ds}{d\sigma^{\rm LO}(s;\Delta\alpha^{\rm had}=0)/ds} \\ &= \left[\frac{1-\Delta\alpha^{\rm QED}(s)-\Delta\alpha^{\rm weak}(s)}{1-\Delta\alpha^{\rm QED}(s)-\Delta\alpha^{\rm weak}(s)-\Delta\alpha^{\rm had}(s)}\right]^2 \\ &= \left[\frac{1}{1-\frac{\Delta\alpha^{\rm had}(s)}{1-\Delta\alpha^{\rm QED}(s)-\Delta\alpha^{\rm weak}(s)}}\right]^2 \approxeq \left[1+\Delta\alpha^{\rm had}(s)\right]^2 \approxeq\ 1+2\Delta\alpha^{\rm had}(s).\end{aligned} \tag{5.70}$$

The quantity $\Delta\alpha^{\rm had}(s)$ varies only mildly across the kinematic range covered in this experiment and in the region that contributes most to the magnetic anomaly its value is of the order of 10^{-3}. For this measurement of the hadronic contribution to a_μ to be useful, a precision of at least $\mathcal{O}(10^{-2})$ is needed. Therefore a precision of $\mathcal{O}(10^{-5})$ in the measurement of the differential cross section must be achieved in this experiment. (See, for instance, [63].)

Note that in the equation (5.70) we have implicitly used the fact that in the space-like region the quantity $\Delta\alpha(s)$, as well as the individual contributions $\Delta\alpha^{\rm QED}(s)$, $\Delta\alpha^{\rm weak}(s)$, and $\Delta\alpha^{\rm had}(s)$, are all real. The differential cross section $d\sigma^{\rm LO}(s)/ds$ depends only on the absolute value $|\alpha(s)|$, and does not tell us anything about the phase of $\alpha(s)$ just by itself. If we considered the possibility that the individual contributions to $\Delta\alpha(s)$ could be complex, then under the same approximation that we used in (5.70) that equation would instead read

$$\begin{aligned}\tilde{R}^{\rm LO}(s) &= \left|\frac{1}{1-\frac{\Delta\alpha^{\rm had}(s)}{1-\Delta\alpha^{\rm QED}(s)-\Delta\alpha^{\rm weak}(s)}}\right|^2 \\ &\approxeq \left|1+\Delta\alpha^{\rm had}(s)\right|^2 \approxeq\ 1+2\Re\left\{\Delta\alpha^{\rm had}(s)\right\}.\end{aligned}$$

This would not invalidate the rough estimate of the required precision that has been presented in the previous paragraph, but the interpretation and usefulness of the results of such an experiment would be dependent on our ability to say something about the phase of $\alpha^{\rm had}(s)$. It is therefore a good thing that we know that in the relevant region $\alpha^{\rm had}(s)$ is real. We provide a proof of this property in Sec. 5.5.

Before we close this topic, let us note that only a few measurements of $\alpha(s)$ in the space-like region have been accomplished to date. A notable example is the measurement of $\alpha(s)$ from the small-angle Bhabha scattering by the OPAL collaboration [64]. This experiment was based on the proposal suggested earlier in [51]. The MUonE experiment is in principle very similar,

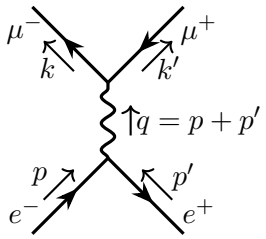


Figure 5.9: The tree diagram for the process $e^+e^- \to \mu^+\mu^-$.

but instead of the Bhabha scattering it considers the process[21] $\mu + e \to \mu + e$. One of the advantages of this approach is that since the two particles have very different masses, it is in a sense easier to analyze the final state data.

5.4 Imaginary part of the hadronic contribution to the vacuum polarization

We encountered the relation (5.54) between $R(s)$ and the imaginary part of $\pi^{\rm had}(s)$ already in the section 5.2, where we used it to calculate the lowest order hadronic contribution to the magnetic anomaly from cross section measurements of the processes $e^+e^- \to$ hadrons. In the present section we are going to look at the equation (5.54) a little bit more closely. This is an elementary topic, but in addition to filling in some details it will also help us to fix notation, since in this third part we are using the more standard notation, which is used also by the Particle Data Group. For a reference, this notation was also employed, for instance, in the books [14] or [42]. In our treatment we are going to be entirely straightforward, and will use only the most basic formulation of the optical theorem, which relates cross sections to forward scattering elements of the S-matrix.

Before we start with the main calculation, it might be useful to first calculate the total cross section for the process $e^+e^- \to \mu^+\mu^-$ in the Born approximation, with all the fermion masses set to zero. This is the quantity in the denominator of the definition (5.50) of R. There is only a single tree diagram that corresponds to this process, depicted in Fig. 5.9. We will work in the center-of-mass frame, and denote the four-momentum of the incoming electron by $p = (|\mathbf{p}|, \mathbf{p})$, and the four-momentum of the outgoing muon by $k = (|\mathbf{k}|, \mathbf{k})$. (Note that we are already making use of the approximation that $m_e = m_\mu = 0$.) The four-momenta p' and k' of the incoming positron and outgoing anti-muon, respectively, are therefore $p' = \mathcal{P}p = (|\mathbf{p}|, -\mathbf{p})$ and $k' = \mathcal{P}k = (|\mathbf{k}|, -\mathbf{k})$. (Here, $\mathcal{P}$ represents the Lorentz transformation of space inversion.) Also, we have $|\mathbf{p}| = |\mathbf{k}|$. The four-momentum q flowing through the intermediate photon line is $q = (2|\mathbf{p}|, 0) = (\sqrt{s}, 0)$, where s is

[21] Note that $\mu^- + e^- \to \mu^- + e^-$ and $\mu^+ + e^+ \to \mu^+ + e^+$ are for our purposes equivalent.

one of the Mandelstam variables, $s = (p + p')^2$. The usual Feynman rules[22] then associate with the diagram Fig. 5.9 the M-matrix given by

$$iM\left(p,\sigma;p',\sigma' \to k,\tilde{\sigma};k',\tilde{\sigma}'\right) = \left[\bar{u}(\mathbf{k},\tilde{\sigma})\left(-ie\gamma^{\mu}\right)v(-\mathbf{k},\tilde{\sigma}')\right]\frac{-i\eta_{\mu\nu}}{s+i\epsilon}\left[\bar{v}(-\mathbf{p},\sigma')\left(-ie\gamma^{\nu}\right)u(\mathbf{p},\sigma)\right]. \quad (5.71)$$

We have used the Feynman gauge for the propagator. Note that although we neglect the masses of the electron and the muon, that approximation is really justified only if $s >> m_{\mu}^2 > m_e^2 > 0$, which means that we should ignore any possible infra-red singularity at $s = 0$. For this reason we will assume that $s > 0$ and we can therefore drop the "$i\epsilon$" term in the denominator of the photon propagator. We can re-write (5.71) in a slightly simplified form

$$M = \frac{e^2}{s}\left[\bar{u}(\mathbf{k},\tilde{\sigma})\gamma^{\mu}v(-\mathbf{k},\tilde{\sigma}')\right]\left[\bar{v}(-\mathbf{p},\sigma')\gamma_{\mu}u(\mathbf{p},\sigma)\right], \quad (5.72)$$

where we now display the M-matrix simply as M. Using the standard manipulations[23] we can write

$$|M|^2 = \frac{e^4}{s^2}\mathrm{Tr}\left\{\left[u(\mathbf{k},\tilde{\sigma})\bar{u}(\mathbf{k},\tilde{\sigma})\right]\gamma^{\mu}\left[v(-\mathbf{k},\tilde{\sigma}')\bar{v}(-\mathbf{k},\tilde{\sigma}')\right]\gamma^{\nu}\right\} \times \mathrm{Tr}\left\{\left[v(-\mathbf{p},\sigma')\bar{v}(-\mathbf{p},\sigma')\right]\gamma_{\mu}\left[u(\mathbf{p},\sigma)\bar{u}(\mathbf{p},\sigma)\right]\gamma_{\nu}\right\}. \quad (5.73)$$

Our aim is to calculate the total cross section, so we will sum over the spin z-components $\tilde{\sigma}$ and $\tilde{\sigma}'$ of the final particles. We also average over all four distinct spin configurations of the initial particles. We denote the resulting quantity by $\overline{|M|^2}$,

$$\overline{|M|^2} = \frac{1}{4}\sum_{\sigma,\sigma',\tilde{\sigma},\tilde{\sigma}'}|M|^2. \quad (5.74)$$

This allows us to make use of the properties

$$\sum_{\sigma}u(\mathbf{p},\sigma)\bar{u}(\mathbf{p},\sigma) = \not{p} + m, \quad (5.75)$$

and

$$\sum_{\sigma}v(\mathbf{p},\sigma)\bar{v}(\mathbf{p},\sigma) = \not{p} - m, \quad (5.76)$$

[22] In our current convention they are specified for instance in the Appendix A of [14].

[23] See any textbook of quantum field theory for details. We use

$$\left[\bar{u}(\mathbf{k},\tilde{\sigma})\gamma^{\mu}v(-\mathbf{k},\tilde{\sigma}')\right]^{\dagger} = \left[\bar{v}(-\mathbf{k},\tilde{\sigma}')\gamma^{\mu}u(\mathbf{k},\tilde{\sigma})\right],$$

and a similar equation for the other factor; and the usual trick

$$\left[\bar{u}\gamma^{\mu}v\right]\left[\bar{v}\gamma^{\nu}u\right] = \mathrm{Tr}\left\{\left[\bar{u}\gamma^{\mu}v\right]\left[\bar{v}\gamma^{\nu}u\right]\right\} = \mathrm{Tr}\left\{u\bar{u}\gamma^{\mu}v\bar{v}\gamma^{\nu}\right\},$$

where in the last step we used the cyclic property of the trace.

where m stands for the mass of the fermion corresponding to the coefficient functions u, v. In our case $m_\mu = m_e = 0$, so we obtain the following simple result,

$$\overline{|M|^2} = \frac{e^4}{4s^2}\mathrm{Tr}\left\{\not{k}\gamma^\mu\hat{\not{k}}\gamma^\nu\right\}\mathrm{Tr}\left\{\hat{\not{p}}\gamma_\mu\not{p}\gamma_\nu\right\}, \tag{5.77}$$

where $\hat{\not{p}}$ and $\hat{\not{k}}$ represent $\gamma_\mu\mathcal{P}^\mu{}_\nu p^\nu$ and $\gamma_\mu\mathcal{P}^\mu{}_\nu k^\nu$, respectively. Employing now another standard result,

$$\mathrm{Tr}\left\{\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\right\} = 4\left(\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\rho\nu}\right), \tag{5.78}$$

we have

$$\mathrm{Tr}\left\{\not{k}\gamma^\mu\hat{\not{k}}\gamma^\nu\right\} = 4\left(k^\mu\left(\mathcal{P}k\right)^\nu + k^\nu\left(\mathcal{P}k\right)^\mu - \eta^{\mu\nu}\frac{s}{2}\right), \tag{5.79}$$

where we have used $k\cdot\mathcal{P}k = \left(k^0\right)^2 + \mathbf{k}^2 = 2|\mathbf{k}|^2 = s/2$. Similarly

$$\mathrm{Tr}\left\{\hat{\not{p}}\gamma_\mu\not{p}\gamma_\nu\right\} = 4\left(\left(\mathcal{P}p\right)^\mu p^\nu + \left(\mathcal{P}p\right)^\nu p^\mu - \eta^{\mu\nu}\frac{s}{2}\right). \tag{5.80}$$

The product of those two traces is

$$\mathrm{Tr}\left\{\not{k}\gamma^\mu\hat{\not{k}}\gamma^\nu\right\}\mathrm{Tr}\left\{\hat{\not{p}}\gamma_\mu\not{p}\gamma_\nu\right\} = 16\left[2\left(k\cdot\mathcal{P}p\right)^2 + 2\left(k\cdot p\right)^2\right], \tag{5.81}$$

which means that

$$\overline{|M|^2} = \frac{8e^4}{s^2}\left[\left(k\cdot\mathcal{P}p\right)^2 + \left(k\cdot p\right)^2\right]. \tag{5.82}$$

Now, to make even better use of our approximation of massless fermions we write $p\cdot k = -(p-k)^2/2 = \left(\mathbf{p}-\mathbf{k}\right)^2/2$, where the time component of $p-k$ vanishes because we are in the center-of-mass frame. Similarly, $\mathcal{P}p\cdot k = -(\mathcal{P}p-k)^2/2 = \left(\mathbf{p}+\mathbf{k}\right)^2/2$. If we denote by θ the angle between $\mathbf{p}$ and $\mathbf{k}$, we obtain

$$p\cdot k = \frac{1}{2}\left(\mathbf{p}^2 - 2\mathbf{p}\cdot\mathbf{k} + \mathbf{k}^2\right) = |\mathbf{p}|^2\left(1-\cos\theta\right) = \frac{s}{4}\left(1-\cos\theta\right), \tag{5.83}$$

and

$$\mathcal{P}p\cdot k = \frac{1}{2}\left(\mathbf{p}^2 + 2\mathbf{p}\cdot\mathbf{k} + \mathbf{k}^2\right) = |\mathbf{p}|^2\left(1+\cos\theta\right) = \frac{s}{4}\left(1+\cos\theta\right). \tag{5.84}$$

Using this in (5.82) yields

$$\overline{|M|^2} = \frac{e^4}{2}\left[\left(1+\cos\theta\right)^2 + \left(1-\cos\theta\right)^2\right]. \tag{5.85}$$

If the incoming particles are massless, then the differential cross section in the center-of-mass frame is given by [14]

$$\left.\frac{d\sigma}{d\Omega}\right|_{\rm C.M.} = \frac{1}{64(p^0)^2}\frac{1}{(2\pi)^2}\overline{|M|^2} = \frac{1}{64\pi^2 s}\overline{|M|^2}. \tag{5.86}$$

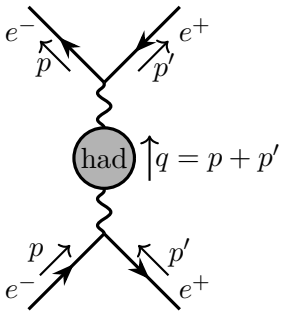


Figure 5.10: *The s-channel forward scattering diagram for the process* $e^+e^- \to e^+e^-$. At the lowest order in electromagnetism, this is the only (hadronic) forward scattering diagram for which the amplitude is not zero.

Thus, we found that

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{C.M.}} = \frac{e^4}{128\pi^2 s}\left[(1+\cos\theta)^2 + (1-\cos\theta)^2\right]. \tag{5.87}$$

Now,

$$\int d\Omega\ (1\pm\cos\theta)^2 = 2\pi\int_{-1}^{1}(d\cos\theta)\,(1\pm\cos\theta)^2 = \frac{16\pi}{3}, \tag{5.88}$$

which means that after we integrate the differential cross section over the angular variables we obtain the final result

$$\sigma_{\text{total}}^{\text{Born}}\left(e^+e^- \to \mu^+\mu^-\right) = \frac{e^4}{12\pi s} = \frac{4\pi\alpha^2}{3s}. \tag{5.89}$$

Let us now move on to the calculation of the M-matrix for the forward scattering of the e^+e^--pair at the lowest order in the electromagnetic interaction, but using the one-particle-irreducible insertion (5.38) restricted to hadrons. That is, we consider the forward scattering to the lowest order in electromagnetism but to all orders in strong interactions. We will again work in the approximation of massless fermions and consider the scattering in the center-of-mass frame. As before, we denote the four-momentum of the incoming electron by $p = (|\mathbf{p}|, \mathbf{p})$. The incoming positron has the four-momentum $p' = \mathcal{P}p = (|\mathbf{p}|, -\mathbf{p})$ and since we are now restricting our attention to the forward scattering, the outgoing electron and positron have four-momenta p and p'. We denote $q = p + p'$.

There are now two relevant Feynman diagrams, one corresponding to the s-channel process and the other one to the t-channel process, depicted in Fig. 5.10 and Fig. 5.11, respectively. However, the four-momentum flowing through the photon line in the t-channel diagram is zero, and this diagram

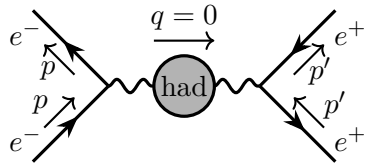


Figure 5.11: *The t-channel forward scattering diagram for the process* $e^+e^- \to e^+e^-$. Because the process is a forward scattering the four-momentum transfer q is zero. The value of the t-channel diagram therefore vanishes, because $\pi^{\rm had}(0) = 0$.

therefore does not contribute, since $\pi^{\rm had}(0) = 0$. We can therefore limit our attention to the diagram Fig. 5.10 only. Its corresponding contribution is

$$iM\left(p,\sigma;p',\sigma' \to p,\sigma;p',\sigma'\right) = \left[\bar{v}(-\mathbf{p},\sigma')\left(-ie\gamma^\mu\right)u(\mathbf{p},\sigma)\right]\frac{-i\eta_{\mu\alpha}}{s}$$
$$\times\left[-i\pi^{\rm had}(s)\left\{-\left(s\eta^{\alpha\beta}-q^\alpha q^\beta\right)\right\}\right]\frac{-i\eta_{\beta\nu}}{s}\left[\bar{u}(\mathbf{p},\sigma)\left(-ie\gamma^\nu\right)v(-\mathbf{p},\sigma')\right], \tag{5.90}$$

where, according to our discussion in the subsection 5.1.3, we inserted the value[24] $-i\Pi^{\alpha\beta}_{\rm had}(q)$ for the hadronic vacuum polarization insertion and used the formula (5.38) to express $\Pi^{\alpha\beta}_{\rm had}(q)$ in terms of the function $\pi^{\rm had}(s)$. We can rewrite this result more simply as

$$M = \frac{e^2}{s^2}\pi^{\rm had}(s)\left[\bar{v}(-\mathbf{p},\sigma')\gamma^\mu u(\mathbf{p},\sigma)\right]\left[\bar{u}(\mathbf{p},\sigma)\gamma^\nu v(-\mathbf{p},\sigma')\right]\left(s\eta_{\mu\nu}-q_\mu q_\nu\right). \tag{5.91}$$

This will be more convenient to calculate if we rewrite it in terms of a trace,

$$M = \frac{e^2}{s^2}\pi^{\rm had}(s){\rm Tr}\left\{v(-\mathbf{p},\sigma')\bar{v}(-\mathbf{p},\sigma')\gamma^\mu u(\mathbf{p},\sigma)\bar{u}(\mathbf{p},\sigma)\gamma^\nu\right\}\left(s\eta_{\mu\nu}-q_\mu q_\nu\right). \tag{5.92}$$

Next, we average over the spin z-components, which yields

$$\overline{M} = \frac{1}{4}\frac{e^2}{s^2}\pi^{\rm had}(s){\rm Tr}\left\{\hat{\not{p}}\gamma^\mu\not{p}\gamma^\nu\right\}\left(s\eta_{\mu\nu}-q_\mu q_\nu\right), \tag{5.93}$$

or, using our earlier result (5.80),

$$\overline{M} = \frac{e^2}{s^2}\pi^{\rm had}(s)\left(\left(\mathcal{P}p\right)^\mu p^\nu + \left(\mathcal{P}p\right)^\nu p^\mu - \eta^{\mu\nu}\frac{s}{2}\right)\left(s\eta_{\mu\nu}-q_\mu q_\nu\right). \tag{5.94}$$

Recalling now that $\left(\mathcal{P}p\right)^\mu p_\mu = s/2$ we obtain

$$\overline{M} = \frac{e^2}{s^2}\pi^{\rm had}(s)\left(-s^2 - 2\left(p\cdot q\right)\left(\mathcal{P}p\cdot q\right) + q^2\frac{s}{2}\right). \tag{5.95}$$

[24]Note that in order to conform to the Feynman rules of the Appendix A of [14] we had to cancel the $(2\pi)^4$-factor in the vacuum polarization insertion with the $(2\pi)^{-4}$ factor in one of the propagators.

However, since $q^2 = s$ and $(p \cdot q) = (\mathcal{P}p \cdot q) = s/2$, we obtain the final result

$$\overline{M} = -e^2 \pi^{\text{had}}(s) = -4\pi\alpha\, \pi^{\text{had}}(s). \tag{5.96}$$

Note that this quantity is of the order e^4 in the electromagnetic coupling, because two orders of e are hidden inside $\pi^{\text{had}}(s)$.

In the center-of-mass frame the optical theorem can be expressed in the form [14]

$$\Im\left\{M\left(\alpha \to \alpha\right)\right\} = 2\sqrt{s}\,|\mathbf{p}|\,\sigma_{\text{total}}\left(\alpha \to \text{anything}\right), \tag{5.97}$$

where $\mathbf{p}$ is the three-momentum of either of the two particles in the center-of-mass frame. In our case, when the particle is massless, we have $|\mathbf{p}| = \sqrt{s}/2$. Applying this to (5.96) yields

$$-4\pi\alpha\,\Im\left\{\pi^{\text{had}}(s)\right\} = 2\sqrt{s}\frac{\sqrt{s}}{2}\sigma^{\text{Born}}_{\text{total}}\left(e^+e^- \to \text{hadrons}\right). \tag{5.98}$$

The cross section on the right-hand side must correspond to the Born approximation, because it must be in the fourth order in e. There are only hadrons in the final state, because the other final states correspond to contributions to $\Im\{\pi(s)\}$ where leptons and electroweak bosons are in the intermediate states. We can rewrite this as

$$\Im\left\{\pi^{\text{had}}(s)\right\} = -\frac{\alpha}{3}\frac{\sigma^{\text{Born}}_{\text{total}}\left(e^+e^- \to \text{hadrons}\right)}{\frac{4\pi\alpha^2}{3s}}, \tag{5.99}$$

or, even more simply,

$$\Im\left\{\pi^{\text{had}}(s)\right\} = -\frac{\alpha}{3}R(s), \tag{5.100}$$

which is the formula (5.54) which we have used earlier in the section 5.2.

5.5 Spectral representation of the photon propagator

In this section we are going to investigate some of the analytic properties of the running fine structure constant $\alpha(s)$. In particular, we will provide an argument for the statement that $\alpha(s)$ is real in the space-like region, and derive the dispersion relation (5.56) for the function $\pi(q^2)$. Recall that the quantities $\alpha(s)$, $\pi(q^2)$ and the full propagator of the photon $i\Delta'_{\mu\nu}(q)$ are all related by the equation (5.42),

$$\Delta'_{\mu\nu}(q) = -\frac{\eta_{\mu\nu} - \left[\xi(q^2)\left(1 - \pi(q^2)\right) + \pi(q^2)\right]\frac{q_\mu q_\nu}{q^2 + i\epsilon}}{\left[1 - \pi(q^2)\right]\left[q^2 + i\epsilon\right]},$$

the identification $\pi(s) = \Delta\alpha(s)$, and the equation (5.47),

$$\alpha(s) = \frac{\alpha}{1 - \Delta\alpha(s)}.$$

Our first goal is to show that the propagator $\Delta'_{\mu\nu}(q)$ is real for space-like values of q. In fact, this is just a special case of the Hermitian analyticity that we mentioned earlier in Sec. 2.5, but compared to the more general properties of Hermitian analyticity or generalized unitarity, the reality condition for the propagator is easier to prove. Probably the most straightforward way to address this issue is to use the *Källén–Lehmann* spectral representation of the full propagator. This is a standard topic that is covered in most textbooks of quantum field theory, and I will therefore attempt to be brief in its presentation.

What we are really interested in is the photon propagator, or the two-point function

$$i\Delta'_{\mu\nu}(x, y) = (\Psi_0, \mathrm{T}\left\{A_\mu(x)A_\nu(y)\right\}\Psi_0)\,. \tag{5.101}$$

The study of this function is complicated by the fact that the field A_μ does not transform as a four-vector under Lorentz transformations. For this reason, it will be a little bit more convenient to consider instead the Green's function

$$(\Psi_0, \mathrm{T}\left\{J_\mu(x)J_\nu(y)\right\}\Psi_0)\,, \tag{5.102}$$

where J_μ is the electromagnetic current. Unlike the gauge field, the operator J_μ is Lorentz covariant. Furthermore, it is Hermitian and satisfies the current conservation condition $\partial_\mu J^\mu = 0$. These properties will make our work easier. Now, the gauge field A couples only to the current J, and this means that the two Green's functions (5.101) and (5.102) are closely related. By considering them in perturbation theory one can see that the relation between those two Green's functions can be expressed in terms of the equation (5.35),

$$\begin{aligned} i\Delta'_{\mu\nu}(q) = {} & i\Delta_{\mu\nu}(q) - [i\Delta_{\mu\alpha_1}(q)]\, M_{00}^{\alpha_1\alpha_2}(q,-q)\, [i\Delta_{\alpha_2\nu}(q)] \\ & + [i\Delta_{\mu\alpha_1}(q)]\, (-i(Z_3-1))\left[q^2\eta^{\alpha_1\alpha_2} - q^{\alpha_1}q^{\alpha_2}\right] [i\Delta_{\alpha_2\nu}(q)]\,. \end{aligned} \tag{5.103}$$

5.5.1 A naive derivation of the spectral representation

In order to obtain the spectral representation for the two-point function (5.102) we need first to consider instead the fixed-ordered vacuum expecta-

tion value[25]

$$(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0)\,. \tag{5.104}$$

To obtain the spectral representation of this fixed-order matrix element we start by inserting between the operators $J^\mu(x)$ and $J^\nu(0)$ a sum over a complete set of physical states $\{\Psi_n\}$ of definite momenta,

$$P^\mu\Psi_n = p_n^\mu\Psi_n. \tag{5.105}$$

Then we can use the translation property (2.104) to write

$$\begin{aligned}(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0) &= \sum_n \mathrm{e}^{-ip_n x}\,(\Psi_0, J^\mu(0)\Psi_n)\,(\Psi_n, J^\nu(0)\Psi_0)\\ &= \sum_n \mathrm{e}^{-ip_n x}\,(\Psi_0, J^\mu(0)\Psi_n)\,(\Psi_0, J^\nu(0)\Psi_n)^*\,. \end{aligned}\tag{5.106}$$

Note that the meaning of the sum "$\sum_n$" is fixed by the normalization condition and Lorentz transformation properties of the set $\{\Psi_n\}$. If we denote by n_Λ the four-momentum and spin labels of the state that is obtained by the Lorentz transformation Λ acting on the state Ψ_n, then if the set $\{\Psi_n\}$ was constructed according to the convention that we are currently using for asymptotic states — that is, following the covariant normalization — then we have

$$U(\Lambda)\Psi_n = \Psi_{n_\Lambda}. \tag{5.107}$$

Accordingly, to keep the sum over intermediate states Lorentz invariant, the sum must contain for each particle an integral with the Lorentz invariant measure

$$\int \frac{d^3\mathbf{p}}{(2\pi)^3 2\sqrt{\mathbf{p}^2+m^2}}. \tag{5.108}$$

The factor $(2\pi)^{-3}$ is present only to conform to our current normalization convention and will not concern us here.

An important step now is to express (5.106) in the form

$$\begin{aligned}&(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0) =\\ &\qquad\int d^4p\,\mathrm{e}^{-ipx}\sum_n \delta^4\,(p-p_n)\,(\Psi_0, J^\mu(0)\Psi_n)\,(\Psi_0, J^\nu(0)\Psi_n)^*\,.\end{aligned}\tag{5.109}$$

The point of rewriting it in this form is that the expression

$$\sum_n \delta^4\,(p-p_n)\,(\Psi_0, J^\mu(0)\Psi_n)\,(\Psi_0, J^\nu(0)\Psi_n)^*$$

[25] To discard the explicit dependence on the coordinate y we made use of the translation property (2.104), which implies that

$$(\Psi_0, J^\mu(x)J^\nu(y)\Psi_0) = (\Psi_0, J^\mu(x-y)J^\nu(0)\Psi_0)\,,$$

and then renamed $x-y$ to x.

is a tensor function of p, which we will denote by $\mathcal{J}^{\mu\nu}(p)$. It is not too difficult to see that $\mathcal{J}^{\mu\nu}(p)$ is indeed a tensor, or, in other words, that $\mathcal{J}^{\mu\nu}(\Lambda p) = \Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta}\mathcal{J}^{\alpha\beta}(p)$. We have

$$\begin{aligned}\mathcal{J}^{\mu\nu}(\Lambda p) &= \sum_n \delta^4\left(\Lambda p - p_n\right)\left(\Psi_0, J^{\mu}(0)\Psi_n\right)\left(\Psi_0, J^{\nu}(0)\Psi_n\right)^* \\ &= \sum_n \delta^4\left(\Lambda p - \Lambda p_{n_{\Lambda^{-1}}}\right)\left(\Psi_0, U(\Lambda)^{-1}J^{\mu}(0)U(\Lambda)\Psi_{n_{\Lambda^{-1}}}\right) \\ &\qquad \times \left(\Psi_0, U(\Lambda)^{-1}J^{\nu}(0)U(\Lambda)\Psi_{n_{\Lambda^{-1}}}\right)^*, \quad (5.110)\end{aligned}$$

where we expressed p_n as the result of the Lorentz transformation Λ acting on the four-momentum corresponding to the state $n_{\Lambda^{-1}}$, and in a similar way we wrote $\Psi_n = U(\Lambda)\Psi_{n_{\Lambda^{-1}}}$. We also used the Lorentz invariance property of the vacuum state, $\Psi_0 = U(\Lambda)\Psi_0$. Now, because J^{μ} is a vector operator, it transform as

$$U(\Lambda)^{-1}J^{\mu}(x)U(\Lambda) = \Lambda^{\mu}{}_{\nu}J^{\nu}(\Lambda^{-1}x). \quad (5.111)$$

Furthermore, under the covariant normalization the sum $\sum_n$ is Lorentz invariant, so we can just as well write it as the sum over $n_{\Lambda^{-1}}$. Altogether, we find

$$\begin{aligned}&\mathcal{J}^{\mu\nu}(\Lambda p) = \\ &\sum_{n_{\Lambda^{-1}}} \delta^4\left(\Lambda p - \Lambda p_{n_{\Lambda^{-1}}}\right)\left(\Psi_0, \Lambda^{\mu}{}_{\alpha}J^{\alpha}(0)\Psi_{n_{\Lambda^{-1}}}\right)\left(\Psi_0, \Lambda^{\nu}{}_{\beta}J^{\beta}(0)\Psi_{n_{\Lambda^{-1}}}\right)^* \\ &= \Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta}\sum_{n_{\Lambda^{-1}}} \delta^4\left(p - p_{n_{\Lambda^{-1}}}\right)\left(\Psi_0, J^{\alpha}(0)\Psi_{n_{\Lambda^{-1}}}\right)\left(\Psi_0, J^{\beta}(0)\Psi_{n_{\Lambda^{-1}}}\right)^* \\ &\qquad = \Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta}\mathcal{J}^{\alpha\beta}(p), \quad (5.112)\end{aligned}$$

where we also made use of the fact that the measure d^4p is Lorentz invariant, so that

$$\delta^4\left(\Lambda p - \Lambda p_{n_{\Lambda^{-1}}}\right) = \delta^4\left(p - p_{n_{\Lambda^{-1}}}\right).$$

Now, if we apply the current conservation condition $\partial_{\mu}J^{\mu}(x) = 0$ to (5.109), we find that $\mathcal{J}^{\mu\nu}$ must satisfy the condition

$$p_{\mu}\mathcal{J}^{\mu\nu}(p) = 0. \quad (5.113)$$

But because $\mathcal{J}^{\mu\nu}$ is a tensor that depends only on p, this condition implies that it must be of the form

$$\mathcal{J}^{\mu\nu}(p) = \left(p^{\mu}p^{\nu} - p^2\eta^{\mu\nu}\right)F(p),$$

where $F(p)$ is some scalar function of p. F can, therefore, depend only on p^2 and for time-like values of p also on the sign of p^0.[26] In fact, $F(p)$ must

[26]Note that we are considering now only transformation properties with respect to restricted Lorentz transformations. In particular, when in (5.110) we expressed p_n as $\Lambda p_{n_{\Lambda^{-1}}}$ we assumed that Λ does not reverse the direction of time.

be proportional to $\theta(p^0)$, because all the physical states in the sum

$$\mathcal{J}^{\mu\nu}(p) = \sum_n \delta^4\left(p - p_n\right)\left(\Psi_0, J^\mu(0)\Psi_n\right)\left(\Psi_0, J^\nu(0)\Psi_n\right)^* \tag{5.114}$$

have non-negative energy, $p_n^0 \geq 0$. Hence, if $p^0 < 0$ then $\mathcal{J}^{\mu\nu}(p)$ should vanish. We can therefore write

$$\mathcal{J}^{\mu\nu}(p) = \frac{1}{(2\pi)^3}\left(p^\mu p^\nu - p^2\eta^{\mu\nu}\right)\rho(p^2)\theta(p^0), \tag{5.115}$$

where we extracted the factor $(2\pi)^{-3}$ for later convenience. Now, using the definition (5.114) and the formula (5.115), we can immediately infer some important properties of the spectral function ρ. First, the definition (5.114) implies that any diagonal component of $\mathcal{J}^{\mu\nu}$ must be real and non-negative. The formula (5.115) then implies that $\rho(p^2)$ must also be real and non-negative. The fact that $\rho(p^2)$ is real follows immediately from the fact that all other factors in (5.115) are real. And if we consider any diagonal component of the formula (5.115) a little bit more closely, we can also easily notice that ρ cannot be negative. For instance, if we look at the component 00 of (5.115), the property $p^0p^0 - p^2\eta^{00} = (p^0)^2 - p^2 = \mathbf{p}^2 \geq 0$, together with $\mathcal{J}^{00} \geq 0$, implies that $\rho(p^2) \geq 0$. And lastly, if for some p^2 there is no physical state n such that $p_n^2 = p^2$, then (5.114) implies that $\rho(p^2) = 0$. In particular, $\rho(p^2) = 0$ for all $p^2 < 0$, since physical states carry only null or time-like four-momenta.[27]

If we now insert (5.115) back into (5.109), we obtain

$$\left(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0\right) = \frac{1}{(2\pi)^3}\int d^4p\, \mathrm{e}^{-ipx}\left(p^\mu p^\nu - p^2\eta^{\mu\nu}\right)\rho(p^2)\theta(p^0). \tag{5.116}$$

To proceed further we need to express $\rho(p^2)$ in (5.116) as

$$\rho(p^2) = \int_0^\infty d\mu^2\delta(p^2 - \mu^2)\rho(\mu^2)$$

and exchange the order of integration over d^4p and $d\mu^2$. Two remarks are at place here. First, note that in specifying in the above equation the integration of μ^2 as starting at 0 and going to $+\infty$ we made use of our earlier observation that $\rho(\mu^2) = 0$ for $\mu^2 < 0$. Second, strictly speaking changing the order of integration is not allowed here (see [9]). For this reason we will later derive the reality property also by other methods. For now, however, let us ignore this issue and interchange the order of integrals anyway. We

[27] This is the reason why the factor $\theta(p^0)$ in the formula (5.115) does not ruin the Lorentz covariance properties of $\mathcal{J}^{\mu\nu}(p)$. The sign of p^0 is generally not invariant under restricted Lorentz transformations, but $\mathcal{J}^{\mu\nu}(p)$ is non-zero only for p null or space-like, and for such four-momenta the sign of p^0 indeed is invariant.

obtain

$$\begin{aligned}(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0) \\ &= \int_0^\infty d\mu^2 \rho(\mu^2) \frac{1}{(2\pi)^3} \int d^4p\, \mathrm{e}^{-ipx} \left(p^\mu p^\nu - p^2\eta^{\mu\nu}\right) \delta(p^2-\mu^2)\theta(p^0) \\ &= \int_0^\infty d\mu^2 \rho(\mu^2) \left(-\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^2\right) \frac{1}{(2\pi)^3} \int d^4p\, \mathrm{e}^{-ipx} \delta(p^2-\mu^2)\theta(p^0),\end{aligned} \tag{5.117}$$

where the derivatives are with respect to the coordinate x. Thus, our final result is

$$(\Psi_0, J^\mu(x)J^\nu(0)\Psi_0) = \int_0^\infty d\mu^2 \rho(\mu^2) \left(-\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^2\right) \Delta_+^{(\mu)}(x), \tag{5.118}$$

where $\Delta_+^{(\mu)}(x)$ stands for the familiar function

$$\begin{aligned}\Delta_+^{(\mu)}(x) &= \frac{1}{(2\pi)^3} \int d^4p\, \mathrm{e}^{-ipx} \delta(p^2-\mu^2)\theta(p^0) \\ &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2\sqrt{\mathbf{p}^2+\mu^2}}\, \mathrm{e}^{-i\sqrt{\mathbf{p}^2+\mu^2}x^0 + i\mathbf{p}\cdot\mathbf{x}}.\end{aligned} \tag{5.119}$$

The function $\Delta_+^{(\mu)}(x)$, when multiplied by $\theta(x^0)$ and combined with its space-time inverted counterpart, yields the free propagator of the scalar field

$$\theta(x^0)\Delta_+^{(\mu)}(x) + \theta(-x^0)\Delta_+^{(\mu)}(-x) = i\Delta_{\mathrm{F}}^{(\mu)}(x). \tag{5.120}$$

Here, $\Delta_{\mathrm{F}}^{(\mu)}$ is the Feynman propagator

$$\Delta_{\mathrm{F}}^{(\mu)}(x) = \int \frac{d^4q}{(2\pi)^4} \frac{\mathrm{e}^{-iqx}}{q^2-\mu^2+i\epsilon}. \tag{5.121}$$

We can carry out a completely analogous analysis also of the matrix element

$$(\Psi_0, J^\nu(0)J^\mu(x)\Psi_0)\,. \tag{5.122}$$

Instead of (5.109), we find

$$(\Psi_0, J^\nu(0)J^\mu(x)\Psi_0) = \int d^4p\, \mathrm{e}^{ipx} \overline{\mathcal{J}}^{\mu\nu}(p), \tag{5.123}$$

where $\overline{\mathcal{J}}^{\mu\nu}(p)$ is given by

$$\overline{\mathcal{J}}^{\mu\nu}(p) = \sum_n \delta^4(p-p_n)\,(\Psi_n, J^\mu(0)\Psi_0)\,(\Psi_n, J^\nu(0)\Psi_0)^*\,. \tag{5.124}$$

Just as before with $\mathcal{J}^{\mu\nu}$ we can show that $\overline{\mathcal{J}}^{\mu\nu}$ is a tensor function of p, and because of the current conservation condition and the positivity of energy of physical states it must be of the form

$$\overline{\mathcal{J}}^{\mu\nu} = \frac{1}{(2\pi)^3}\left(p^\mu p^\nu - p^2\eta^{\mu\nu}\right)\overline{\rho}(p^2)\theta(p^0). \tag{5.125}$$

Now, from the formulas (5.115) and (5.125) we see that both $\mathcal{J}^{\mu\nu}$ and $\overline{\mathcal{J}}^{\mu\nu}$ are symmetric under the interchange of μ and ν. But that implies that the right-hand sides of their definitions (5.114) and (5.124) are equal to each other. For instance, after we exchange the indices in (5.124) we can use the Hermiticity of the current J to obtain the right-hand side of (5.114). This implies that $\mathcal{J}^{\mu\nu} = \overline{\mathcal{J}}^{\mu\nu}$, and therefore[28]

$$\overline{\rho}(p^2) = \rho(p^2). \tag{5.126}$$

If we now carry out the steps that led from (5.116) to (5.118), but starting with (5.123), we obtain

$$(\Psi_0, J^\nu(0)J^\mu(x)\Psi_0) = \int_0^\infty d\mu^2\rho(\mu^2)\left(-\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^2\right)\Delta_+^{(\mu)}(-x). \tag{5.127}$$

We can combine the results (5.118) and (5.127) to obtain

$$\begin{aligned}(\Psi_0, \mathrm{T}\left\{J^\mu(x)J^\nu(0)\right\}\Psi_0) &= \int_0^\infty d\mu^2\rho(\mu^2)\left(-\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^2\right)\left[\theta(x^0)\Delta_+^{(\mu)}(x) + \theta(-x^0)\Delta_+^{(\mu)}(-x)\right] \\ &= \int_0^\infty d\mu^2\rho(\mu^2)\left(-\partial^\mu\partial^\nu + \eta^{\mu\nu}\partial^2\right) i\Delta_{\mathrm{F}}^{(\mu)}(x). \end{aligned} \tag{5.128}$$

This is the *Källén–Lehmann* spectral representation of the two-point function (5.102) that we were looking for,

$$(\Psi_0, \mathrm{T}\left\{J^\mu(x)J^\nu(0)\right\}\Psi_0) = \int_0^\infty d\mu^2\rho(\mu^2)\int\frac{d^4q}{(2\pi)^4}\frac{i\left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right)\mathrm{e}^{-iqx}}{q^2 - \mu^2 + i\epsilon}. \tag{5.129}$$

We can now insert this result into the equation (5.35)

$$\begin{aligned}i\Delta'_{\mu\nu}(q) = i\Delta_{\mu\nu}(q) &- \left[i\Delta_{\mu\alpha_1}(q)\right]M_{00}^{\alpha_1\alpha_2}(q,-q)\left[i\Delta_{\alpha_2\nu}(q)\right] \\ &+ \left[i\Delta_{\mu\alpha_1}(q)\right](-i(Z_3-1))\left[q^2\eta^{\alpha_1\alpha_2} - q^{\alpha_1}q^{\alpha_2}\right]\left[i\Delta_{\alpha_2\nu}(q)\right].\end{aligned}$$

[28] Even if J^μ was not Hermitian we would still be able to proceed, but we would need to use the condition of local commutativity and the symmetry property $\Delta_+^{(\mu)}(x) = \Delta_+^{(\mu)}(-x)$ which holds if x is space-like.

If we use the Feynman gauge for the propagator, we find

$$\Delta'_{\mu\nu}(q) = \frac{-\eta_{\mu\nu}}{q^2+i\epsilon} + \frac{-\eta_{\mu\nu}+\frac{q_\mu q_\nu}{q^2+i\epsilon}}{q^2+i\epsilon}\left[\int_0^\infty d\mu^2\rho(\mu^2)\frac{1}{q^2-\mu^2+i\epsilon} - (Z_3-1)\right]. \tag{5.130}$$

We are considering the full propagator $\Delta'_{\mu\nu}(q)$ in the physical renormalization scheme, which means that the gauge-invariant part of its residue at $q^2 = 0$ must be the same as that of the free propagator. In other words, when $q^2 = 0$ the expression in the square brackets must vanish, which allows us to express the constant Z_3 in terms of the spectral function $\rho(\mu^2)$. Instead of (5.130), we can write

$$\Delta'_{\mu\nu}(q) = \frac{-\eta_{\mu\nu}}{q^2+i\epsilon} + \frac{-\eta_{\mu\nu}+\frac{q_\mu q_\nu}{q^2+i\epsilon}}{q^2+i\epsilon}\left[\int_0^\infty d\mu^2\rho(\mu^2)\frac{1}{q^2-\mu^2+i\epsilon} + \int_0^\infty d\mu^2\frac{\rho(\mu^2)}{\mu^2}\right]. \tag{5.131}$$

Now, the only way this expression can yield a complex value is if either q^2 in the denominator factor $q^2+i\epsilon$ or $q^2-\mu^2$ in the denominator factor $q^2-\mu^2+i\epsilon$ vanishes, because those are the only circumstances under which the "$i\epsilon$"-terms matter, and everything else in that formula is real. But for q space-like we have $q^2 < 0$, which means that none of the denominators vanishes and the "$i\epsilon$"-terms play no role. Hence, the value of (5.131) is manifestly real for space-like values of q. This, together with the formula (5.42), implies that in the space-like region $\pi(q^2) = \Delta\alpha(q^2)$ is real.

This might be a good place to take a step back and compare more closely our result (5.131) to the expression (5.42) for the full photon propagator that we found earlier. In (5.131) we used the Feynman gauge, which means that in (5.42) the function $\xi(q^2)$ is identically zero, and that formula therefore reads

$$\Delta'_{\mu\nu}(q) = \frac{-\eta_{\mu\nu}+\pi(q^2)\frac{q_\mu q_\nu}{q^2+i\epsilon}}{[1-\pi(q^2)]\,[q^2+i\epsilon]}. \tag{5.132}$$

Comparing the coefficients of $\eta_{\mu\nu}$ we find

$$\frac{1}{1-\pi(q^2)} = 1 + \int_0^\infty d\mu^2\rho(\mu^2)\left[\frac{1}{q^2-\mu^2+i\epsilon}+\frac{1}{\mu^2}\right], \tag{5.133}$$

or

$$\pi(q^2) = \frac{\int_0^\infty d\mu^2\rho(\mu^2)\left[\frac{1}{q^2-\mu^2+i\epsilon}+\frac{1}{\mu^2}\right]}{1+\int_0^\infty d\mu^2\rho(\mu^2)\left[\frac{1}{q^2-\mu^2+i\epsilon}+\frac{1}{\mu^2}\right]}. \tag{5.134}$$

Thus we can also see that the factor $\pi(q^2)/(1-\pi(q^2))$ that stands in front of the $q_\mu q_\nu/[q^2+i\epsilon]^2$ in (5.132) is the same as the corresponding factor $\int_0^\infty d\mu^2\rho(\mu^2)\left[\frac{1}{q^2-\mu^2+i\epsilon}+\frac{1}{\mu^2}\right]$ in (5.131), indicating that our manipulations

have been internally consistent. However, one should be cautious when making judgments regarding the properties of $\Delta'_{\mu\nu}(q)$ or $\pi(q^2)$ from the results (5.131)–(5.134), because we derived those results using mathematically unsound arguments. For this reason we will now present another argument that $\pi(q^2)$ is real in the space-like region.

5.5.2 Derivation of the spectral representation by dispersion-theoretical methods

The argument that we are going to present now relies on standard dispersion theoretical methods, and the approach presented here more or less follows the treatment of a different problem presented in [9], although I attempt to discuss some details that seem to be often ignored in the literature. We will be interested in the analytic properties of the Green's function $M_{00}^{\alpha_1\alpha_2}(q,-q)$. To avoid any confusion, let us recall that in (5.33) we defined this function as

$$(2\pi)^4\delta^4\left(q_1+q_2\right)\, M_{00}^{\mu\nu}(q_1,q_2) = \int d^4x_1\, d^4x_2\, \mathrm{e}^{iq_1x_1}\mathrm{e}^{iq_2x_2}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x_1)J^\nu(x_2)\right\}\Psi_0\right). \quad (5.135)$$

The translation invariance of the matrix element on the right-hand side implies that

$$\begin{aligned}
&\int d^4x_1\, d^4x_2\, \mathrm{e}^{iq_1x_1}\mathrm{e}^{iq_2x_2}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x_1)J^\nu(x_2)\right\}\Psi_0\right)\\
&\quad=\int d^4x_2\, d^4x_1\, \mathrm{e}^{iq_1(x_1-x_2)}\mathrm{e}^{i(q_2+q_1)x_2}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x_1-x_2)J^\nu(0)\right\}\Psi_0\right)\\
&\qquad=\int d^4x_2\, d^4x\, \mathrm{e}^{iq_1x}\mathrm{e}^{i(q_2+q_1)x_2}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x)J^\nu(0)\right\}\Psi_0\right)\\
&\qquad\quad=(2\pi)^4\delta^4\left(q_1+q_2\right)\int d^4x\, \mathrm{e}^{iq_1x}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x)J^\nu(0)\right\}\Psi_0\right).
\end{aligned}$$

That is, instead of (5.135) we can use the simpler definition

$$M_{00}^{\mu\nu}(q,-q)=\int d^4x\, \mathrm{e}^{iqx}\left(\Psi_0,\mathrm{T}\left\{J^\mu(x)J^\nu(0)\right\}\Psi_0\right). \quad (5.136)$$

Let us start by showing that the quantity $M_{00}^{\mu\nu}(q,-q)$ is a tensor function of q.[29] We have

$$\begin{aligned} M_{00}^{\mu\nu}(\Lambda q,-\Lambda q) &= \int d^4x\, \mathrm{e}^{i(\Lambda q)x}\, (\Psi_0, \mathrm{T}\,\{J^\mu(x)J^\nu(0)\}\,\Psi_0) \\ &= \int d^4x\, \mathrm{e}^{i(\Lambda q)x}\Big(\Psi_0, \theta(x^0)\left[U(\Lambda)J^\mu(x)U(\Lambda)^{-1}\right]\left[U(\Lambda)J^\nu(0)U(\Lambda)^{-1}\right] \\ &\qquad + \theta(-x^0)\left[U(\Lambda)J^\nu(0)U(\Lambda)^{-1}\right]\left[U(\Lambda)J^\mu(x)U(\Lambda)^{-1}\right]\Psi_0\Big), \quad (5.137) \end{aligned}$$

where we have used the Lorentz invariance of the vacuum to suitably insert the operators $U(\Lambda)$, $U(\Lambda)^{-1}$. The point of these steps is that we want to express the right-hand side as an integral over $\Lambda^{-1}x$, and we have $U(\Lambda)J^\mu(x)U(\Lambda)^{-1} = \Lambda^\mu{}_\alpha J^\alpha(\Lambda^{-1}x)$ and $U(\Lambda)J^\nu(x)U(\Lambda)^{-1} = \Lambda^\nu{}_\beta J^\beta(0)$. We would also like to write $\theta\left((\Lambda^{-1}x)^0\right)$ and $\theta\left(-(\Lambda^{-1}x)^0\right)$ instead of $\theta(x^0)$ and $\theta(-x^0)$, but that requires some discussion, because the sign of x^0 is not Lorentz invariant. However, a restricted Lorentz transformation Λ can change the sign of x^0 only if x is space-like, and for such values of x the order of $J^\mu(x)$ and $J^\nu(0)$ is immaterial, because of the property of local commutativity. For that reason we are indeed allowed to change $\theta(x^0)$ and $\theta(-x^0)$ to $\theta\left((\Lambda^{-1}x)^0\right)$ and $\theta\left(-(\Lambda^{-1}x)^0\right)$ in (5.137). After we apply these steps in (5.137) we find

$$\begin{aligned} M_{00}^{\mu\nu}(\Lambda q,-\Lambda q) &= \int d^4x\, \mathrm{e}^{i(\Lambda q)x}\Big(\Psi_0, \theta\left(\left(\Lambda^{-1}x\right)^0\right)\Lambda^\mu{}_\alpha J^\alpha(\Lambda^{-1}x)\Lambda^\nu{}_\beta J^\beta(0) \\ &\qquad + \theta\left(-\left(\Lambda^{-1}x\right)^0\right)\Lambda^\nu{}_\beta J^\beta(0)\Lambda^\mu{}_\alpha J^\alpha(\Lambda^{-1}x)\Psi_0\Big) \\ &= \Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta \int d^4\left(\Lambda^{-1}x\right)\, \mathrm{e}^{iq\left(\Lambda^{-1}x\right)}\Big(\Psi_0, \theta\left(\left(\Lambda^{-1}x\right)^0\right) J^\alpha(\Lambda^{-1}x)J^\beta(0) \\ &\qquad + \theta\left(-\left(\Lambda^{-1}x\right)^0\right) J^\beta(0)J^\alpha(\Lambda^{-1}x)\Psi_0\Big), \quad (5.138) \end{aligned}$$

where we used the Lorentz invariance of the measure d^4x and the property $\Lambda q\cdot x = q\cdot\Lambda^{-1}x$. But this means that

$$M_{00}^{\mu\nu}(\Lambda q,-\Lambda q) = \Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta M_{00}^{\alpha\beta}(q,-q). \quad (5.139)$$

The gauge invariance property (5.34) then implies that $M_{00}^{\mu\nu}(q,-q)$ is of the form

$$M_{00}^{\mu\nu}(q,-q) = \left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right)\tilde{M}(q), \quad (5.140)$$

where $\tilde{M}$ is a scalar function of q. Thus, it can depend on q^2 and in principle also on the sign of q^0. However, unlike the functions $\mathcal{J}^{\mu\nu}$ and $\overline{\mathcal{J}}^{\mu\nu}$ that we

[29] Recall that when we speak about quantities being "tensors" we presently consider their transformation properties only with respect to restricted Lorentz transformations.

encountered earlier, the function $\tilde{M}$ is symmetric under the inversion of q. We have

$$M_{00}^{\mu\nu}(-q,q) = \int d^4x\, \mathrm{e}^{i(-q)\cdot x}\, (\Psi_0, \mathrm{T}\,\{J^\mu(x)J^\nu(0)\}\,\Psi_0) \\ = \int d^4x\, \mathrm{e}^{i(-q)\cdot x}\, (\Psi_0, \mathrm{T}\,\{J^\mu(0)J^\nu(-x)\}\,\Psi_0)\,, \quad (5.141)$$

because of the translation invariance of the Green's function

$$(\Psi_0, \mathrm{T}\,\{J^\mu(x)J^\nu(0)\}\,\Psi_0)\,.$$

If we now substitute $x \to -x$ and use the property $d^4x = d^4(-x)$, we obtain

$$M_{00}^{\mu\nu}(-q,q) = \int d^4(-x)\, \mathrm{e}^{i(-q)\cdot(-x)}\, (\Psi_0, \mathrm{T}\,\{J^\nu(x)J^\mu(0)\}\,\Psi_0) \\ = \int d^4x\, \mathrm{e}^{iq\cdot x}\, (\Psi_0, \mathrm{T}\,\{J^\nu(x)J^\mu(0)\}\,\Psi_0) = M_{00}^{\nu\mu}(q,-q). \quad (5.142)$$

If we combine this result with the formula (5.140) we find that $\tilde{M}(-q) = \tilde{M}(q)$. Thus, $\tilde{M}(q)$ depends only on q^2 and not on the sign of q^0, and instead of (5.140) we can write

$$M_{00}^{\mu\nu}(q,-q) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M(q^2). \quad (5.143)$$

To gain insight into analytic properties of $M_{00}^{\mu\nu}(q,-q)$ we will use the method which we have briefly described at the end of Sec. 2.1. We can rewrite the time-ordered product $\mathrm{T}\,\{J^\mu(x)J^\nu(0)\}$ in two different ways, either as

$$\mathrm{T}\,\{J^\mu(x)J^\nu(0)\} = J^\mu(x)J^\nu(0) - \theta(-x^0)\,[J^\mu(x), J^\nu(0)]\,, \quad (5.144)$$

or as

$$\mathrm{T}\,\{J^\mu(x)J^\nu(0)\} = J^\nu(0)J^\mu(x) + \theta(x^0)\,[J^\mu(x), J^\nu(0)]\,. \quad (5.145)$$

The point of this is that because of the condition of local commutativity (also known as microscopic causality) the commutator terms vanish for x space-like. This, combined with the step function factors that appear in front of those commutators, restricts the support of those terms to values of x that are either in the causal future or the causal past of the point $x = 0$. More specifically, the support of the term $\theta(-x^0)\,[J^\mu(x), J^\nu(0)]$ is restricted to past-pointing null or time-like vectors x, or, in the notation of Sec. 2.3, it is restricted to $\bar{V}_-$. The support of $\theta(x^0)\,[J^\mu(x), J^\nu(0)]$ is restricted to future-pointing null or time-like values of x, that is, to $x \in \bar{V}_+$. The results that we discussed in Sec. 2.3 then imply that the Laplace transform

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, -\theta(-x^0)\,[J^\mu(x), J^\nu(0)]\,\Psi_0\right) \quad (5.146)$$

is analytic for all q in the tube Γ_-, that is, for those four-vectors q that have arbitrary real components but whose imaginary part is inside the past light-cone. In other words, it is analytic on the set of four-vectors q of the form $q = k + il$, where k and l are real four-vectors, with k being otherwise arbitrary and l satisfying the conditions $l^2 > 0$ and $l^0 < 0$. We also saw that such a Laplace transform is bounded by a polynomial for large values of q. Similarly, the support of $\theta(x^0)\,[J^\mu(x), J^\nu(0)]$ is restricted to $x \in \bar{V}_+$, which implies that the Laplace transform

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, \theta(x^0)\,[J^\mu(x), J^\nu(0)]\,\Psi_0\right) \tag{5.147}$$

is analytic in the tube Γ_+ and polynomially bounded for large q.

Now, we already know that $M_{00}^{\mu\nu}$ is of the form (5.143),

$$M_{00}^{\mu\nu}(q,-q) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M(q^2),$$

and if we use the decomposition (5.144), we find that

$$\begin{aligned}\left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M(q^2) = M_{00}^{\mu\nu}(q,-q) = \\ \int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, \mathrm{T}\,\{J^\mu(x) J^\nu(0)\}\,\Psi_0) = \int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, J^\mu(x) J^\nu(0) \Psi_0) \\ + \int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, -\theta(-x^0)\,[J^\mu(x), J^\nu(0)]\,\Psi_0\right). \end{aligned} \tag{5.148}$$

We have already encountered the matrix element that appears in the first term of the rightmost expression: it can be expressed in terms of the formula (5.116),

$$(\Psi_0, J^\mu(x) J^\nu(0) \Psi_0) = \frac{1}{(2\pi)^3} \int d^4p\, \mathrm{e}^{-ipx} \left(p^\mu p^\nu - p^2 \eta^{\mu\nu}\right) \rho(p^2) \theta(p^0).$$

We can thus easily evaluate the integral,

$$\int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, J^\mu(x) J^\nu(0) \Psi_0) = 2\pi \left(q^\mu q^\nu - q^2 \eta^{\mu\nu}\right) \rho(q^2) \theta(q^0). \tag{5.149}$$

Now, when we insert (5.149) into (5.148) we find that the commutator term must be of the form

$$\begin{aligned}\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, -\theta(-x^0)\,[J^\mu(x), J^\nu(0)]\,\Psi_0\right) \\ = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_-(q^2, \mathrm{sgn}(q^0)), \end{aligned} \tag{5.150}$$

where M_- depends both on q^2 and the sign of q^0.

Similarly, our earlier results (5.123) and (5.125) show that

$$(\Psi_0, J^\nu(0) J^\mu(x) \Psi_0) = \frac{1}{(2\pi)^3} \int d^4p\, \mathrm{e}^{ipx} \left(p^\mu p^\nu - p^2 \eta^{\mu\nu}\right) \overline{\rho}(p^2) \theta(p^0). \tag{5.151}$$

Using the Hermiticity of the current J, combined with the symmetry of the above expression under the interchange of μ and ν, we also found that $\rho(p^2) = \overline{\rho}(p^2)$. This implies

$$\int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, J^\nu(0) J^\mu(x) \Psi_0) = 2\pi \left(q^\mu q^\nu - q^2 \eta^{\mu\nu}\right) \rho(q^2) \theta(-q^0). \quad (5.152)$$

If we combine this with the variant of (5.148) that we obtain if we use (5.145) instead of (5.144),

$$\left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M(q^2) = M_{00}^{\mu\nu}(q, -q) = \int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, J^\nu(0) J^\mu(x) \Psi_0) + \int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, \theta(x^0) \left[J^\mu(x), J^\nu(0)\right] \Psi_0\right), \quad (5.153)$$

we find that

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, \theta(x^0) \left[J^\mu(x), J^\nu(0)\right] \Psi_0\right) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2, \mathrm{sgn}(q^0)). \quad (5.154)$$

Let us now take a step back so that we do not get confused. Just a moment ago we claimed that the left-hand sides of (5.150) and (5.154) are analytic functions of q in Γ_- and Γ_+, respectively. But on the right-hand sides of those equations there appear factors that depend on the sign of q^0, which clearly is not analytic in q. This is because the equations (5.150) and (5.154) are valid only for real four-vectors q, because to derive them we used (5.149) and (5.152), which are valid only for real q. The tubes Γ_- and Γ_+ do not contain the real hyperplane of q, and the equations (5.150) and (5.154) represent only the boundary values of functions analytic in the regions Γ_- and Γ_+, respectively.

Now, the question arises, what do the equations (5.150) and (5.154) tell us about the form of $\int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, -\theta(-x^0) \left[J^\mu(x), J^\nu(0)\right] \Psi_0)$ in Γ_-, respectively the form of $\int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, \theta(x^0) \left[J^\mu(x), J^\nu(0)\right] \Psi_0)$ in Γ_+? Let us focus first on the case of $\int d^4x\, \mathrm{e}^{iqx}\, (\Psi_0, -\theta(-x^0) \left[J^\mu(x), J^\nu(0)\right] \Psi_0)$. Consider some closed connected region R of the real hyperplane of q on which none of the components of $(q^\mu q^\nu - \eta^{\mu\nu} q^2)$ vanishes and which does not intersect the hyperplane $q^0 = 0$. Since q are real, we can use the formula (5.150) but since $\mathrm{sgn}(q^0)$ has the same value on the whole region R, the function $M_-(q^2, \mathrm{sgn}(q^0))$ depends only on q^2 there. Let us denote it by $\tilde{M}_-(q^2)$. On this region, $(q^\mu q^\nu - \eta^{\mu\nu} q^2)\, \tilde{M}_-(q^2)$ is the boundary value of a function analytic in the tube Γ_-, which we temporarily denote by $\mathcal{F}^{\mu\nu}(q)$. But the region R is closed and $(q^\mu q^\nu - \eta^{\mu\nu} q^2)$ does not vanish on it, which means that there exists some open neighborhood (in the complex space of q) $\mathcal{N}$ of R on which $(q^\mu q^\nu - \eta^{\mu\nu} q^2)$ does not vanish. Hence, the function $\mathcal{F}^{\mu\nu}(q) / (q^\mu q^\nu - \eta^{\mu\nu} q^2)$ is analytic on the open set $\Gamma_- \cap \mathcal{N}$ and $\tilde{M}_-(q^2)$ is its boundary value. (Although it might not be a function and exist only as a tempered distribution.)

But that means that the function $\mathcal{F}^{\mu\nu}(q)/\left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right)$, defined on $\Gamma_- \cap \mathcal{N}$, is an analytic function only of q^2. Thus, on $\Gamma_- \cap \mathcal{N}$ the function $\mathcal{F}^{\mu\nu}(q)$ is of the form

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, -\theta(-x^0)\left[J^\mu(x), J^\nu(0)\right]\Psi_0\right) = \left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right) M_-(q^2), \quad (5.155)$$

where $M_-(q^2)$ is an analytic function on $\Gamma_- \cap \mathcal{N}$. But since the left-hand side is an analytic function on the whole tube Γ_-, it must be possible to continue $M_-(q^2)$ onto the whole tube and the formula (5.155) holds on the whole of Γ_-. Using the same argument, we can show that

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, \theta(x^0)\left[J^\mu(x), J^\nu(0)\right]\Psi_0\right) = \left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right) M_+(q^2) \quad (5.156)$$

is valid on the whole tube Γ_+ and $M_+(q^2)$ is an analytic function there.

Now, the equations (5.150) and (5.154) show that when q approaches a time-like or null point[30] on the real hyperplane then the functions $M_\pm(q^2)$ have different limits in the half-space $q^0 > 0$ than in the half-space $q^0 < 0$. This implies that $M_\pm(q^2)$ are not analytic on the whole plane of q^2, and have a branch cut there. As defined here, the branch cut goes along the real semi-axis $q^2 \geq 0$ and $M_\pm(q^2)$ have different values depending on whether they approach the branch cut from above or from below.

Now that we have established that the tensor functions which we are studying are all of the form

$$\left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right) \times \text{“a scalar function of } q^2 \text{ and } \mathrm{sgn}(q^0)\text{”},$$

we can simplify the analysis by focusing only on a particular one-dimensional subset of the tubes Γ_+ and Γ_-. We choose an arbitrary future time-like vector l, $l^0 > 0$, that is normalized to satisfy $l^2 = 1$, and restrict our attention to four-momenta q of the form ωl, where ω is a complex number. If $\Im\{\omega\} > 0$ then $q \in \Gamma_+$ and if $\Im\{\omega\} < 0$ then $q \in \Gamma_-$. That is,

$$\mathcal{F}^{\mu\nu}_-(\omega) = \int d^4x\, \mathrm{e}^{i\omega l\cdot x} \left(\Psi_0, -\theta(-x^0)\left[J^\mu(x), J^\nu(0)\right]\Psi_0\right), \quad (5.157)$$

considered now as a function of ω is analytic and bounded by a polynomial in the lower half-plane of ω and

$$\mathcal{F}^{\mu\nu}_+(\omega) = \int d^4x\, \mathrm{e}^{i\omega l\cdot x} \left(\Psi_0, \theta(x^0)\left[J^\mu(x), J^\nu(0)\right]\Psi_0\right) \quad (5.158)$$

[30] We can see that this dependence on $\mathrm{sgn}(q^0)$ exists only for causal four-vectors q by considering the Lorentz transformation properties of $M^{\mu\nu}_{00}(q,-q)$ and the other quantities that appear after we use the decomposition (5.144) or the decomposition (5.145). Because all those quantities transform as tensors under restricted Lorentz transformations, there can be no dependence on $\mathrm{sgn}(q^0)$ when q is space-like. Below, we will see that this is correct also by another consideration.

is analytic and polynomially bounded in the upper half-plane of ω. Using now the equations (5.155) and (5.156) we can write

$$\mathcal{F}_{-}^{\mu\nu}(\omega) = (l^\mu l^\nu - \eta^{\mu\nu})\omega^2 M_{-}(\omega^2), \tag{5.159}$$

if ω is in the lower half-plane, and

$$\mathcal{F}_{+}^{\mu\nu}(\omega) = (l^\mu l^\nu - \eta^{\mu\nu})\omega^2 M_{+}(\omega^2), \tag{5.160}$$

if ω is in the upper half-plane. Accordingly, the functions $M_{-}(\omega^2)$ and $M_{+}(\omega^2)$ are analytic on the whole ω^2-plane except for the real positive axis, $\omega^2 \geq 0$. They are also bounded by a polynomial. In fact, it turns out that we can provide a full characterization of the function $M_{00}^{\mu\nu}(q, -q)$ by considerations of either one of the functions $M_{-}(\omega^2)$ and $M_{+}(\omega^2)$. To make the presentation more simple, let us now pick one of those two functions, say $M_{+}(\omega^2)$, and focus our attention only on it. If we instead used $M_{-}(\omega^2)$ the treatment would be completely analogous.

Before we continue I would like to briefly discuss one point that can be somewhat confusing.[31] In the following discussion we are going to study the function M_{+}, and we will be particularly interested in its values at real ω^2. According to the equation (5.160), the function $M_{+}(\omega^2)$ is defined only when ω is in the upper half-plane. But that means that on the definition domain of $M_{+}(\omega^2)$, ω^2 can never attain a positive real value. Clearly, to speak about the value of M_{+} for $\omega^2 > 0$ we need to consider some suitable limit, which corresponds to $q = \omega l$ approaching the real hyperplane. However, unlike positive real values, negative real values of ω^2 can be attained inside the definition domain of $M_{+}(\omega^2)$. Let us focus our present discussion on this case. To obtain a negative real value of ω^2 in the upper half-plane of ω, ω must be purely imaginary. This corresponds to a four-vector $q = +i|\omega|l$, which is deep inside the tube Γ_{+}, far off the real hyperplane. Ultimately, however, we are interested in real four-momenta: that is, in the boundary value of M_{+} as given in the equation (5.154). How do we know that for a real q, with $q^2 < 0$, these values are the same as those for purely imaginary ω, far off the real axis? The answer consists of two components, both of which we have already used, at least implicitly, earlier. First, in Sec. 2.3 we discussed Theorem 5 which says that if we approach from inside the tube Γ_{+} the real hyperplane of q then the corresponding limit of $M_{+}(q^2)$ is the value $M_{+}(q^2, \mathrm{sgn}(q^0))$ defined by the equation (5.150). Second, inside Γ_{+} the function $M_{+}(q^2)$ is analytic and depends only on q^2. Thus, as we approach a space-like real q, the function $M_{+}(q^2)$ must approach the same value which it attains at the imaginary four-vector $q = +i|\omega|l$. A similar argument applies for a real q with $q^2 > 0$, except that there is no corresponding ω at which $M_{+}(\omega^2)$ attains that value, and we have two different limits, depending on the direction from which we approach the positive real axis of q^2.

[31] I apologize for repeating some of the arguments mentioned earlier.

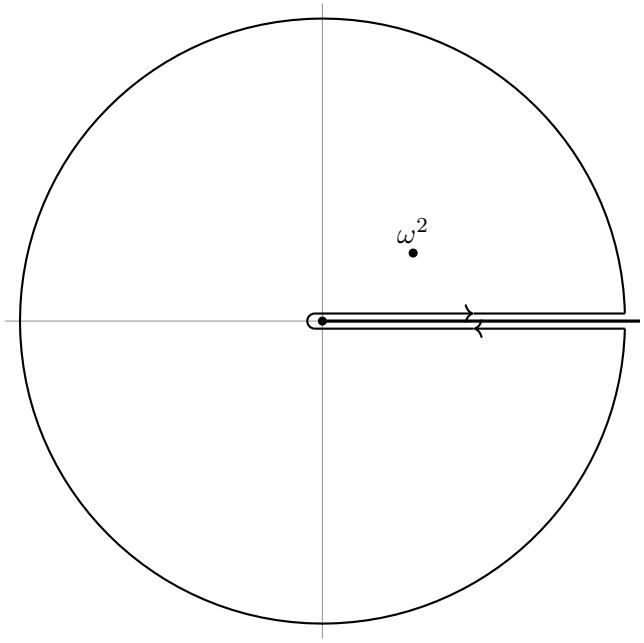


Figure 5.12: *The integration contour that can be used to obtain the dispersion relation (5.161) for the function* $M_+(\omega^2)$. The integration contour consists of four parts. First, a straight line segment that goes right above the real axis, starts at $\omega'^2 = -\delta + i\epsilon$ and ends at $\omega'^2 = +R + i\epsilon$. (Here δ, ϵ, and R are all real positive numbers, $0 < \epsilon, \delta << R$.) Second, a large circle of radius R that starts at $\omega'^2 = +R + i\epsilon$ and takes a counterclockwise circular path to $\omega'^2 = +R - i\epsilon$. Third, a straight line segment just below the real axis, from $\omega'^2 = +R - i\epsilon$ to $\omega'^2 = -\delta - i\epsilon$. Fourth, a small semicircle of radius ϵ that starts at $\omega'^2 = -\delta - i\epsilon$ and goes clockwise to $\omega'^2 = -\delta + i\epsilon$. (Of course, δ and ϵ must be chosen small enough, and R large enough, for the point $\omega'^2 = \omega^2$ to be located inside the contour of integration.) We then increase the radius R to infinity, $R \to +\infty$, at which limit the integral over the large circle vanishes. (Assuming that we made appropriate subtractions.) To obtain (5.161) we then consider the limit of $\epsilon, \delta \to 0^+$.

As we have seen, $M_+(\omega^2)$ is analytic in the whole plane of ω^2 except for the positive real axis, $\omega^2 \geq 0$. We also know that $M_+(\omega^2)$ is bounded by a polynomial for large values of $|\omega^2|$, which means that can formulate for it a dispersion relation. To keep the discussion as simple as possible, let us assume for now that $M_+ \to 0$ as $|\omega^2| \to \infty$. We will return to the general case later. Let us choose a value of ω^2 such that $\omega^2 \notin [0, +\infty)$ and integrate $M_+(\omega'^2)/(\omega'^2 - \omega^2)$, as a function of ω'^2, over the contour depicted in Fig. 5.12. As one can see in the figure, the contour start at $\omega'^2 = -\delta + i\epsilon$, where δ and ϵ are very small positive real numbers, continues along a straight line just above the real axis to $\omega'^2 = +R + i\epsilon$, where $R >> \delta, \epsilon > 0$, then takes nearly a full circle of radius R in the counterclockwise direction to the point $\omega'^2 = +R - i\epsilon$, after which it goes along a straight line just below the real axis to the point $\omega'^2 = -\delta - i\epsilon$, and finally it closes with a small semicircle of radius ϵ going clockwise to $\omega'^2 = -\delta + i\epsilon$. We then stretch the large circle to infinity. In this limit the integral over this large circle vanishes. We also take the limit of $\epsilon \to 0^+$. The integral over the small semicircle vanishes, because there is no singularity at $\omega^2 = -\delta$, and the part of the integral over the straight line segments becomes the integral of the discontinuity along the interval $[-\delta, +\infty)$. Now, on the interval $[-\delta, 0)$ the discontinuity is zero, but we formally retain this limit to make sure that any possible singularity at $\omega'^2 = 0$ is included. In this formulation this happens through the integral of the discontinuity, which might not be a regular function. Altogether, we obtain the dispersion relation[32]

$$M_+(\omega^2) = \frac{1}{2\pi i} \lim_{\delta \to 0^+} \int_{-\delta}^{\infty} d\xi^2 \frac{M_+(\xi^2 + i\epsilon) - M_+(\xi^2 - i\epsilon)}{\xi^2 - \omega^2}. \tag{5.161}$$

Now, even though M_+ itself is defined only for ω^2 that are not on the positive real semi-axis, the limits of $M_+(\xi^2 + i\epsilon)$ and $M_+(\xi^2 - i\epsilon)$ for $\epsilon \to 0^+$ do exist. Since the function $M_+(\omega^2)$ corresponds to values of ω that are in the upper half-plane, if $\omega^2 = \xi^2 + i\epsilon$ approaches the positive real axis this corresponds to the limit $\omega = +|\xi| + i\epsilon$. That is, ω approaches the positive real semi-axis, and the corresponding limit of M_+ is therefore the boundary value $M_+(\omega^2, +1)$ defined in (5.154). If, on the other hand, $\omega^2 = \xi^2 - i\epsilon$ approaches the positive real axis from below, then this corresponds to the limit $\omega = -|\xi| + i\epsilon$. That is, ω approaches the negative real semi-axis and the corresponding limit of M_+ is the boundary value $M_+(\omega^2, -1)$. In order

[32] Alternatively, we could also consider the function $\mathcal{F}^{\mu\nu}(\omega)$ defined by

$$\mathcal{F}^{\mu\nu}(\omega) = \begin{cases} \mathcal{F}^{\mu\nu}_+(\omega) & \text{if } \Im\{\omega\} > 0, \\ \mathcal{F}^{\mu\nu}_-(\omega) & \text{if } \Im\{\omega\} < 0. \end{cases}$$

This function is analytic in whole plane of ω except for the real axis. We could then consider the integration contour depicted in Fig. 5.13 to obtain a dispersion relation directly for $\mathcal{F}^{\mu\nu}(\omega)$. The discontinuity across the real axis is given by the equations (5.148), (5.153), (5.149), and (5.152).

to calculate the discontinuity $M_+(\xi^2+i\epsilon)-M_+(\xi^2-i\epsilon)$ across the positive real semi-axis, we therefore need to evaluate $M_+(\xi^2,+1)-M_+(\xi^2,-1)$. For this purpose, we can use the equation (5.153),

$$\left(q^\mu q^\nu-\eta^{\mu\nu}q^2\right)M(q^2)=M_{00}^{\mu\nu}(q,-q)$$
$$=\int d^4x\,\mathrm{e}^{iqx}\,(\Psi_0,J^\nu(0)J^\mu(x)\Psi_0)+\left(q^\mu q^\nu-\eta^{\mu\nu}q^2\right)M_+(q^2,\mathrm{sgn}(q^0)),$$

where we put $q=\pm|\xi|l$. Now, the left-hand side of this equation does not change under the inversion of its argument, $q\to -q$. This implies that

$$\xi^2\,(l^\mu l^\nu-\eta^{\mu\nu})\left[M_+(\xi^2,\mathrm{sgn}(\xi)=+1)-M_+(\xi^2,\mathrm{sgn}(\xi)=-1)\right]$$
$$=\int d^4x\,\mathrm{e}^{i(-|\xi|l)x}\,(\Psi_0,J^\nu(0)J^\mu(x)\Psi_0)$$
$$-\int d^4x\,\mathrm{e}^{i(+|\xi|l)x}\,(\Psi_0,J^\nu(0)J^\mu(x)\Psi_0)\,. \quad (5.162)$$

Using now the result (5.152),

$$\int d^4x\,\mathrm{e}^{iqx}\,(\Psi_0,J^\nu(0)J^\mu(x)\Psi_0)=2\pi\left(q^\mu q^\nu-q^2\eta^{\mu\nu}\right)\rho(q^2)\theta(-q^0),$$

we see that this discontinuity equals

$$M_+(\xi^2,\mathrm{sgn}(\xi)=+1)-M_+(\xi^2,\mathrm{sgn}(\xi)=-1)$$
$$=2\pi\rho(\xi^2)\left[\theta\left(-q^0\right)\Big|_{q=-|\xi|l}-\theta\left(-q^0\right)\Big|_{q=+|\xi|l}\right]=2\pi\rho(\xi^2). \quad (5.163)$$

The dispersion relation (5.161) can therefore be written also as

$$M_+(\omega^2)=-i\lim_{\delta\to 0^+}\int_{-\delta}^{\infty}d\xi^2\frac{\rho(\xi^2)}{\xi^2-\omega^2}. \quad (5.164)$$

This dispersion relation can be used to calculate the value of $M_+(q^2)$ for any four-vector q. The only remaining ambiguity is that to calculate the value for a real time-like q^2, we can approach the cut on the positive real axis from two directions, leading to two different results.

Now we are ready to formulate our result for the quantity $M_{00}^{\mu\nu}(q,-q)$. As we saw earlier, when we use the decomposition (5.145) to rewrite the time-ordered product that appears in $M_{00}^{\mu\nu}(q,-q)$, then for a real four-vector q we obtain the formula (5.153). Using then equations (5.152) and (5.154), we can write that formula in the form

$$M_{00}^{\mu\nu}(q,-q)=\left(q^\mu q^\nu-\eta^{\mu\nu}q^2\right)M(q^2)=2\pi\left(q^\mu q^\nu-q^2\eta^{\mu\nu}\right)\rho(q^2)\theta(-q^0)$$
$$+\left(q^\mu q^\nu-\eta^{\mu\nu}q^2\right)M_+(q^2,\mathrm{sgn}(q^0)), \quad (5.165)$$

where $M_+(q^2, \mathrm{sgn}(q^0))$ is the appropriate boundary value of $M_+(q^2)$. As we have already discussed, for real space-like values of $q^2 < 0$ this means that $M_+(q^2, \mathrm{sgn}(q^0))$ is directly the function $M_+(q^2)$ given by the dispersion relation (5.164), and does not depend on the sign of q^0. Note that this is consistent with the equation (5.165). There the quantity $M_{00}^{\mu\nu}(q, -q)$ that figures on the left-hand side is invariant under the interchange $q \to -q$ and thus cannot depend on the sign of q^0. But for q space-like the first term on the right-hand side, $2\pi \left(q^\mu q^\nu - q^2 \eta^{\mu\nu}\right) \rho(q^2) \theta(-q^0)$, vanishes, because for $q^2 < 0$ we have $\rho(q^2) = 0$. The equation (5.165) therefore implies that in the space-like region the second term must not depend on the sign of q^0. For a real time-like $q^2 > 0$, on the other hand, the matter is slightly more complicated. The left-hand side of (5.165) does not depend on the sign of q^0, but both terms on the right hand side do. When we approach the branch cut of the function $M_+(q^2)$ from above, that limit corresponds to the positive value of $q^0 = \xi l^0$. The first term on the right-hand side then vanishes, because in that limit $\theta(-q^0) = 0$. If, on the other hand, we approach the branch cut of $M_+(q^2)$ from below, then the first term on the right-hand side yields the value $2\pi \left(q^\mu q^\nu - q^2 \eta^{\mu\nu}\right) \rho(q^2)$. This is exactly what is needed to cancel the discontinuity across the branch cut of $M_+(q^2)$, which by the methods that we introduced in Sec. 2.5 can be calculated from the dispersion relation (5.164) as

$$M_+(q^2 + i\epsilon) - M_+(q^2 - i\epsilon) = -i\,(2\pi i)\,\mathrm{Res}\left(\frac{\rho(\xi^2)}{\xi^2 - q^2}; \xi^2 = q^2\right) = 2\pi\rho(q^2).$$

In other words, we have

$$\begin{aligned}\left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2 + i\epsilon) \\ = 2\pi \left(q^\mu q^\nu - q^2 \eta^{\mu\nu}\right) \rho(q^2) + \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2 - i\epsilon),\end{aligned}$$

which shows that the left-hand side of (5.165) is indeed invariant under the inversion of its argument, just as it had to be. Of course, this is not surprising, because the dispersion relation (5.164) has been formulated explicitly to yield the discontinuity which ensures this property.

To summarize, we found that if we approach the branch cut of $M_+(q^2)$ from above, we obtain the quantity which after supplying the factor

$$q^\mu q^\nu - \eta^{\mu\nu} q^2$$

corresponds directly to $M_{00}^{\mu\nu}(q, -q)$. Thus, instead of (5.165), we can simply write

$$M_{00}^{\mu\nu}(q, -q) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2 + i\epsilon). \tag{5.166}$$

Note that although this equation has been derived only for real four-vectors q, the right-hand side is an analytic function of q for all values of q^2, except

for the possible presence of singularities along the positive real axis $q^2 \geq 0$.[33] If we substitute into (5.166) our dispersion relation (5.164), we obtain

$$M_{00}^{\mu\nu}(q,-q) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) \lim_{\delta\to 0^+} \int_{-\delta}^{\infty} d\mu^2 \frac{i\rho(\mu^2)}{q^2 - \mu^2 + i\epsilon}, \tag{5.167}$$

where we renamed the variable ξ^2 to μ^2 and extracted a factor of -1 from the denominator in order to obtain the form identical to our earlier result (5.129). We could now repeat our earlier discussion, which followed the result (5.129), to show that the full photon propagator, as well as $\alpha(s)$, are real in the space-like region.

The argument that we just presented constitutes a mathematically rigorous proof of the reality of the full propagator in the space-like region. However, there is a problem with it: we assumed that $M_+(q^2) \to 0$ as $|q^2| \to \infty$, so that we could formulate an unsubtracted dispersion relation for $M_+(q^2)$. But the only thing that we really know is that $M_+(q^2)$ is bounded by a polynomial, so in general we must account for the possibility that subtractions are necessary. As we discussed in the first part of this thesis, and as we will illustrate on a simple example below, when we formulate a subtracted dispersion relation the final formula depends on a number of undetermined subtraction constants, and it is not obvious that those subtraction constants do not ruin the reality property of the propagator. (These subtraction constants are not determined by the analytic properties that are used in the formulation of dispersions relations, such as the extent of the region of analyticity or asymptotic properties of the function. It is of course possible that they might be determined or at least constrained by some other means.) As we will now show, the truth is that the reality property of the propagator is completely independent of the presence of possible subtractions. In the above proof for the unsubtracted case we actually failed to explicitly notice the reason why the propagator must be real in the space-like region. Let us now set this issue right. First, notice that nowhere in the derivation of the formula (5.166),

$$M_{00}^{\mu\nu}(q,-q) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2 + i\epsilon), \tag{5.168}$$

did we use any asymptotic properties of $M_+(q^2)$. This formula therefore holds generally, and it implies that we can determine the reality property

[33] Nothing that has been said so far does really determine where exactly the branch cut starts. We do not really know where and how many branch points are there, except that our earlier discussion suggests that there should be a branch point for each normal threshold. In any case, the answer to this question should be fully determined by the exact form of the function $\rho(q^2)$. If we start at a space-like value of $q^2 < 0$ and gradually move along the real axis to larger values of q^2, the lowest branch point should be at the point where $\rho(q^2)$ ceases to be zero. In the full standard model, which contains massless particles in physical states, this should happen already at $q^2 = 0$. On the other hand, if we calculate the quantity $M_{00}^{\mu\nu}(q,-q)$ in a theory that contains only the strong force, then the lowest branch point should be at the two-pion threshold.

of the propagator by investigating the function $M_+(q^2)$. From our earlier discussion that followed (5.129) we know that in order for the propagator to satisfy the reality condition, the same condition must be satisfied by the function $-iM_+(q^2)$.[34] That means that we need $M_+(q^2)$ to satisfy the relation

$$-iM_+(z^*) = [-iM_+(z)]^* . \tag{5.169}$$

Also notice that our earlier description of the analytic structure of $M_+(q^2)$ does still apply, even if $M_+(q^2)$ does not vanish for $|q^2| \to \infty$. In particular, it remains true that $M_+(q^2)$ is analytic in the whole complex plane of q^2 except for the positive semi-axis $q^2 \in [0, \infty)$, where its discontinuity is determined by the spectral function $\rho(q^2)$. Consider now the equation (5.154), which holds for real four-momenta q, and which has been derived independently of any assumptions about the asymptotic properties of $M_+(q^2)$ for large arguments. The equation reads

$$\int d^4x\, \mathrm{e}^{iqx} \left(\Psi_0, \theta(x^0)\, [J^\mu(x), J^\nu(0)]\, \Psi_0\right) = \left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2, \mathrm{sgn}(q^0)). \tag{5.170}$$

Recall that $M_+(q^2, +1)$ corresponds to the limit when we approach the real axis from above, and $M_+(q^2, -1)$ corresponds to the limit when we approach the real axis from below. When $q^2 < 0$, both those limits are the same, and $M_+(q^2, \mathrm{sgn}(q^0))$ does not depend on the sign of q^0. Now, it is easy to see that this equation implies the desired property (5.169) of $M_+(q^2)$. When we complex conjugate the equation (5.170), we obtain

$$\begin{aligned}
\left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2, \mathrm{sgn}(q^0))^* &= \int d^4x\, \mathrm{e}^{-iqx} \left(\Psi_0, \theta(x^0)\, [J^\mu(x), J^\nu(0)]\, \Psi_0\right)^* \\
&= \int d^4x\, \mathrm{e}^{-iqx} \left(\theta(x^0)\, [J^\mu(x), J^\nu(0)]\, \Psi_0, \Psi_0\right) \\
&= \int d^4x\, \mathrm{e}^{i(-q)x} \left(\Psi_0, \theta(x^0)\, [J^\nu(0), J^\mu(x)]\, \Psi_0\right) \\
&= -\left(q^\mu q^\nu - \eta^{\mu\nu} q^2\right) M_+(q^2, -\mathrm{sgn}(q^0)).
\end{aligned} \tag{5.171}$$

From this we see that $M_+(q^2)$ satisfies the property

$$M_+(q^2, -\mathrm{sgn}(q^0)) = -M_+(q^2, \mathrm{sgn}(q^0))^*. \tag{5.172}$$

The function $-iM_+(q^2)$ therefore satisfies (5.169), from which follows the reality property of the photon propagator. In fact, since for $q^2 < 0$ the function $M_+(q^2, \mathrm{sgn}(q^0))$ does not depend on $\mathrm{sgn}(q^0)$, we can see right away that in the space-like region $-iM_+(q^2)$ must be real.

[34] This follows directly from the equation (5.35). We are also going to see an example of this property explicitly at the end of the present section.

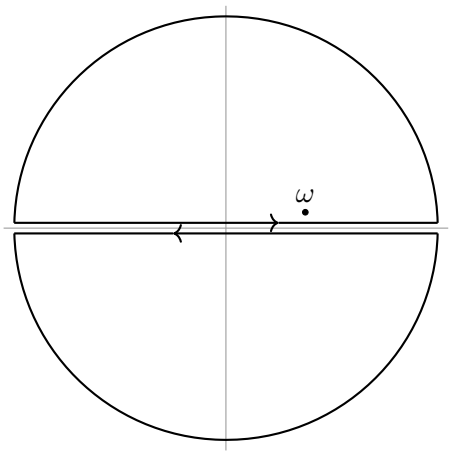


Figure 5.13: The integration contour that could be employed instead of the one depicted in Fig. 5.12 if we wanted to formulate a dispersion relation directly for the function $\mathcal{F}^{\mu\nu}(\omega)$ that is defined as $\mathcal{F}^{\mu\nu}_{+}(\omega)$ in the upper half-plane of ω and as $\mathcal{F}^{\mu\nu}_{-}(\omega)$ in the lower half-plane of ω.

There are still two remaining issues that we ought to discuss before we close this topic. The first one is the derivation of subtracted variants of the dispersion relation (5.161). As we have already mentioned, since we only know that $M_+(\omega^2)$ is bounded by a polynomial, we might need to consider a dispersion relation with one or more subtractions.[35] This possibility makes the treatment somewhat more complicated, but it does not stop us from formulating a dispersion relation of some kind. We know that there exists an integer $N > 0$, such that if $P(\omega^2)$ is a polynomial of the N-th degree then $M_+(\omega^2)/P(\omega^2) \to 0$ when $|\omega^2| \to \infty$. We can therefore formulate the dispersion relation for $M_+(\omega^2)/P(\omega^2)$ instead of $M_+(\omega^2)$ and proceed otherwise in exactly the same way as we did before. The only difference is that when we integrate

$$\frac{M_+(\omega'^2)}{P(\omega'^2)(\omega'^2 - \omega^2)} \tag{5.173}$$

over the contour depicted in Fig. 5.12, we obtain an additional residue for each zero of $P(\omega'^2)$ that lies inside of the contour of integration.

For illustration, let us consider the simple case of special interest, when $P(\omega^2) = \omega^2$. Now, since the zero of $P(\omega'^2)$ is outside of the contour of integration, namely at $\omega'^2 = 0$, we could just directly write down a dispersion relation similar to (5.164). Note, however, that $M_+(\omega'^2)/P(\omega'^2)$ contains a pole factor at $\omega'^2 = 0$, which means that the discontinuity will not be a regular function there. It is not completely straightforward to investigate

[35] This suggests that our earlier derivation of the Källén–Lehmann representation may indeed be flawed, since no subtractions appear in the equation (5.129).

the effect of this new pole directly, because there can be various singularities on the positive real axis and we also expect there to be a branch cut. But there is a trick that allows us to avoid all these issues and remain in the realm of an elementary complex analysis. Instead of ω'^2 we consider the polynomial $P(\omega'^2) = \omega'^2 + \alpha$, where α is a small and positive real number. That is, we move the zero of P slightly to the left of the point $\omega'^2 = 0$. If we choose $\alpha > \delta$ then the zero is inside of the contour of integration, and we can evaluate its contribution by the application of the basic version of the residue theorem. Afterwards, we want to take the limit $\alpha \to 0^+$, while maintaining the relation $\alpha > \delta > 0$.

Let us follow this approach a little bit further. We fix a small real α, satisfying $\alpha > \delta > 0$. According to our assumption, $M_+(\omega^2)/\left(\omega^2 + \alpha\right) \to 0$ when $|\omega^2| \to \infty$, so if we integrate (5.173), with $P(\omega'^2) = \omega'^2 + \alpha$, over the contour in Fig. 5.12 and send $R \to \infty$, we obtain

$$\lim_{\delta\to 0^+} \int_{-\delta}^{\infty} \frac{d\xi^2}{\xi^2+\alpha} \frac{2\pi\rho(\xi^2)}{\xi^2-\omega^2} = 2\pi i \frac{M_+(\omega^2)}{\omega^2+\alpha} + 2\pi i \,\mathrm{Res}\left(\frac{1}{\omega'^2+\alpha}\frac{M_+(\omega'^2)}{\omega'^2-\omega^2}; \omega'^2 = -\alpha\right). \quad (5.174)$$

Now, at the space-like $\omega'^2 = -\alpha$ the function $M_+(\omega'^2)$ is analytic, which means that the residue on the right-hand side is simply $M_+(-\alpha)/\left(-\alpha - \omega^2\right)$. Thus, if we maintain $\alpha > \delta > 0$ and take the limit $\alpha \to 0^+$, we obtain the result that can be written as

$$M_+(\omega^2) - \lim_{\alpha\to 0^+} M_+(-\alpha) = -i\omega^2 \lim_{\delta\to 0^+} \int_{-\delta}^{\infty} \frac{d\xi^2}{\xi^2} \frac{\rho(\xi^2)}{\xi^2-\omega^2}. \quad (5.175)$$

Or, if we assume some regularity properties of $M_+(\omega^2)$ that we have not really demonstrated, we may write this equation more simply as

$$M_+(\omega^2) - M_+(0) = -i\omega^2 \int_0^{\infty} \frac{d\xi^2}{\xi^2} \frac{\rho(\xi^2)}{\xi^2-\omega^2}. \quad (5.176)$$

We can now obtain an equation for $M_{00}^{\mu\nu}(q,-q)$ analogous to (5.167), but for once-subtracted dispersion relation (5.175),

$$M_{00}^{\mu\nu}(q,-q) = i\left(q^\mu q^\nu - \eta^{\mu\nu}q^2\right) \times \left[q^2 \int_0^{\infty} \frac{d\mu^2}{\mu^2} \frac{\rho(\mu^2)}{q^2-\mu^2+i\epsilon} - iM_+(0)\right]. \quad (5.177)$$

If we now use this result in the equation (5.35) for the full propagator,

just as we did in (5.130), we obtain

$$\Delta'_{\mu\nu}(q) = \frac{-\eta_{\mu\nu}}{q^2+i\epsilon} + \frac{-\eta_{\mu\nu}+\frac{q_\mu q_\nu}{q^2+i\epsilon}}{q^2+i\epsilon}\Big[q^2\int_0^\infty \frac{d\mu^2}{\mu^2}\frac{\rho(\mu^2)}{q^2-\mu^2+i\epsilon} - iM_+(0) - (Z_3-1)\Big]. \quad (5.178)$$

Now, again, since we are considering the full propagator $\Delta'_{\mu\nu}(q)$ in the physical renormalization scheme, the gauge-invariant part of the residue at $q^2=0$ must be the same as that of the free propagator. From this point of view the subtraction constant $M_+(0)$ is related to the renormalization scheme. Note, however, that it is very important that we have already demonstrated above that the constant $-iM_+(0)$ is real.[36] For $q^2=0$ the expression in the square brackets must vanish, so we obtain a single condition for the value of the constant $-iM_+(0)-(Z_3-1)$. In fact, since for $q^2=0$ we should have

$$\left[q^2\int_0^\infty \frac{d\mu^2}{\mu^2}\frac{\rho(\mu^2)}{q^2-\mu^2+i\epsilon}\right]_{q^2=0} = 0,$$

in the physical renormalization scheme we put

$$Z_3 = 1 - iM_+(0). \quad (5.179)$$

This leads to

$$\Delta'_{\mu\nu}(q) = \frac{-\eta_{\mu\nu}}{q^2+i\epsilon} + \left[-\eta_{\mu\nu}+\frac{q_\mu q_\nu}{q^2+i\epsilon}\right]\int_0^\infty \frac{d\mu^2}{\mu^2}\frac{\rho(\mu^2)}{q^2-\mu^2+i\epsilon}. \quad (5.180)$$

This brings us to the second issue, which is the derivation of the dispersion relation for the one-particle-irreducible function $\pi(q^2)$. By the same method that we used below the equation (5.130), we see that

$$\frac{1}{1-\pi(q^2)} = 1 + q^2\int_0^\infty \frac{d\mu^2}{\mu^2}\frac{\rho(\mu^2)}{q^2-\mu^2+i\epsilon}, \quad (5.181)$$

which means that the function $\pi(q^2)$ should possess the same analytic and reality properties as does the function $M_+(q^2)$. That means that we can use the same contour of integration as we did in the derivation of (5.175). Now, the perturbation theory suggests that we need to use once-subtracted dispersion relations, in order to suppress large logarithms. Thus, the dispersion relation must be of the form (5.175), except that instead of $2\pi\rho(\mu^2)$ we use

[36] Recall that $M_+(0)$ represents $\lim_{\alpha\to 0^+} M_+(-\alpha)$ combined with the assumption that $M_+(q^2)$ is continuous at $q^2=0$. What we know is that for $\alpha>0$ the function $-iM_+(-\alpha)$ is real, and this is very important, for otherwise we would not be able to achieve a physical renormalization of the propagator using a real renormalization constant Z_3.

the more general form $2i\Im\left\{\pi(\mu^2+i\epsilon)\right\}$ for the discontinuity, and that we can use the physical renormalization condition $\pi(0)=0$. Thus, we obtain the dispersion relation

$$\pi(q^2)=\frac{q^2}{\pi}\int_0^\infty \frac{d\mu^2}{\mu^2}\frac{\Im\left\{\pi(\mu^2+i\epsilon)\right\}}{\mu^2-q^2+i\epsilon}. \tag{5.182}$$

If we specialize to the theory of strong force only, we obtain the dispersion relation (5.56), that we have employed earlier in Sec. 5.2.

5.6 Utilizing the U&A model

In the remaining part of this section I will introduce some of the results that I have obtained in collaboration with my colleagues, and which we plan to publish in the article [65]. This work was originally inspired by the recent paper of Stamen *et al.* [66], in which the authors constructed models of electromagnetic form factors of charged and neutral kaons and used them to calculate several quantities of interest. These quantities were the corresponding charge radii, the correction to the Dashen's theorem, the lowest order contribution of kaons to the muon's magnetic anomaly a_μ, and the contribution of light-by-light kaon boxes to a_μ. Models constructed by Stamen *et al.* in [66] are appropriate as approximate phenomenological models, but there is a reasonable perspective from which they do not appear very attractive. Their models for isoscalar and isovector components of kaon form factors are in their core each different, and this difference is due to practical convenience rather than underlying physics. A more dangerous aspect of their approach is that they combine two mutually exclusive approaches within a single model, namely the dispersion theoretical approach based on the Omnès function [67] and the VMD model.[37] While this does not by any means invalidate their work, we thought it could be worthwhile to redo their analysis using the more universal approach of the U&A model. Furthermore, in addition to kaon form factors we also calculated the same four quantities from the form factor of the charged pion.

Another specificity of our approach was that we decided to fit our model on exclusively time-like experimental values. This was despite the fact that

[37] These two approaches each describe the same thing from a different point of view. The Omnès function approach describes the form factor in terms of its discontinuity, as determined by relevant asymptotic physical states, while the VMD approach approximates the form factor using poles of vector meson resonances. A resonance is not an asymptotic state, and corresponds instead to a particular superposition of such states. If one combines those two approaches, one encounters various troubles related to possible double counting. Now, this is not say that it is impossible to combine them correctly, but they are in any case exclusive: once we account for some contribution by one approach we should not do so again by the other approach. From a practical point of view, however, and especially in an approximate model, it is possible to keep these issues under control, and it might make a good sense to use an approach such as the one presented in [66].

three out of four above-mentioned quantities which we aimed to calculate depended on space-like values of form factors. Our reasoning was that it is quite challenging to measure space-like values of the form factors of the pion and kaons. For the charged pion and the charged kaon the main experimental method is the scattering of those mesons on atomic electrons, which is complicated by the fact that the mass of the target electron is much smaller than the mass of the projectile: some 280 times smaller in the case of $\pi\, e$ scattering and nearly thousand times smaller for the $K\, e$ scattering. Furthermore, the interpretation of such measurements is also complicated by the presence of the external field of the target nucleus. (And the measurement of space-like values of the form factor of the neutral kaon is even more difficult.) One could observe these difficulties in early measurements of the charge radius of the pion, which all yielded quite disparate values.[38] For the charged kaon form factor there are only two available sets of space-like data [71, 72], and they are of relatively low precision. We therefore decided to estimate our model parameters on selected time-like data and use the analytic continuation inherent the U&A model to predict space-like values of form factors. These predicted values of the form factor of the charged kaon are depicted here in Fig. 5.14, where we also display the available experimental data. Unlike kaon form factors, which we constructed anew, we adopted the model of the charged pion form factor from an earlier work [73], on which I did not collaborate, but which also fitted the model on only time-like data.

At the level of precision appropriate for this work it is possible to operate under the approximation that isospin is a precise symmetry of strong interactions.[39] This allows us to relate the form factor $F_{K^\pm}$ of the charged kaon to the form factor F_{K^0} of the neutral kaon. Under isotopic spin transformations the operator of the electromagnetic current transforms as the sum of a component that transforms as a scalar and a component that transforms as a vector. That is, we can write $J = J^s + J^v$, where J^s is the isoscalar and J^v the isovector component. Each form factor can be accordingly decomposed into an isoscalar and an isovector component. This, combined with the fact that the pair K^+ and K^0 forms the isospin doublet $\{\Psi_{K^+}, \Psi_{K^0}\}$, and K^- and $\bar{K}^0$ the doublet $\{-\Psi_{\bar{K}^0}, \Psi_{K^-}\}$, can then be used to show that the isoscalar and isovector parts of the charged kaon form factor and the isoscalar and isovector parts of the neutral kaon form factor are closely related. As one can verify, both form factors can be expressed in terms of

[38] In 1977 Adylov *et al.* [68] published the value $\langle r^2\rangle_{\pi^\pm} = 0.61 \pm 0.16$ fm^2 and Dally *et al.* [69] the value $\langle r^2\rangle_{\pi^\pm} = 0.31 \pm 0.04$ fm^2. Later, in 1982, Dally *et al.* [70] published the value $\langle r^2\rangle_{\pi^\pm} = 0.439 \pm 0.030$ fm^2.

[39] There is an exception that during the fitting of the model for the pion form factor one must account in cross section data for the contribution of the isospin violating electromagnetic decay $\omega(782) \to \pi^+\pi^-$. Details can be found, for instance, in [73], from which we adopted the pion model.

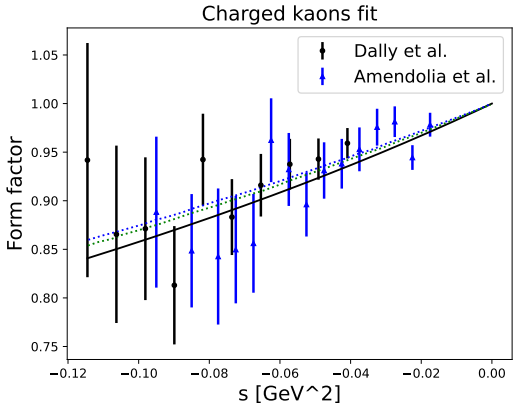


Figure 5.14: Predictions of the model (the black solid line) plotted against available charged kaons data from the space-like region: Dally *et al.* [71], Amendolia *et al.* [72]. These data were not used to determine model parameters. The blue dotted line is the "neutral fit" and the green dotted line the "charged fit" from Stamen *et al.* [66]. (Taken from [65].)

their common isoscalar part F_K^s and their common isovector part F_K^v as

$$F_{K^\pm}(s) = F_K^s(s) + F_K^v(s), \tag{5.183}$$

$$F_{K^0}(s) = F_K^s(s) - F_K^v(s). \tag{5.184}$$

Our assumption that isospin is a precise symmetry of the strong force implies that isoscalar form factor F_K^s couples only to isoscalar resonances, and the isovector form factor F_K^v only to isovector resonances. Unfortunately, there is no clear-cut way to decide which resonances to include in the model and which not. In principle, all resonances that carry the correct quantum numbers should be included, but available data are not sufficient to determined all their parameters. For any given form factor, some resonances contribute more than other and, furthermore, if a resonance pole is located far away from the region of interest then its effect is strongly suppressed. For this reason we usually want to decide carefully which resonances to include and which to omit. Unless there is some specific reason to do otherwise, the natural rule of thumb is to include as few resonances as possible while maintaining a good description of data. In practice, one usually makes the decision based on the inspection of data which will be used to estimate the model parameters, and by experimenting with various reasonably looking choices. In our work [65] we decided to include all three resonances ϕ, ϕ', ϕ'' into our model of F_K^s, and leave all their parameters free. This is because kaons carry non-zero strangeness so one expects them to couple strongly the ϕ resonances. We noticed in data an indication of a possible contribution of the $\omega''(1650)$ resonance, which we therefore also included, but with its mass

fixed at 1.67 GeV and only its width left as a free parameter. We excluded the resonance $\omega'(1420)$, whose contribution was not at all apparent in the data. Furthermore, in order to correctly describe data at the low energy it is in general necessary to include the ground state resonances. Thus, we also included the $\omega(782)$-resonance, with its parameters fixed at the values from [74]. Thus, the model of F_K^s used in our work [65] contains resonances ω, ϕ, ϕ', ω'' and ϕ'', with the mass and width of ω and the mass of ω'' fixed. In the model of the isovector part of the kaon form factor we considered all three relevant ρ-mesons, and fixed the parameters of the ground resonance ρ at their values from [74].

With this selection of resonances, one can then proceed and construct the model according to the rules described in the chapter 4. For the isoscalar component $F_K^s(s)$ the function $\hat{F}_K^s(W)$, corresponding to the function $\hat{F}_{\mathrm{U\&A}}$ of (4.26) and (4.27), is given by

$$\hat{F}_K^s(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^2 \sum_{i=\omega,\phi,\phi',\omega'',\phi''} \frac{f_{iKK}}{f_i} \prod_{p\in\{\pm W_i,\pm W_i^*\}} \frac{W_N - p}{W - p}. \quad (5.185)$$

The relation between s and W is given by the equations (4.9), (4.10), and (4.11), with $s_0 = 9m_\pi^2$ and s_{in} determined by the fit of data. The form factor $F_K^s(s)$ is then given by the function $\hat{F}_K^s(W(q(s)))$. The form (5.185) corresponds to the case when $\sqrt{s_{\mathrm{in}}} < m_i$ for all included resonances, which is the condition that happened to hold for our fit. For the isovector part the corresponding function $\hat{F}_K^v$ was of the form

$$\hat{F}_K^v(W) = \left(\frac{1-W^2}{1-W_N^2}\right)^2 \left[\frac{f_{\rho KK}}{f_\rho} \prod_{p\in\{W_\rho,W_\rho^*,1/W_\rho,1/W_\rho^*\}} \frac{W_N - p}{W - p} + \sum_{i=\rho',\rho''} \frac{f_{iKK}}{f_i} \prod_{p\in\{\pm W_i,\pm W_i^*\}} \frac{W_N - p}{W - p}\right]. \quad (5.186)$$

Here the map from the four-sheeted Riemann surface of s to W is given again by (4.9), (4.10), (4.11), but this time with $s_0 = 4m_\pi^2$ and a different value of s_{in}, again determined by the fit. The form above is for the case when $m_\rho < \sqrt{s_{\mathrm{in}}} < m_{\rho'}, m_{\rho''}$. We adopted the model of $F_{\pi^\pm}$ from [73].

Among the available measurements of the process $e^+e^- \to K^0\bar{K}^0$ we chose to use data measured by collaborations CMD-2 [75], CMD-3 [76], BaBar [77], and BESIII [78]. For the process $e^+e^- \to K^+K^-$ we considered data by CMD-3 [79], BaBar [80, 81], and BESIII [82]. Publications [75, 80, 78, 82] presented "undressed" data; that is, data with the vacuum polarization effect on the virtual photon line removed. The remaining publications presented only "dressed" cross sections with the vacuum polarization effects included. Since for our purposes "undressed" form factors were appropriate, we removed the vacuum polarization effects from [79, 76, 77, 81]

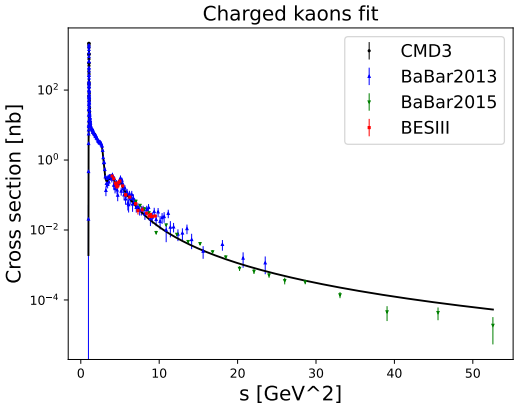


Figure 5.15: The U&A model (5.185), (5.186) with parameters from Table 5.1 displayed against all the experimental charged kaon data that were used to fit the model. The y-axis is depicted in the logarithmic scale. Labels correspond to the following sources: CMD3 [79], BaBar2013 [80], BaBar2015 [81], BESIII [82]. (Taken from [65].)

before further analysis. Furthermore, because our model assumes the conservation of the isospin symmetry, it was important to remove all final state radiation effects from data of charged kaons. We did that using the final state radiation factor $1 + \frac{\alpha}{\pi}\eta$. The reader can find a discussion of final state radiation corrections, as well as the explicit form of the η function, in [45] or [46]. The original reference is [83], with a misprint corrected in [84]. The parameters of our fit are displayed in the Table 5.1. Plots of the fit, displayed against the data, can be found in figures Fig. 5.15, Fig. 5.16, Fig. 5.17, and Fig. 5.18.

After we constructed the models of F_K^s and F_K^v, defined by (5.185) and (5.186), respectively, and estimated their parameters as depicted in Table 5.1, we used those models (together with the charged pion model of [73]) to calculate several quantities of interest. Below, I am going to briefly present our results.

5.6.1 Charge radii

One of the quantities that we calculated were the charge radii of the charged pion and of the charged and neutral kaons. The mean squared charge radius of a spin zero particle is defined in terms of the derivative of its electromagnetic form factor,

$$\langle r^2 \rangle = 6 \frac{dF(s)}{ds}\bigg|_{s=0}. \tag{5.187}$$

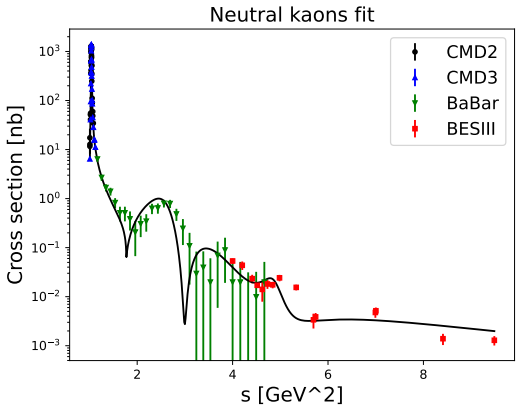


Figure 5.16: The U&A model (5.185), (5.186) with parameters from Table 5.1 displayed against all the experimental neutral kaon data that were used to fit the model. The y-axis is depicted in the logarithmic scale. Labels correspond to the following sources: CMD2 [75], CMD3 [76], BaBar [77], BESIII [78]. (Taken from [65].)

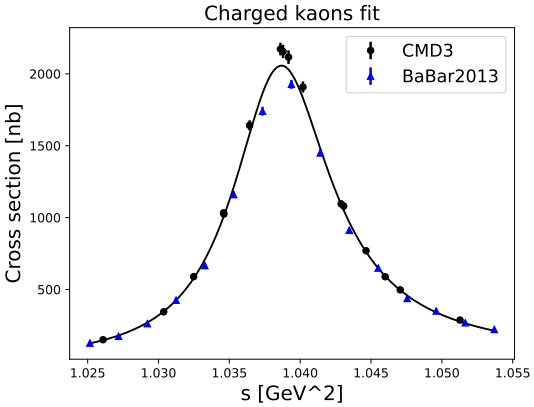


Figure 5.17: The model displayed against the data of charged kaons near the ϕ-resonance peak. (Taken from [65].)

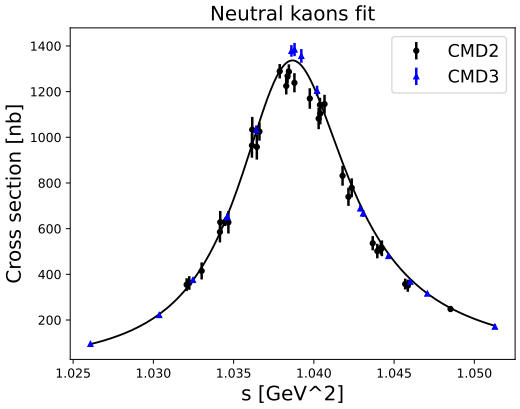


Figure 5.18: The model displayed against the data of neutral kaons near the ϕ-resonance peak. (Taken from [65].)

Table 5.1: Parameters obtained in [65] for the model of the kaon form factors described the equations (5.185) and (5.186).

name	value	name	value
s_0^s	0.1753190 GeV^2 (fixed)	s_0^v	0.0779196 GeV^2 (fixed)
s_{in}^s	0.6125 ± 0.0001 GeV^2	s_{in}^v	1.7720 ± 0.0035 GeV^2
$f_{\omega KK}/f_\omega$	0.1977 ± 0.0013		
m_ω	0.78266 GeV (fixed)	Γ_ω	0.00868 GeV (fixed)
$f_{\omega'' KK}/f_{\omega''}$	0.1580 ± 0.0034		
$m_{\omega''}$	1.67 GeV (fixed)	$\Gamma_{\omega''}$	0.329 ± 0.017 GeV
$f_{\phi KK}/f_\phi$	0.32427 ± 0.00059		
m_ϕ	1.0190367 ± 0.0000079 GeV	Γ_ϕ	0.004139 ± 0.000015 GeV
$f_{\phi' KK}/f_{\phi'}$	-0.1833 ± 0.0033		
$m_{\phi'}$	1.6396 ± 0.0019 GeV	$\Gamma_{\phi'}$	0.2322 ± 0.0079 GeV
$m_{\phi''}$	2.2063 ± 0.0025 GeV	$\Gamma_{\phi''}$	0.1000 ± 0.0051 GeV
$f_{\rho KK}/f_\rho$	0.5308 ± 0.0013		
m_ρ	0.75823 GeV (fixed)	Γ_ρ	0.14456 GeV (fixed)
$f_{\rho' KK}/f_{\rho'}$	-0.1308 ± 0.0080		
$m_{\rho'}$	1.467 ± 0.020 GeV	$\Gamma_{\rho'}$	0.796 ± 0.020 GeV
$m_{\rho''}$	1.8640 ± 0.0093 GeV	$\Gamma_{\rho''}$	0.481 ± 0.018 GeV

We have obtained the following values [65]

$$\langle r^2\rangle_{\pi^\pm} = 0.428 \pm 0.010 \text{ fm}^2, \tag{5.188}$$

$$\langle r^2\rangle_{K^0} = -0.123 \pm 0.010 \text{ fm}^2, \tag{5.189}$$

$$\langle r^2\rangle_{K^\pm} = 0.403 \pm 0.005 \text{ fm}^2. \tag{5.190}$$

The uncertainties of these quantities, as well as of other results presented below, were estimated with the help of a Monte Carlo method. According to the published statistical and systematic errors of the experimental data, we generated a number of datasets, and fitted each of those with our model. From each of such a fit we then calculated all the quantities of interest, and used those results to estimate the errors.

Note that while the value (5.188) for the charge radius of the pion is in agreement with its PDG average value [74]

$$\langle r^2\rangle_{\pi^\pm}|_{[74]} = 0.434 \pm 0.005 \text{ fm}^2, \tag{5.191}$$

the values for kaon charge radii obtained here have significantly larger magnitude than their respective PDG averages

$$\langle r^2\rangle_{K^\pm}|_{[74]} = 0.314 \pm 0.035 \text{ fm}^2. \tag{5.192}$$

and

$$\langle r^2\rangle_{K^0}|_{[74]} = -0.077 \pm 0.010 \text{ fm}^2, \tag{5.193}$$

As we discuss in the article, a possible explanation for this discrepancy could be that while the PDG averages (5.192) and (5.193) include measurements carried out directly in the space-like region, we avoided all space-like data, and relied solely on the analytic continuation from the time-like region.

5.6.2 Correction to the Dashen's theorem

When Dashen investigated the consequences of the approximate chiral SU(3)× SU(3) symmetry of strong interactions [85], he predicted that the difference between the electromagnetic contributions to the squared mass of the charged pion and the squared mass of the neutral pion

$$(\Delta m_\pi^2)_{\text{EM}} = (m_{\pi^\pm}^2)_{\text{EM}} - (m_{\pi^0}^2)_{\text{EM}}, \tag{5.194}$$

should be the same as the corresponding difference $(\Delta m_K^2)_{\text{EM}} = (m_{K^\pm}^2)_{\text{EM}} - (m_{K^0}^2)_{\text{EM}}$ for the charged and neutral kaons. The corrections are, however, quite large. We define the quantity ϵ, which describes the size of those corrections, as

$$\epsilon = \frac{(\Delta m_K^2)_{\text{EM}}}{(\Delta m_\pi^2)_{\text{EM}}} - 1. \tag{5.195}$$

To evaluate these corrections, Stamen *et al.* [66] employed the Cottingham formula [86, 87]. They determined the electromagnetic masses using the relation

$$\left(m_p^2\right)_{\rm EM} = \frac{\alpha}{8\pi}\int_0^\infty ds\,[F_p(-s)]^2\left(4W + \frac{s}{m_p^2}(W-1)\right) \tag{5.196}$$

for the electromagnetic mass of a spin 0 particle p. Here, $F_p(-s)$ is the particle's electromagnetic form factor evaluated at the space-like value $-s$, and $W = \sqrt{1-4m_p^2/s}$. When we substituted into this formula our models of form factors we obtained the values [65]

$$(m_{K^\pm}^2)_{\rm EM} = (1.99 \pm 0.03) \times 10^{-3}{\rm GeV}^2, \tag{5.197}$$
$$(m_{K^0}^2)_{\rm EM} = (4.3 \pm 1.1) \times 10^{-5}{\rm GeV}^2, \tag{5.198}$$
$$(m_{\pi^\pm}^2)_{\rm EM} = (1.183 \pm 0.023) \times 10^{-3}{\rm GeV}^2, \tag{5.199}$$

for the electromagnetic masses. Note that $(m_{\pi^0}^2)_{\rm EM}$ vanishes, because the form factor of the neutral pion is identically zero. From these results we then calculated [65]

$$(\Delta m_K^2)_{\rm EM} = (1.95 \pm 0.03) \times 10^{-3}{\rm GeV}^2, \tag{5.200}$$
$$(\Delta m_\pi^2)_{\rm EM} = (1.183 \pm 0.023) \times 10^{-3}{\rm GeV}^2, \tag{5.201}$$
$$\epsilon = 0.65 \pm 0.06, \tag{5.202}$$

where we estimated the error of ϵ by adding the errors linearly.

5.6.3 Pion and kaon box contributions to the magnetic anomaly of the muon

Another application that we considered was the calculation of the contribution to the muon's magnetic anomaly a_μ of the pion and kaon box diagrams for hadronic light-by-light scattering. This corresponds to the diagram Fig. 5.8 of the subsection 5.2. We implemented the procedure described in [88, 89] and obtained the following values [65]

$$a_\mu^{K^\pm-{\rm box}} = (-0.451 \pm 0.022) \times 10^{-11}, \tag{5.203}$$
$$a_\mu^{K^0-{\rm box}} = (-4.6 \pm 1.4) \times 10^{-15}, \tag{5.204}$$
$$a_\mu^{\pi^\pm-{\rm box}} = (-16.0 \pm 0.2) \times 10^{-11}. \tag{5.205}$$

The values for the charged kaon and the charged pion are in agreement with the corresponding values obtained in [66] and [89]

$$a_\mu^{K^\pm-{\rm box}}|_{[66]} = (-0.484 \pm 0.011) \times 10^{-11}, \tag{5.206}$$
$$a_\mu^{\pi^\pm-{\rm box}}|_{[89]} = (-15.9 \pm 0.2) \times 10^{-11}. \tag{5.207}$$

The contribution of neutral kaons that we have obtained is an order of magnitude larger than that of [66]

$$a_\mu^{K^0-\text{box}}|_{[66]} = (-0.50 \pm 0.45) \times 10^{-15}. \tag{5.208}$$

It is not clear to us how to interpret this discrepancy. The relative errors of both results do, however, indicate that this quantity may be difficult to determine from available data.

5.6.4 Lowest order contributions to the muon's magnetic anomaly

We also considered evaluating the lowest order contributions of the processes $e^+e^- \to K^+K^-$ and $e^+e^- \to K_S K_L$ to the value of a_μ, corresponding to the diagram Fig. 5.2. In the end we decided not to include these results in [65], because unlike the previous three sets of quantities, this one is calculated from time-like values of the form factors. Nevertheless, I will mention those results at least here. Using the equations (5.49), (5.50), and (5.51), with the integration region spanning from the thresholds to $s = (1.05\ \text{GeV})^2$, and using our models of form factors to calculate the corresponding values of $R(s)$, I obtained the results

$$a_\mu^{\text{HVP,LO}}\left[e^+e^- \to K^+K^-, \sqrt{s} \le 1.05\ \text{GeV}\right] = (188.7 \pm 3.2) \times 10^{-11}, \tag{5.209}$$

and

$$a_\mu^{\text{HVP,LO}}\left[e^+e^- \to K_S K_L, \sqrt{s} \le 1.05\ \text{GeV}\right] = (115.4 \pm 2.2) \times 10^{-11}. \tag{5.210}$$

Chapter 6

Other applications of the Unitary and Analytic model

6.1 Damped oscillatory structures in the effective form factor of proton

Another topic on which I collaborated was the study of the periodic oscillatory structures that were originally found in the data for the effective form factor of the proton by Andrea Bianconi and Egle Tomasi-Gustafsson [30]. These authors, as well several other researchers, have speculated that those oscillations might represent interference patterns resulting from rescattering processes between the newly-formed hadrons. (See, e.g., [30, 90].) We investigated this issue from a different point of view and published our results in several papers. I collaborated on two of those, [31, 32]. Because both these manuscripts have been already published and all the details can be found there, I will limit myself here to only a short introduction of this topic and a brief discussion of the possible interpretation of our findings.

As we saw in Chapter 3, the interaction of a spin-1/2 particle such as the proton with a single virtual photon can be described in terms of two form factors. If we decide to use the electric form factor $G_E^{(p)}(s)$ and the magnetic form factor $G_M^{(p)}(s)$, then the total cross section $\sigma_{\text{tot}}\left(e^+e^- \to p\bar{p}\right)$ for the annihilation of the electron–positron pair into the proton–anti-proton pair can be written as

$$\sigma_{\text{tot}}\left(e^+e^- \to p\bar{p}\right) = \frac{4\pi\alpha^2}{3s} C_p \beta_p(s) \left[\left| G_M^{(p)}(s) \right|^2 + \frac{2m_p^2}{s} \left| G_E^{(p)}(s) \right|^2 \right], \quad (6.1)$$

where $\beta_p(s) = \sqrt{1 - \frac{4m_p^2}{s}}$ is the velocity of the proton in the center-of-mass frame and C_p is the Sommerfeld–Gamow–Sakharov Coulomb enhancement factor, which can be used to account for the final state radiation effects in the

low energy limit.[1] Note that in this equation the form factors correspond to quantities that result from strong interactions only. This is why we needed to include the effects of the final state electromagnetic radiation explicitly. Furthermore, the cross section $\sigma_{\rm tot}\left(e^+e^-\to p\bar{p}\right)$ is "undressed". This means that the vacuum-polarization effects on the propagator of the intermediate photon should be removed. One can see this from the fact that on the right-hand side of (6.1) there appears the factor $4\pi\alpha^2/3s$ instead of $4\pi\alpha(s)^2/3s$.[2]

At present it is still quite challenging to carry out measurements that would allow us to extract the values of the electric and magnetic form factors individually. For this reason it is often useful to work with the concept of the so-called *effective form factor* $F_{\rm eff}^{(p)}(s)$ of the proton, defined by the equation

$$\sigma_{\rm tot}\left(e^+e^-\to p\bar{p}\right)=\frac{4\pi\alpha^2}{3s}C_p\beta_p(s)\left[1+\frac{2m_p^2}{s}\right]\left|F_{\rm eff}^{(p)}(s)\right|^2. \tag{6.2}$$

This equation would follow from (6.1) if the condition

$$\left|G_E^{(p)}(s)\right|=\left|G_M^{(p)}(s)\right|=\left|F_{\rm eff}^{(p)}(s)\right| \tag{6.3}$$

was satisfied for all $s\geq 0$. (Or at least for all s above the threshold for the production of two protons.) In general, however, the electric and magnetic form factors $G_E^{(p)}(s)$ and $G_M^{(p)}(s)$ are equal only at $s=0$. Once can therefore interpret the formula (6.2) as an approximation in which we extend this condition to non-zero values of the center-of-mass energy s.

Bianconi and Tomasi-Gustafsson [30] fitted the experimental data for the effective form factor of the proton by the function[3]

$$\left|F_{\rm eff}^{(p)}(q^2)\right|=\frac{\mathcal{A}}{(1+\frac{q^2}{m_a^2})\left[1-\frac{q^2}{0.71\ {\rm GeV}^2}\right]^2}, \tag{6.4}$$

with the free parameters estimated as as $\mathcal{A}=7.7\ {\rm GeV}^{-4}$ and $m_a^2=14.8\ {\rm GeV}^2$. Afterwards, they subtracted the fit from the data, and noticed that when one inspects the ensuing residues as a function of the magnitude p of the three-momentum of either one of the final particles in a frame in which the other particle is at rest, then the residues display a regular damped oscillatory pattern. This is depicted in Fig. 6.1, taken from [30]. The authors suggested that these oscillations might result from interference effects of re-scattering processes between outgoing hadrons.

[1] Of course, we could just as well use the more precise final state radiation factor $1+\frac{\alpha}{\pi}\eta$ that we briefly mentioned in the section 5.6. In the region where experimental data are available the difference between these two alternatives is not substantial.

[2] According to our result (5.89), $4\pi\alpha^2/3s$ is the lowest order total cross section for the process $e^+e^-\to\mu^+\mu^-$, in the approximation that all those leptons are massless.

[3] The authors claim that they also investigated what happens if one uses several other simple functions, and their results appeared to be quite robust with respect to the choice of the precise form of the model.

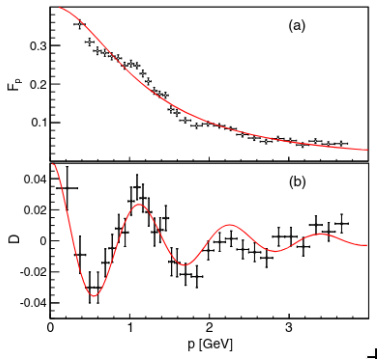


Figure 6.1: The figure (a) depicts time-like values of the effective form factor of proton that were investigated in [30]. The red line represents the fit (6.4). The figure (b) displays the oscillatory pattern that appears when the fit depicted in the figure (a) is subtracted from the data. The variable p on the x-axis is the magnitude of the three-momentum of one of the protons in a frame in which the other proton is at rest. (Taken from [30].)

Our group explored a different kind of interpretation of the observed pattern. We investigated the possibility that these patterns were simply an artifact that appeared as a result of the authors of [30] using an inadequate phenomenological model (6.4) to fit the data. One expects the data to include various "bulges" due to the presence of vector meson resonances, and it is, of course, not surprising that if one fits such data with a simple monotonous function such as (6.4) then oscillations appear in the residues. (Although it is not clear why such oscillations would show a periodic regular pattern.) We therefore described the data with the U&A model to see if the oscillations appear also in the residues of this more appropriate model. The residues of the U&A model showed no discernible oscillations [32]. Furthermore, since the hypothesis presented in [30] was not in any obvious way dependent on the fact that the final particles were protons, one would expect the same phenomenon to appear also for other hadrons. For this reason we also investigated form factors of pions and kaons. (I collaborated on the study of the kaon form factors [31].) When we described the data with a simple model, that was similar to the model (6.4) used by Bianconi and Tomasi-Gustafsson in [30], we indeed observed an oscillatory pattern in the residues to the fit. (But that is not surprising, since in that case the pattern was clearly just a by-product of an inadequate model.) But when we described the same data by the U&A model, those oscillations seemed to have disappeared. For details, please see the paper [31]. This result together with the further exploration presented in [32] suggest that the patterns described in [30] might indeed represent just an accidental side effect of describing data with an overly simplistic model such as (6.4). Let me mention that

we were not the only group to argue for this kind of interpretation. For instance, the authors of [91] have presented a very similar argument.

So where does this leave us? In my opinion it is safe to say that our work — as well as the work of several other authors including [91] — has demonstrated that the oscillatory patterns detected by Bianconi and Tomasi-Gustafsson in [30] can be adequately described by fitting the data by more complex models. It must be emphasized that these more complex models — such as the U&A model used by us or the extended VMD model employed in [91] — stand on quite solid physical foundations. For instance, as we saw in the chapter 4 the U&A model is based on a number of general physical requirements, including the analyticity on the physical sheet, a low-energy approximation of the correct branch structure, the Hermitian analyticity, or the presence of resonance poles. Compared to the U&A model, the model (6.4) of [30] seems somewhat arbitrary, and motivated chiefly by its simplicity and an approximate agreement with data. From this point of view one could claim that because the oscillations disappear when a physically more adequate U&A model is applied, those oscillations do not have any physical significance beyond what is already captured by the U&A model. Continuing this line of reasoning, the simplest interpretation would then be that the oscillations result from the presence of resonances. If that was true, it still would be interesting, however, to study the alleged periodicity in those oscillations — are they really periodic and if yes, what does it mean for resonances? On the other hand, taking a step back, one could also argue that the U&A model has a much larger number of free parameters, and our work was nothing but an exercise in curve-fitting. To answer this objection one must demonstrate that the U&A model (or any other model) that manages to fit those oscillatory patterns does so using a physically admissible values of its free parameters. Unfortunately, many of those parameters, in particular the masses and widths of resonances, are currently not known with a sufficient precision. There is also another, somewhat more fundamental research direction, that one could follow to explore this topic. That is the question of whether these two competing explanations, one of the oscillations as an interference pattern from re-scattering processes and the other one of the oscillations as the effect of resonances, do really represent two distinct phenomena. For instance, it is known that the near-threshold resonances can be understood (at least in a toy model) in terms of a constructive interference when the low-energy quark–anti-quark pair periodically bounces back to the location of where it was created. (See, e.g., Chapter 4 of [7] and references cited therein.) So perhaps the fact that the oscillatory patterns can be described by the U&A model or the extended VMD model does not in any sense imply that the interpretation suggested by Bianconi and Tomasi-Gustafsson is wrong. However, it is quite likely that to answer this question we need first to achieve a better understanding of the process of hadronization itself.

6.2 The ratio $R = \phi \to K^+K^-/K^0_LK^0_S$

In this last section I will briefly present another topic on which I worked during my doctoral study. Just as in the previous section, I will only briefly introduce the topic and discuss some of its significance or interpretation, but I will skip most of the technical details of the work itself, which can be found in the already published article [33].

Our work [33] was concerned with the ratio of the probability that the $\phi(1020)$ resonance decays into a pair of charged kaons to the probability that it decays into a pair of neutral kaons. The processes $\phi \to K^+K^-$ and $\phi \to K^0\bar{K}^0$ are two dominant decay modes of the resonance $\phi(1020)$ and we denote the ratio of their respective probabilities by R. This number can be expressed also in terms of their respective branching ratios,

$$R = \frac{\mathrm{BR}(\phi \to K^+K^-)}{\mathrm{BR}(\phi \to K^0\bar{K}^0)}. \tag{6.5}$$

The reason why we were interested in this quantity is that there has been a long-lasting discrepancy between its measured value and its theoretical predictions. From the 2022 Review of Particle Physics [74] we can calculate the following value for the experimental world average of the ratio R

$$R_{\mathrm{exp}} = 1.4484 \pm 0.0226. \tag{6.6}$$

Efforts to predict the value of R theoretically date at least as far back as 1969 when Eugene Cremmer and Michel Gourdin obtained the value $R \approx 1.60$, based mainly on the considerations of the isotopic spin symmetry and radiative corrections for the final state charged kaons [92]. In 1974 Harmut Pilkuhn used a different method, the penetration factor model, to estimate the value of R to lie between 1.52 and 1.61 [93]. The most recent studies of this topic are the works of Bramon *et al.* [94] and Flores-Baéz and López Castro [95]. These authors reexamined and expanded upon the earlier approach of Cremmer and Gourdin. Their investigations led to the theoretical estimate that R is 1.59 or higher, which is far above the experimental average (6.6).

For this reason we decided to estimate the value of R from the available experimental data for the total cross section of the processes $e^+e^- \to K^+K^-$ and $e^+e^- \to K^0\bar{K}^-$. We hoped that this would provide a further insight into the experimental side of the discrepancy. At the heart of our method lies the assumption that the Breit–Wigner formula for the resonant contribution to the S-matrix holds accurately enough for the $\phi(1020)$ resonance. More specifically, we assumed that the ratio R was the same as the ratio of the corresponding cross sections at the ϕ-resonance peak. Under this assumption we could then estimate the value of R from experimental cross section data by the following program:

1. In the cross section data for the processes $e^+e^- \to K^+K^-$ and $e^+e^- \to K^0\bar{K}^-$ try to isolate the contribution of the processes $e^+e^- \to \phi(1020) \to K^+K^-$ and $e^+e^- \to \phi(1020) \to K^+K^-$ from the contribution of the non-resonant background.

2. In these data for the resonant scattering determine the position — that is the energy — of the resonance peak. In other words, estimate the (mean) mass of the resonance.

3. Estimate the cross sections for both processes at the resonance peak, and calculate their ratio.

Considering that the available data are of limited density and precision, none of these steps is possible without some kind of a model to describe the cross sections. We employed the framework of the U&A model, which is described here in Chapter 4, to describe simultaneously both the data for the process $e^+e^- \to K^+K^-$ and the data for the process $e^+e^- \to K^0\bar{K}^0$. Considering the fact that the U&A model describes the form factor as a sum of contributions from individual resonances, it allows one to isolate from the non-resonant background the contribution of the $\phi(1020)$ resonance. If one fits the data and then focuses only on the part corresponding to the resonance $\phi(1020)$, then at least in principle that should correspond to considering directly the relevant processes $e^+e^- \to \phi \to K^+K^-$ and $e^+e^- \to \phi \to K^0\bar{K}^0$.

To describe the data we used a model similar to the one described in the section 5.6. To more easily relate the data of the charged and neutral kaons we used the assumption of the isotopic spin symmetry, which allowed us to decompose the corresponding form factors into their common isoscalar and isovector components F_K^s and F_K^v,

$$F_{K^\pm} = F_K^s + F_K^v, \tag{6.7}$$

$$F_{K^0} = F_K^s - F_K^v. \tag{6.8}$$

Thus both form factors depended on the same set of parameters: the masses, widths and thresholds were common. The coupling constants were also related. From the equations (6.7) and (6.8) it follows that the couplings to isoscalar resonances were the same in $F_{K^\pm}$ as in F_{K^0}, while the couplings to isovector resonances had opposite signs in $F_{K^\pm}$ than in F_{K^0}. However, there was an important exception that we had to make. If we really used the model as just described then the resulting R ratio would be completely fixed, since the term corresponding to the $\phi(1020)$ resonance would be exactly same in $F_{K^\pm}$ as in F_{K^0}. The R ratio would then be determined by the relative size of the kinematic spaces of K^+K^- and $K^0\bar{K}^0$ at the resonance peak and final state radiation corrections, and we would have inevitably reproduced the theoretical estimate of [92]. In order to actually determine the value of

R from data it is vital that we released the isospin-symmetry constraint on the coupling constants of the $\phi(1020)$ resonance to the charged and neutral kaons. In other words, instead of the equations (6.7) and (6.8) we used

$$F_{K^\pm} = F^s_{K^\pm} + F^v_K, \tag{6.9}$$

$$F_{K^0} = F^s_{K^0} - F^v_K, \tag{6.10}$$

where $F^s_{K^\pm}$ and $F^s_{K^0}$ were just like F^s_K of (6.7), (6.8) but with the difference that the functions $F^s_{K^\pm}$ and $F^s_{K^0}$ had two independent coupling constants to the $\phi(1020)$ resonance, instead of the single common one of F^s_K. (We then also had to release the isospin-symmetry constraint on one more pair of isoscalar coupling constants, in order to preserve the normalization condition. But that is just a technical detail.) In this way we estimated two coupling constants, one for the coupling of $\phi(1020)$ to K^+K^- and the other one for the coupling of $\phi(1020)$ to $K^0\bar{K}^0$. From their ratio, after we accounted for the kinematic space differences and final state radiative corrections, we then calculated the value of R which corresponded to the experimental data. The detail of our analysis can be found in [33], and it proceeded roughly along the following steps:

1. Since our models of form factors and any use of the isotopic spin symmetry required that we consider form factors as determined only by the strong force, we first needed to remove from experimental data all electromagnetic radiative corrections. At least approximately, we needed to remove from the data the effect of the vacuum polarization on the propagator of the intermediate photon, and take care of any initial or final state radiative corrections.

2. We then fitted the data with our U&A model for the form factors. Our fits are depicted in Fig. 6.2 and Fig. 6.3. This allowed us to remove the non-resonant background and focus directly on the ratio of the coupling constants of the resonance $\phi(1020)$ to K^+K^- and to $K^0\bar{K}^0$.

3. To extract from data the ratio R we considered the above mentioned ratio of coupling constant, added the factor corresponding to the ratio of volumes of the kinematic spaces of the charged and neutral kaons at the resonance peak, and added the radiative corrections corresponding to virtual and soft photons that could be exchanged between or emitted by the final state charged kaons K^+K^-.

After all these step we obtained the value

$$R = 1.553 \pm 0.040. \tag{6.11}$$

Interestingly, this was more than 2σ above the PDG value (6.6) and practically in agreement with the theoretical estimate of [94].

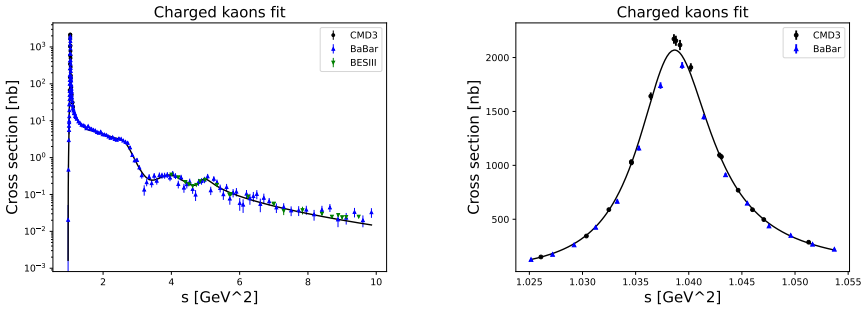


Figure 6.2: Our fit depicted against the charged kaon cross section data. Note that the most important part is the region near the ϕ-resonance peak, which is depicted on the right. (Taken from [33].)

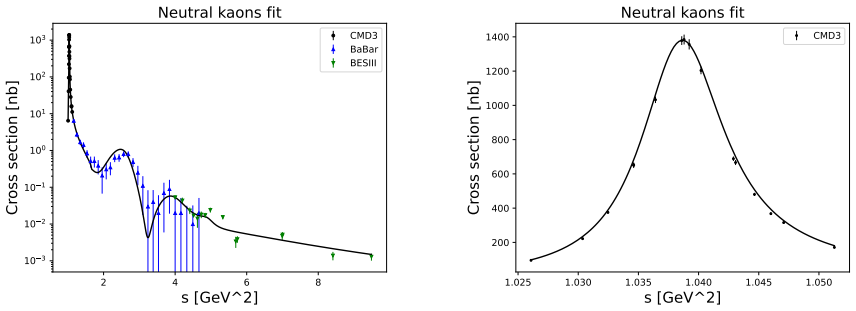


Figure 6.3: Our fit depicted against the neutral kaon cross section data. Note that the most important part is the region near the ϕ-resonance peak. (Taken from [33].)

Before closing this section, let us briefly comment on the broader significance of this issue. First, it is important to note that similar discrepancies have been observed also for other resonances that are located near new-flavor thresholds. They were observed for the ratios of charged to neutral yields of the processes $\psi(3770) \to DD$ and $\Upsilon(4S) \to BB$. Several authors have attempted to explain these discrepancies. For instance, Dubynskiy *et al.* [96] and Voloshin [97] considered the effects of strong-scattering phases in the isoscalar and isovector channels on the corresponding ratios. It turns out that these phases modify the final state radiation effects as well as the mass difference effect on the corresponding cross section. These strong-scattering phases have so far not been calculated, but they are in principle capable of explaining the observed anomalies. Interestingly, there have been also several suggestions that for near-threshold resonances one must appropriately modify the usual formulas in order to obtain correct cross sections. For instance, Fischbach *et al.* [98] considered what appears to be a resonance-appropriate ansatz to study the Fermi's Golden Rule. Their ansatz allowed them to move beyond the first order of perturbation theory and use the unitarity condition to derive an integral equation for the decay rate. They found that if either the size of the kinematic space or the values of matrix elements of the interaction between the resonance and the asymptotic states vary rapidly near the resonance peak, then usual formulas for decay rate must be adjusted. Another interesting instance is the work of Ishikawa *et al.* [99], who studied these processes using a formalism that had been previously developed by some of the authors to describe the scattering of wave-packets. In their work they also argued that the decay rate formula should be adjusted. While [98] considered applications of their modifications only to the anomaly of the $\phi(1020)$ resonance, [99] considered also the anomaly in the decay modes of the resonance $\Upsilon(4S)$. Both of these works claim to have resolved the discrepancies. At the same time, however, both works seem to approach the problem in a different way, and their supposed solutions appear to be unrelated. It seems that a further study is required to reach a better understanding of what, if any, adjustments are necessary and how are the various suggestions related. In relation to this issue, especially in the context of our work [33], let me add two speculative remarks:

- In the derivation of the Breit–Wigner formula it is assumed that the non-resonant background varies negligibly over the region where the resonance dominates. In other words, the formula is derived under the approximation that near the resonance the energy dependence of the scattering amplitude is completely described by the contribution of the resonance pole, possibly up to an additive constant.[4] However, when

[4]This constant represents the energy-independent contribution of the non-resonant background. This is important if the non-resonant contribution to the phase shift is large, in which case the resonance manifests itself as a sharp dip in the cross section, instead of

the resonance is located near a normal threshold, as is the case of the resonances $\phi(1020)$, $\psi(3770)$, and $\Upsilon(4S)$, then this assumption might not be justified. The amplitude might be significantly influenced by the presence of the threshold, and perhaps even the shadow poles that can be accessed by encircling the threshold and crossing the branch cut from another direction. Thus, it is quite possible that some modification of the usual formulas might be necessary. This could be an interesting project for a further study.

- In our work [33] we built our analysis on the assumption that the Breit–Wigner formula is precise, and we obtained from data results that were consistent with theoretical predictions. (Note that the theoretical predictions also ignore the presence of the threshold.) Thus it might be possible that our result (6.11) is the correct value of the ratio of the resonance peak heights, and that this ratio has been correctly predicted by various theoretical works [92, 93, 94, 95], but that at the same time this ratio does not correspond to the ratio of probabilities of the corresponding decay modes, if the Breit–Wigner formula does not hold precisely enough. Then the PDG value (6.6) could also be correct, in the sense of the ratio of corresponding probabilities, although not in the sense of the ratio of peak heights.

Notwithstanding these speculative remarks, recently there has been published a preprint [100] by the BESIII collaboration that studied the anomaly in the ratio R of the $\phi(1020)$ resonance. Interestingly, they measured the value

$$R_{[100]} = 1.675 \pm 0.093, \tag{6.12}$$

which is even higher than our estimate (6.11). This indicates the possibility that the PDG average value (6.6) might not be reliable.

a more common peak. (See, e.g., [9].)

Chapter 7

Conclusion

In Part I of this thesis I presented a gentle introduction to the general topic of the analyticity in quantum field theory. Part II was then dedicated mainly to the introduction of the U&A model. After a review of some basic properties of electromagnetic form factors in Chapter 3, we were ready to introduce the U&A model in Chapter 4. The treatment presented in that chapter introduces the model from a somewhat novel perspective and includes several improvements over the treatments that can be found in older literature.

Then in Part III, the chapter 5 opened with a general discussion of the running of the fine structure constant. We also saw how the hadronic contribution to its imaginary part can be related to the total cross section for the electron–positron annihilation into hadrons. At the present experimental precision one can still work under the approximation of a single intermediate photon, which means that this cross section is directly related to the corresponding electromagnetic form factor, and can be described with the U&A model. Furthermore, using the dispersion relation for $\Delta\alpha(s)$ it is possible to express the value of $\alpha(s)$ in the space-like region in terms of its imaginary part in the time-like region. This means that the U&A model can be also used to describe the behavior of the hadronic contribution to $\alpha(s)$ in the space-like region.

In the same chapter, we also saw how the running of the fine structure constant enters into the evaluation of the leading contribution to the magnetic anomaly of the muon, and how one can choose to calculate this contribution either from the imaginary part of $\alpha(s)$ in the time-like region, or from $\alpha(s)$ itself in the space-like region. In collaboration with my colleagues I have explored some aspects of this topic in the paper [29]. Chapter 5 also contains a brief description of yet unpublished work [65], in which we employed electromagnetic form factors of the charged pion and of both the charged and neutral kaons to calculate various quantities of interest, including the corresponding hadronic light-by-light contribution to the magnetic anomaly of the muon. In this thesis, I also presented my result for the lead-

ing hadronic contribution to the magnetic anomaly of the muon. And lastly, the chapter 5 also contained a detailed proof of the reality of the fine structure constant in the space-like region,[1] and a discussion of its significance.

The thesis was then concluded in Chapter 6, where other published work on which I collaborated during my doctoral study was briefly presented. The section 6.1 introduced the topic of damped oscillatory structures, which were first observed in the effective form factor of the proton by A. Bianconi and E. Tomasi-Gustafsson in [30], and which were studied by us in the papers [31, 32]. Section 6.2 was dedicated to the problem of the discrepancy between the theoretical predictions and experimental estimates of the ratio R of the probabilities of the decay of the resonance $\phi(1020)$ into a pair of charged kaons and into a pair of neutral kaons. I presented there our paper [33], where we used the U&A model to obtain an estimate of the ratio R from the total cross section data for the electron–positron annihilation into the corresponding kaon pairs.

[1] This proof is my own, but I think that all the techniques employed in it have been known since late 1960s, which makes me believe that the result itself and some kind of its proof must have been known to the community at least since that period.

Bibliography

[1] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne. *The Analytic S-Matrix*. Cambridge University Press, 1962.

[2] S. Mandelstam. Dispersion relations in strong-coupling physics. *Rep. Prog. Phys.*, 25(1):99–162, 1962.

[3] H. M. Nussenzveig. *Causality and dispersion relations*, volume 95. Academic Press, New York, London, 1972.

[4] S. Mizera. Physics of the analytic S-matrix. *Physics Reports*, 1047:1–92, 2024. Physics of the Analytic S-Matrix.

[5] R. F. Streater and A. S. Wightman. *PCT, Spin and Statistics, and All That*. Princeton University Press, 1989.

[6] J. Collins. A new and complete proof of the Landau condition for pinch singularities of Feynman graphs and other integrals. 7 2020.

[7] J. Collins. *Foundations of Perturbative QCD*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2023.

[8] E. C. Titchmarsh. *Introduction to the Theory of Fourier Integrals*. Oxford University Press, London, second edition, 1948.

[9] S. Weinberg. *The Quantum Theory of Fields*, volume 1 Foundations. Cambridge University Press, 2005.

[10] J. Collins. States in QFT and a derivation of perturbation theory compatible with Haag's theorem.

[11] J. Collins. A new approach to the LSZ reduction formula. 2019.

[12] R. Haag. On quantum field theories. *Kong. Dan. Vid. Sel. Mat. Fys. Med.*, 29N12:1–37, 1955.

[13] H. Lehmann, K. Symanzik, and W. Zimmermann. Zur formulierung quantisierter feldtheorien. *Nuovo Cim.*, 1(1):205–225, 1955.

[14] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[15] S. Coleman and R. E. Norton. Singularities in the physical region. *Nuovo Cim.*, 38:438–442, 1965.

[16] P. V. Landshoff. Poles and thresholds and unstable particles. *Nuovo Cim.*, 28(1):123–131, 1963.

[17] S. Gasiorowicz. *Elementary Particle Physics*. John Wiley & Sons, 1966.

[18] S. Dubnička and A. Z. Dubničková. Analyticity in a phenomenology of electro-weak structure of hadrons. *Acta Phys. Slovaca*, 60(1):1–153, 2010.

[19] S. Dubnička, I. Furdik, and V. A. Meshcheryakov. An Analytic Formula Describing All Data on the Pion Form-Factor. 11 1976.

[20] J.J. Sakurai. Theory of strong interactions. *Annals of Physics*, 11(1):1–48, 1960.

[21] R. E. Peierls. Complex eigenvalues in scattering theory. *Proc. R. Soc. Lond. A*, 253(1272):16–36, 1959.

[22] J. B. Conway. *Functions of One Complex Variable I*. Graduate Texts in Mathematics. Springer New York, NY, 1994.

[23] R. Penrose and W. Rindler. *Spinors and Space-Time*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1984.

[24] S. J. Brodsky and G. R. Farrar. Scaling laws at large transverse momentum. *Phys. Rev. Lett.*, 31:1153–1156, Oct 1973.

[25] G. Peter Lepage and S. J. Brodsky. Exclusive processes in quantum chromodynamics: Evolution equations for hadronic wavefunctions and the form factors of mesons. *Phys. Lett. B*, 87(4):359–365, 1979.

[26] A. V. Efremov and A. V. Radyushkin. Factorization and asymptotic behaviour of pion form factor in QCD. *Phys. Lett. B*, 94(2):245–250, 1980.

[27] C. Adamuščin, A.-Z. Dubničková, S. Dubnička, R. Pekárik, and P. Weisenpacher. Asymptotic conditions for the electromagnetic form factors of hadrons represented by the VMD model. *Eur. Phys. J. C*, 28:115–118, 2003.

[28] S. Dubnička, A. Z. Dubničková, and P. Weisenpacher. General solution of asymptotic conditions for electromagnetic form factors of hadrons represented by VMD model, 2002.

[29] A. Z. Dubničková, S. Dubnička, L. Holka, and A. Liptaj. Advantages of a novel approach to the numerical evaluation of the muon $g-2$ anomaly. *Phys. Part. Nucl.*, 53, 2022.

[30] A. Bianconi and E. Tomasi-Gustafsson. Periodic interference structures in the timelike proton form factor. *Phys. Rev. Lett.*, 114:232301, Jun 2015.

[31] S. Dubnička, A. Z. Dubničková, L. Holka, and A. Liptaj. Study of damped oscillating structures from charged and neutral K-meson electromagnetic form factors data. *Eur. Phys. J. A*, 59:190, 2023.

[32] E. Bartoš, S. Dubnička, A. Z. Dubničková, L. Holka, and A. Liptaj. Comprehensive insight into regular damped oscillatory structures from effective electromagnetic form factor data of some mesons and nucleons. *Dynamics*, 4(4):804–829, 2024.

[33] S. Dubnička, A. Z. Dubničková, L. Holka, and A. Liptaj. What is the real value of the ratio $R = \phi \to K^+K^-/K_L^0K_S^0$? *Phys. Rev. D*, 110:054019, Sep 2024.

[34] J. Collins. *Renormalization: An Introduction to Renormalization, the Renormalization Group and the Operator-Product Expansion.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2023.

[35] S. Weinberg. *The Quantum Theory of Fields*, volume 2 Modern Applications. Cambridge University Press, 2005.

[36] P.A.M. Dirac. The quantum theory of the electron. *Proc. R. Soc. Lond. A*, 117:610–624, 1928.

[37] Julian Schwinger. On quantum-electrodynamics and the magnetic moment of the electron. *Phys. Rev.*, 73:416–417, Feb 1948.

[38] C. Bouchiat and L. Michel. La résonance dans la diffusion méson π-méson π et le moment magnétique anormal du méson μ. *J. Phys. Radium*, 22(2):121, 1961.

[39] L. Durand. Pionic contributions to the magnetic moment of the muon. *Phys. Rev.*, 128:441–448, Oct 1962.

[40] S. J. Brodsky and E. de Rafael. Suggested boson-lepton pair couplings and the anomalous magnetic moment of the muon. *Phys. Rev.*, 168:1620–1622, Apr 1968.

[41] M. Gourdin and E. De Rafael. Hadronic contributions to the muon g-factor. *Nucl. Phys. B*, 10(4):667–674, 1969.

[42] F. Jegerlehner. *The Anomalous Magnetic Moment of the Muon.* Springer Cham, 2018.

[43] J. Calmet, S. Narison, M. Perrottet, and E. De Rafael. Higher order hadronic corrections to the anomalous magnetic moment of the muon. *Phys. Lett. B*, 61(3):283–286, 1976.

[44] J. Calmet, S. Narison, M. Perrottet, and E. de Rafael. The anomalous magnetic moment of the muon: A review of the theoretical contributions. *Rev. Mod. Phys.*, 49:21–29, 1977.

[45] A. Hoefer, J. Gluza, and F. Jegerlehner. Pion pair production with higher order radiative corrections in low energy e^+e^- collisions. *Eur. Phys. J. C*, 24(1):51–69, 2002.

[46] Kirill Melnikov. On the theoretical uncertainties in the muon anomalous magnetic moment. *Int. J. Mod. Phys. A*, 16(28):4591–4612, 2001.

[47] Sz. Borsanyi, Z. Fodor, J. N. Guenther, C. Hoelbling, S. D. Katz, L. Lellouch, T. Lippert, K. Miura, L. Parato, K. K. Szabo, F. Stokes, B. C. Toth, Cs. Torok, and L. Varnhorst. Leading hadronic contribution to the muon magnetic moment from lattice qcd. *Nature*, 593:51–55, 2021.

[48] A. Gérardin, W. E. A. Verplanke, G. Wang, Z. Fodor, J. N. Guenther, L. Lellouch, K. K. Szabo, and L. Varnhorst. Lattice calculation of the π^0, η and $\eta^{'}$ transition form factors and the hadronic light-by-light contribution to the muon $g-2$. *Phys. Rev. D*, 111:054511, Mar 2025.

[49] G. Abbiendi, C. M. C. Calame, U. Marconi, C. Matteuzzi, G. Montagna, O. Nicrosini, M. Passera, F. Piccinini, R. Tenchini, L. Trentadue, and G. Venanzoni. Measuring the leading hadronic contribution to the muon $g-2$ via μe scattering. *Eur. Phys. J. C*, 77, 2017.

[50] C.M. Carloni Calame, M. Passera, L. Trentadue, and G. Venanzoni. A new approach to evaluate the leading hadronic corrections to the muon $g-2$. *Phys. Lett. B*, 746:325–329, 2015.

[51] A. B. Arbuzov, D. Haidt, C. Matteuzzi, M. Paganoni, and L. Trentadue. The running of the electromagnetic coupling α in small angle Bhabha scattering. *Eur. Phys. J. C*, 34:267–275, 2004.

[52] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio. Complete tenth-order QED contribution to the muon $g-2$. *Phys. Rev. Lett.*, 109:111808, Sep 2012.

[53] A. Kurz, T. Liu, P. Marquard, and M. Steinhauser. Hadronic contribution to the muon anomalous magnetic moment to next-to-next-to-leading order. *Phys. Lett. B*, 734:144–147, 2014.

[54] J. Prades, E. de Rafael, and A. Vainshtein. *The Hadronic Light-by-Light Scattering Contribution to the Muon and Electron Anomalous Magnetic Moments*, pages 303–317. 2009.

[55] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[56] B. Abi et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.

[57] S. Dubnička, A. Z. Dubničková, and A. Liptaj. Application of the U&A model for evaluation of the kaonic and other LO hadronic contributions to the muon magnetic anomaly with reduced uncertainty. *Nucl. Phys. B Proc. Suppl.*, 219–220:271–275, 2011. The Proceedings of the 5th Joint International Hadron Structure '11 Conference.

[58] A. Anastasi et al. Measurement of the running of the fine structure constant below 1 GeV with the KLOE detector. *Physics Letters B*, 767:485–492, 2017.

[59] H. Czyż, A. Grzelińska, J.H. Kühn, and G. Rodrigo. Electron–positron annihilation into three pions and the radiative return. *Eur. Phys. J. C*, 47:617–624, 2006.

[60] G Abbiendi. Letter of Intent: the MUonE project, 2019.

[61] G. Ballerini et al. A feasibility test run for the MUonE project. *Nucl. Instrum. Meth. A*, 936:636–637, 2019.

[62] G. Abbiendi et al. A study of muon-electron elastic scattering in a test beam. *JINST*, 16(06), 2021.

[63] G Abbiendi. Status of the MUonE experiment. *Phys. Scr.*, 97(5):054007, 2022.

[64] G. Abbiendi et al. Measurement of the running of the QED coupling in small-angle Bhabha scattering at LEP. *Eur. Phys. J. C*, 45:1–21, 2006.

[65] E. Bartoš, S. Dubnička, A. Z. Dubničková, and L. Holka. Kaon and pion electromagnetic form factors and a prediction of some physical quantities from their space-like behaviour.

[66] D. Stamen, D. Hariharan, M. Hoferichter, B. Kubis, and P. Stoffer. Kaon electromagnetic form factors in dispersion theory. *Eur. Phys. J. C*, 82(5):432, 2022.

[67] E. Omnès. On the solution of certain singular integral equations of quantum field theory. *Nuovo Cimento*, 8(2):316–326, 1958.

[68] G. T. Adylov et al. A measurement of the electromagnetic size of the pion from direct elastic pion scattering data at 50 GeV/c. *Nucl. Phys. B*, 128(3):461–505, 1977.

[69] E. B. Dally et al. Direct measurement of the π^- form factor. *Phys. Rev. Lett.*, 39:1176–1179, Nov 1977.

[70] E. B. Dally, J. M. Hauptman, J. Kubic, D. H. Stork, A. B. Watson, Z. Guzik, et al. Elastic-scattering measurement of the negative-pion radius. *Phys. Rev. Lett.*, 48:375–378, Feb 1982.

[71] E. B. Dally, J. M. Hauptman, J. Kubic, D. H. Stork, A. B. Watson, Z. Guzik, et al. Direct Measurement of the Negative-Kaon Form Factor. *Phys. Rev. Lett.*, 45:232–235, 1980.

[72] S. R. Amendolia et al. A Measurement of the Kaon Charge Radius. *Phys. Lett. B*, 178:435–440, 1986.

[73] E. Bartoš, S. Dubnička, and A. Z. Dubničková. Search for damped oscillating structures from charged pion electromagnetic form factor data. *Dynamics*, 3:137–151, 2023.

[74] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[75] R. R. Akhmetshin et al. Measurement of ϕ meson parameters in $K_L^0 K_S^0$ decay mode with cmd-2. *Phys. Lett. B*, 466(2):385–391, 1999.

[76] E. A. Kozyrev et al. Study of the process $e^+e^- \to K_S^0 K_L^0$ in the center-of-mass energy range 1004–1060 MeV with the CMD-3 detector at the VEPP-2000 e^+e^- collider. *Phys. Lett. B*, 760:314–319, 2016.

[77] J. P. Lees et al. Cross sections for the reactions $e^+e^- \to K_S^0 K_L^0$, $K_S^0 K_L^0 \pi^+\pi^-$, $K_S^0 K_S^0 \pi^+\pi^-$, and $K_S^0 K_S^0 K^+ K^-$ from events with initial-state radiation. *Phys. Rev. D*, 89(9):092002, 2014.

[78] Medina Ablikim et al. Cross section measurement of $e^+e^- \to K_S^0 K_L^0$ at $\sqrt{s} = 2.00 - 3.08\ GeV$. *Phys. Rev. D*, 104(9):092014, 2021.

[79] E. A. Kozyrev et al. Study of the process $e^+e^- \to K^+K^-$ in the center-of-mass energy range 1010–1060~MeV with the CMD-3 detector. *Phys. Lett. B*, 779:64–71, 2018.

[80] J. P. Lees et al. Precision measurement of the $e^+e^- \to K^+K^-(\gamma)$ cross section with the initial-state radiation method at BABAR. *Phys. Rev. D*, 88(3):032013, 2013.

[81] J. P. Lees et al. Study of the $e^+e^- \to K^+K^-$ reaction in the energy range from 2.6 to 8.0 GeV. *Phys. Rev. D*, 92(7):072008, 2015.

[82] M. Ablikim et al. Measurement of $e^+e^- \to K^+K^-$ cross section at $\sqrt{s} = 2.00 - 3.08$ GeV. *Phys. Rev. D*, 99(3):032001, 2019.

[83] J. S. Schwinger. *Particles, Sources, And Fields*, volume 3. Perseus Books, Reading, Massachusetts, 1998.

[84] Manuel Drees and Ken-Ichi Hikasa. Scalar top production in e^+e^- annihilation. *Phys. Lett. B*, 252(1):127–134, 1990.

[85] Roger Dashen. Chiral SU(3) $\bigotimes$ SU(3) as a symmetry of the strong interactions. *Phys. Rev.*, 183:1245–1260, Jul 1969.

[86] W.N Cottingham. The neutron proton mass difference and electron scattering experiments. *Ann. Phys.*, 25(3):424–432, 1963.

[87] Haim Harari. Superconvergent dispersion relations and electromagnetic mass differences. *Phys. Rev. Lett.*, 17:1303–1306, Dec 1966.

[88] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer. Dispersion relation for hadronic light-by-light scattering: theoretical foundations. *J. High Energy Phys.*, 2015(9):74, 2015.

[89] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer. Dispersion relation for hadronic light-by-light scattering: two-pion contributions. *J. High Energy Phys.*, 2017(4):161, 2017.

[90] Egle Tomasi-Gustafsson and Simone Pacetti. Interpretation of recent form factor data in terms of an advanced representation of baryons in space and time. *Phys. Rev. C*, 106:035203, Sep 2022.

[91] Bing Yan, Cheng Chen, Xia Li, and Ju-Jun Xie. Understanding oscillating features of the timelike nucleon electromagnetic form factors within the extending vector meson dominance model. *Phys. Rev. D*, 109:036033, Feb 2024.

[92] E. Cremmer and M. Gourdin. Radiative corrections to $\phi \to K\bar{K}$ decay. *Nucl. Phys. B*, 9(3):451–459, 1969.

[93] H. Pilkuhn. Coulomb corrections to hadron widths. *Nucl. Phys. B*, 82(3):365–369, 1974.

[94] A. Bramon, R. Escribano, J.L. Lucio M., and G. Pancheri. The ratio $\Phi \to K^+K^-/K^0\bar{K}^0$. *Phys. Lett. B*, 486(3–4):406–413, August 2000.

[95] F. V. Flores-Baéz and G. López Castro. Structure-dependent radiative corrections to $\phi \to K^+K^-/K_LK_S$ decays. *Phys. Rev. D*, 78:077301, Oct 2008.

[96] S. Dubynskiy, A. Le Yaouanc, L. Oliver, J.-C. Raynal, and M. B. Voloshin. Isospin breaking in the yield of heavy meson pairs in e^+e^- annihilation near threshold. *Phys. Rev. D*, 75:113001, Jun 2007.

[97] M. B. Voloshin. Charged-to-neutral meson yield ratio at $\psi(3770)$ and $\Upsilon(4S)$ revisited. *Phys. Rev. D*, 97:114023, Jun 2018.

[98] E. Fischbach, A.W. Overhauser, and B. Woodahl. Corrections to Fermi's golden rule in $\phi \to K\bar{K}$ decays. *Phys. Lett. B*, 526(3):355–359, 2002.

[99] K. Ishikawa, O. Jinnouchi, K. Nishiwaki, and K. Oda. Wave-packet effects: a solution for isospin anomalies in vector-meson decay. *Eur. Phys. J. C*, 83(10):978, 2023.

[100] M. Ablikim et al. Study of $\phi \to K\bar{K}$ and $K^0_S - K^0_L$ asymmetry in the amplitude analysis of $D^+_s \to K^0_SK^0_L\pi^+$ decay, 2025.