Chapter 34

Causality and Analyticity

The relationship between causality and analyticity; its implications for the propagation of electromagnetic waves in dispersive media; dispersion relations and their relevance in particle physics.

Last Update: 8/4/12

Consider some device which is supplied with an input as a function of time, $f(t)$, and as a result outputs some signal $g(t)$. Suppose the input at time t' is sustained for a short period dt' and that the output at some later time t is proportional to the input and hence of the form,

$$dg(t) = h(t,t') \cdot f(t')dt' \tag{1}$$

Here the function h describes the operation of the device. Assuming that its operation has no explicit time dependence, i.e., no built-in clock which changes its behaviour, then the relationship between input and output in (1) will depend only upon the time interval $t - t'$ and not upon the absolute time t. Hence we can replace $h(t,t')$ with a function of a single variable, $H(t-t')$. This may be called the *response function* of the device.

The assumption of causality, namely that cause precedes effect, implies an output at time t is obtained only due to inputs at or before t. Hence (1) applies only for $t' \le t$, or equivalently that $h(t,t') = 0$ for $t' > t$, or $H(\tau) = 0$ for $\tau = t - t' < 0$. Integrating (1) gives the total signal at time t,

$$g(t) = \int_{-\infty}^{t} h(t,t') f(t') dt' = \int_{0}^{\infty} H(\tau) f(t-\tau) d\tau \tag{2}$$

Surprisingly non-trivial results follow from looking at the frequency spectrum of the response function. This is defined in the usual manner via the Fourier transform,

$$\tilde{H}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(t) e^{i\omega t} dt \qquad \text{and its inverse,} \qquad H(t) = \int_{-\infty}^{\infty} \tilde{H}(\omega) e^{-i\omega t} d\omega \tag{3}$$

But the causality requirement $H(\tau) = 0$ for $\tau < 0$ reduces this to,

$$\text{Causality:} \qquad \tilde{H}(\omega) = \frac{1}{2\pi}\int_{0}^{\infty} H(t) e^{i\omega t} dt \tag{4}$$

Recall that this defines the Fourier transformed function $\tilde{H}(\omega)$ for complex values of the frequency, not just real values. Provided that $H(t)$ is always finite, it follows from (4) that $\tilde{H}(\omega)$ is also well defined and finite for any ω with positive imaginary part. This is because the exponential in (4) consists of an oscillatory part of unit magnitude times $\exp\{-\omega_i t\}$, where $\omega = \omega_r + i\omega_i$, and the latter factor ensures unconditional convergence of the t-integral in (4) in the upper half plane of complex ω, i.e., when $\omega_i > 0$. This means that the response function $\tilde{H}(\omega)$ is analytic in the upper half of the complex frequency plane. This conclusion, known as Titchmarsh's theorem, Titchmarsh (1948), extends to the real frequency axis only if we add the condition that $H(\tau \to \infty) \to 0$ which is sometimes true and sometimes not true in physical applications.

Consequently, Cauchy's theorem allows us to write,

$$\tilde{H}(\omega) = \frac{1}{2\pi i} \oint_{\Gamma} \frac{\tilde{H}(\omega')}{\omega' - \omega} d\omega' \qquad (5)$$

where ω is any point on the upper half plane and Γ is any closed contour containing ω and lying entirely within the upper half plane or on the real axis if analyticity extends that far. Let us assume the latter is the case for now. Hence we can take Γ to lie along the whole of the real line and to be closed by a semi-circle at infinity in the upper half plane. But the integral around the infinite semi-circle will be zero because, as noted above, the integrand of (4) for $\tilde{H}(\omega')$ contains the factor $\exp\{-\omega_i' t\}$ which drives $\tilde{H}(\omega')$ strongly to zero for $\omega_i' \to \infty$. Now taking the limit that ω approaches the real axis, we need to deform the contour infinitesimally so that it departs from the real line along a vanishingly small semi-circle around, and below, the pole at $\omega' = \omega$. The integral around this little semi-circle accounts for half the residue of the pole whilst the rest of the integral becomes the principal part, defined by,

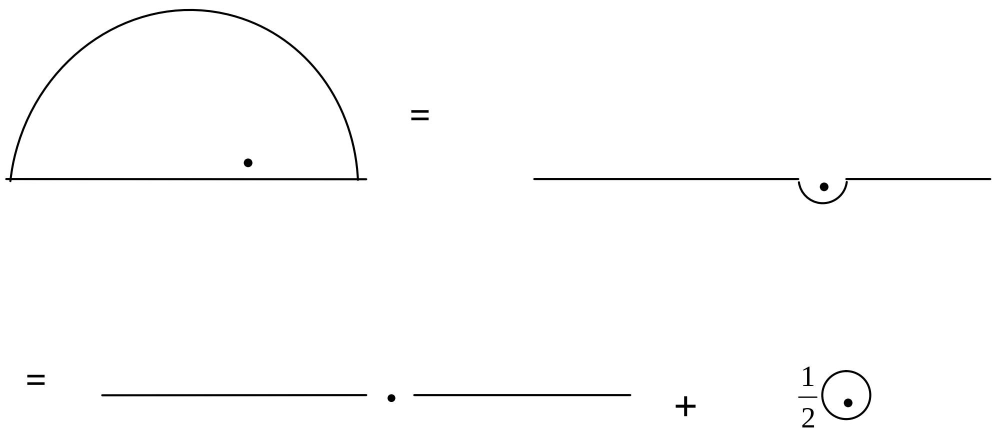
$$P \int_{-\infty}^{+\infty} \frac{\tilde{H}(\omega')}{\omega' - \omega} d\omega' \equiv LIM_{\varepsilon \to 0} \left(\int_{-\infty}^{\omega - \varepsilon} + \int_{\omega + \varepsilon}^{+\infty} \right) \frac{\tilde{H}(\omega')}{\omega' - \omega} d\omega' \qquad (6)$$

where ω is now real. Hence, (5) becomes,

$$\tilde{H}(\omega) = \frac{1}{\pi i} P \int_{-\infty}^{+\infty} \frac{\tilde{H}(\omega')}{\omega' - \omega} d\omega' \qquad (7)$$

In terms of the contours, the transformation of (5) into (7) can be understood as illustrated by Figure 1.

Figure 1: The Integration Contours in the Complex ω-Plane Leading to Equ.(7)



Equ.(7) is a constraint on the frequency spectrum of a causal response function. It is perhaps better appreciated when expressed in terms of real and imaginary parts,

$$\tilde{H}_r(\omega) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\tilde{H}_i(\omega')}{\omega' - \omega} d\omega' \qquad (8)$$

$$\tilde{H}_i(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\tilde{H}_r(\omega')}{\omega' - \omega} d\omega' \qquad (9)$$

The physical import of this is best understood via a specific application. Take, for example, the propagation of electromagnetic waves through some medium. The speed of the wave through the medium is c/n where c is the speed of light in vacuum and n is the refractive index of the medium. But $c = 1/\sqrt{\varepsilon_0 \mu_0}$ where ε_0 and μ_0 are the permittivity and permeability of the vacuum respectively, and similarly $c/n = 1/\sqrt{\varepsilon_r \varepsilon_0 \mu_r \mu_0}$ so that the refractive index can be expressed in terms of the relative permittivity, ε_r, and the relative permeability, μ_r, as $n = \sqrt{\varepsilon_r \mu_r}$. In most cases the relative permeability is close to unity and hence this becomes $n \approx \sqrt{\varepsilon_r}$. The reason that the medium affects the electromagnetic wave propagation is that the electric field causes polarisation of the material. The polarisation is related to the electric field by,

$$\overline{P} = \varepsilon_0 \chi_e \overline{E} \qquad (10)$$

where χ_e is the electric susceptibility. The electric displacement is the sum of that in vacuum and that due to the polarisation,

$$\overline{D} = \varepsilon_0 \overline{E} + \overline{P} = \varepsilon_0 (1 + \chi_e) \overline{E} = \varepsilon_0 \varepsilon_r \overline{E} \qquad (11)$$

Hence, in this application, the “input” is $\overline{E}$ and the material’s response is $\overline{P}$. Consequently the relevant “response function” is the electric susceptibility, $\chi_e = \varepsilon_r - 1$, since this is the factor of proportionality between them, Equ.(10). We labour this point because incorrect results would be obtained if either the relative permittivity, ε_r, or the refractive index, n, were treated as the response function.

The complex relative permittivity is conventionally written $\varepsilon_r = \varepsilon' + i\varepsilon''$, so inserting $\chi_e = \varepsilon_r - 1$ into Equs.(8,9) gives,

$$\varepsilon'(\omega) = 1 + \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\varepsilon''(\omega')}{\omega' - \omega} d\omega' \qquad (12)$$

$$\varepsilon''(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\varepsilon'(\omega') - 1}{\omega' - \omega} d\omega' \qquad (13)$$

The physical meaning of these relations becomes apparent when the implications of a complex refractive index $n = n_r + in_i$ are considered. The wave vector in vacuum $k = \omega/c$ becomes $k = n\omega/c$ in the medium. Hence the wave is proportional to,

$$\exp i(kx - \omega t) = \exp i\omega\left(\frac{nx}{c} - t\right) = \exp\left\{-\frac{n_i \omega x}{c}\right\} \exp i\omega\left(\frac{n_r x}{c} - t\right) \qquad (14)$$

A wave propagating through a medium whose refractive index has a positive imaginary part therefore decays exponentially with distance. The amplitude decays by a factor of e for every distance $c/n_i\omega$ travelled (the *penetration depth*). But since

$n \approx \sqrt{\varepsilon_r}$ the real and imaginary parts of the relative permittivity are related to those of the refractive index by,

$$\varepsilon' = n_r^2 - n_i^2 \quad \text{and} \quad \varepsilon'' = 2n_r n_i \tag{15}$$

Consequently Equs.(12,13) can be regarded as constraints between the real refractive index and its imaginary part which measures absorptivity. It is rather remarkable that these two properties of the medium should be related since *a priori* they would appear to be quite separate things. More remarkable still, this connection follows from little more than causality and hence is fundamental.

In the case of optics, it is the frequency dependence of the real part of the refractive index of a transparent medium which causes it to separate the colours from an initially white beam of light. The frequency dependence of n_r is therefore referred to as dispersion, and this terminology is passed on to Equs.(12,13) which are called *dispersion relations*.

Note that Equ.(4) implies that the real and imaginary parts of ε_r are respectively even and odd functions of ω, i.e., $\varepsilon'(-\omega) = \varepsilon'(\omega)$ and $\varepsilon''(-\omega) = -\varepsilon''(\omega)$. This means that the dispersion relations, (12,13), can be written as integrals over positive frequencies only,

$$\varepsilon'(\omega) = 1 + \frac{2}{\pi} P\int_0^\infty \frac{\omega'\varepsilon''(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{16}$$

$$\varepsilon''(\omega) = -\frac{2\omega}{\pi} P\int_0^\infty \frac{\varepsilon'(\omega') - 1}{\omega'^2 - \omega^2} d\omega' \tag{17}$$

To emphasise the physical utility of such relations, they provide a means of determining the frequency dependence of the real refractive index by measuring only the absorption over all frequencies.

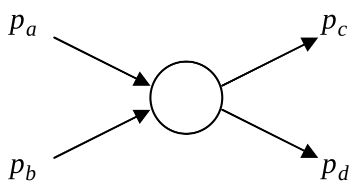
Particle physics provides a second example where dispersion relations have been used to constrain what otherwise might have been regarded as completely separate physical quantities. Note that the term *dispersion relation*, borrowed from optics, continues to be used in particle physics, despite having no direct meaning in this context. The correct mathematical term might be *Cauchy representation* or *Hilbert transform*.

In the 1960s, quantum field theory was at its nadir as regards strong interaction physics. The fundamental problem was that the low energy effective strong coupling constant is large, $g_s \sim 14$, which makes a perturbation expansion meaningless. The 1970s saw a renaissance of strong interaction quantum field theory, thanks to the rise of the gauge theories and the proof of renormalisation. With that the standard model of particle physics was consolidated. However, in the 1960s the vogue was to derive as much as possible about particle interactions without a detailed dynamical model, relying heavily upon the analytic properties of the reaction amplitudes. This becomes particularly productive when coupled with the crossing principle.

Consider a particle reaction,

s-channel: $$a + b \to c + d \tag{18}$$

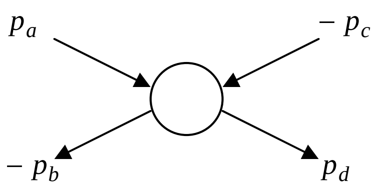
If a sensible field theory for the interaction existed, the amplitude for this process could be written in terms of a sum of all Feynman diagrams of the form,



where the arrows are labelled with the corresponding 4-momenta. The amplitude for (18) can thus be written $f_s(p_a, p_b; p_c, p_d)$. If (18) is a physical process, the energies of all four particles, i.e., the first components of their 4-momenta, must be positive. In fact they must exceed the rest mass of the corresponding particle. However we know from field theory that we can also interpret the above Feynman graph as that for the "crossed" process,

u-channel: $$a + \bar{c} \to \bar{b} + d \qquad (19)$$

where a bar denotes the anti-particle. The Feynman diagram for this reaction is algebraically identical but reinterpreted as,



Consequently the amplitude for this process, called the u-channel, is "the same" function as for the original (s-channel) process, $f_s(p_a, p_b; p_c, p_d)$, except that to be physical for the u-channel we now require the energies E_b and E_c to be negative, and algebraically less than minus the corresponding rest mass. Normally the amplitude for the u-channel would be written differently, as a different function, $f_u(p_a, p'_c; p'_b, p_d)$, where $p'_c = -p_c$ and $p'_b = -p_b$ and all four energies are now required to be positive. The crossing relation holds that $f_u(p_a, p'_c; p'_b, p_d) = f_s(p_a, -p'_b; -p'_c, p_d)$. But if we choose values for the 4-momenta which are physical in the u-channel they are necessarily unphysical in the s-channel, and *vice-versa*.

Since the regions of 4-momentum space in which the two reactions are physical do not overlap, is there any content to the assertion that the two functions are, in some sense, "the same"? There would not be if the functions were otherwise arbitrary. After all, if I tell you what values a real function $g(x)$ takes for $x \geq 0$, what can you conclude about its values for $x < 0$ given no other information? The answer is strictly nothing, since I am free to define the function for negative x however I wish. But if I were also to tell you that the function has a Taylor series expansion which is valid for all x, then you can deduce its value for $x < 0$ by deriving the Taylor series in the region $x \geq 0$. The generalisation of this example for functions of a complex variable is to be told that the function is analytic except at some specified singular points. Specification of the function over some continuous region is sufficient to permit its value elsewhere to be found by analytic continuation, a generalisation of the use of a Taylor series. Analytic continuation can be accomplished only within limitations defined by the singularities of the function.

The postulate that particle reaction amplitudes are analytic defines what is really meant by the loose statement that the functions f_s and f_u for the two processes (18) and (19) are "the same". Within the limitations due to essential singularities, analytic continuation provides the means by which knowledge of the s-channel amplitude can be used to make deductions about the u-channel. What is meant by the "essential singularities" will be discussed shortly, but suffice it to say that the singularities should not be regarded as irritants since they contain the key physics. That scattering amplitudes are maximally analytic otherwise is the central postulate of analytic S-matrix theory, e.g., Eden et al (1966). Whilst not rigorously provable, this hypothesis has a firmer basis than a mere guess. It is inspired, of course, by Titchmarsh's theorem discussed above. In quantum field theory, causality is formulated via the commutativity of the fields at spacelike separations. It is possible, for certain systems and for a limited range of parameters, to derive dispersion relations in a manner independent of a specific Lagrangian via the axiomatic approach to field theory, see for example Klein (1961) and Martin (1966).

To discuss the essential singularities, and their role in dispersion relations in particle physics, consider a specific example: the kaon-nucleon system. The three channels related by crossing are,

s channel: $$K + N \rightarrow K + N \tag{20}$$

u channel: $$\overline{K} + N \rightarrow \overline{K} + N \tag{21}$$

t channel: $$K + \overline{K} \rightarrow N + \overline{N} \tag{22}$$

Hence, the s and u channels are elastic scattering, whereas the t channel is an annihilation channel. K stands for either kaon, K^+ or K^0, and hence $\overline{K}$ refers to K^- or $\overline{K}^0$, whilst N means either nucleon. Consequently there are a multiplicity of different reactions like (20-22), and corresponding scattering amplitudes, depending upon isospin and whether the scalar or vector coupling amplitude is meant (or equivalently, whether the nucleon spin flips or not). We shall gloss over these finer details for clarity.

The conventional variables used to describe these reactions are the three squared 4-momenta,

$$s = (p_K + p_N)^2 ; \qquad u = (p_{\overline{K}} + p_N)^2 ; \qquad t = (p_K + p_{\overline{K}})^2 \tag{23}$$

Only two of these are independent since we see that,

$$s + u + t = 2\left(M_K^2 + M_N^2\right) \tag{24}$$

In the s, u and t channels the corresponding variable, s, u, t, is the square of the centre of mass energy. In addition to the energy, the final state of a channel is specified by an angular scattering variable, say $z_s = \cos\theta_s$ for the s channel. These angular variables can also be expressed in terms of any two of s, u, t. For example,

$$z_s = 1 + \frac{t}{2q_s^2} \quad \text{where,} \quad q_s^2 = \frac{1}{4s}\left[s - (M_N + M_K)^2\right]\left[s - (M_N - M_K)^2\right] \tag{25}$$

The physical region of the s channel is specified by $s \geq (M_K + M_N)^2$ and $|z_s| \leq 1$, and similarly for the u and t channels, none of which overlap.

There are two kinds of singularities which we should expect. The first type are simple poles (Born terms) due to the existence of particles which could fulfil the role of an intermediate combined state. Thus, the u channel has such intermediate states in the form of the hyperons, $\Lambda(1116)$ and $\Sigma(1189)$, and the first resonant state of the sigma, $\Sigma(1385)$. These have the right quantum numbers to form an intermediate state of the u channel, $\overline{K}N$, specifically the same strangeness. This contrasts with the s channel, KN, which has no such possible intermediate state. Note that we are concerned here with states that lie below the physical threshold, i.e., below an energy of $M_K + M_N = 1432$ MeV, since the contribution of above threshold states will be taken into account below. Hence, $\Lambda(1116)$, $\Sigma(1189)$ and $\Sigma(1385)$ all contribute below-threshold u channel poles.

The second type of singularity are branch cuts (or, rather, the points from which the cuts emanate). These singularities are legislated by unitarity. From Chapter 29 we know that the unitarity of the S-matrix, $S^+S = I$, when written in terms of the T-matrix, $S = I + 2iT$, implies $T^+T = \mathrm{Im}(T)$. One corollary of this is the optical theorem, but more generally we have for any initial and final state,

$$\sum_j \langle f|T^+|j\rangle\langle j|T|i\rangle = \mathrm{Im}\langle f|T|i\rangle \tag{26}$$

The sum on the LHS is over all physical intermediate states. When there are no physical intermediate states the LHS is zero and hence so must the imaginary part of the amplitude be zero. Conversely, as the energy is increased, each time the threshold of a new intermediate state is crossed, the imaginary part of the amplitude can increase discontinuously. This suggests that each such threshold is associated with a singularity. These singularities are branch cuts, see Eden et al (1966) for details. By convention these branch cuts run along the positive real u axis from the threshold energy.

By “physical” we mean that the intermediate state must involve particles which are on mass-shell whilst also having the same 4-momentum and total quantum numbers as the initial and final states. Note that this can happen below the physical threshold of the channel. Thus, for the u channel, whose physical threshold is $u \geq (M_K + M_N)^2$, the lowest mass intermediate state is $\Lambda + \pi$, since $M_\Lambda + M_\pi = 1251$ MeV is less than $M_K + M_N = 1432$ MeV. For the t channel, the lowest mass intermediate state is that of two pions, so branch cuts obtain for $t \geq 4M_\pi^2$.

Knowing the singularity structure we can now deduce a dispersion relation simply by performing a Cauchy integration like (5). However, because there are singularities in the form of simple poles and branch points with associated cuts lying on the real axis, that part of the contour is displaced slightly. We must also assume, as before, that the amplitudes decrease sufficiently fast at high energy that the contribution from the infinite semi-circle is zero. However, we now have two degrees of freedom to define the variable over which we are integrating: namely the path in s,u,t space. The simplest path is to take one of these variables, say t, as constant. The resulting fixed-t dispersion relation is of the form,

$$\Re(F(s,t)) = poles + \frac{P}{\pi}\int_{(M_\Lambda+M_\pi)^2}^{\infty}\left[\frac{1}{s'-s}+\frac{1}{s'-u}\right]\mathrm{Im}(F(s',t))\cdot ds' \tag{27}$$

where, $$poles = \alpha_\Lambda \left(\frac{1}{M_\Lambda^2 - s} + \frac{1}{M_\Lambda^2 - u} \right) + \alpha_\Sigma \left(\frac{1}{M_\Sigma^2 - s} + \frac{1}{M_\Sigma^2 - u} \right) \tag{28}$$

Assuming a negative value of t is chosen, Equ.(27) is a constraint involving the s and u channels only, not the t channel. Equ.(27) is illustrative only, the exact form of the dispersion relation depending upon which of the eight amplitudes with differing spin/isospin combinations is considered. A pole term for the $\Sigma(1385)$ resonance can also be added to (28). The $\alpha_\Lambda, \alpha_\Sigma$ are related to the couplings of these hyperons to $\bar{K}N$, see Bradford and Martin (1979) for details.

Dispersion relations like (27) can be useful in extracting the maximum information from scattering data. For example, scattering data tends to define the imaginary part of a scattering amplitude rather better than the real part in some circumstances. So (27) will improve the determination of the real part compared with, say, simple phase-shift analysis.

However, the most impressive implications follow from considering certain curved paths in s, u, t space. These paths have to be chosen carefully so as not to introduce additional branch cuts. One possibility is to use hyperbolae defined by $(s-a)(u-a) = b$ for constant a and b. With a suitable choice of these parameters the hyperbola passes through the physical region of either the s or u channel as well as that of the t channel. The dispersion relation typically becomes,

$$\begin{aligned} \Re(F(s, t(s,a))) = poles + \frac{P}{\pi} \int_{(M_\Lambda + M_\pi)^2}^{\infty} \left[\frac{1}{s'-s} + \frac{1}{s'-u} - \frac{1}{s'-a} \right] \mathrm{Im}(F(s',t')) \cdot ds' \\ + \frac{s-u}{\pi} \int_{4M_\pi^2}^{\infty} \mathrm{Im}\left(\frac{F_t(t',s')}{s'-u'} \right) \frac{dt'}{t'-t} \end{aligned} \tag{29}$$

where,

$$poles = \alpha_\Lambda \left(\frac{1}{M_\Lambda^2 - s} + \frac{1}{M_\Lambda^2 - u} - \frac{1}{M_\Lambda^2 - a} \right) + \alpha_\Sigma \left(\frac{1}{M_\Sigma^2 - s} + \frac{1}{M_\Sigma^2 - u} - \frac{1}{M_\Sigma^2 - a} \right) \tag{30}$$

In (29) it is understood that $t = t(s,a)$ lies on the hyperbola at the point indicated by variable s. Similarly, within the integrals the variables s', u', t' all correspond to the same point on the hyperbola. Again (29,30) are only illustrative, the exact form of the dispersion relation depending upon the particular amplitude chosen.

In (29) the last term relates to the t channel. Consequently (29) provides a constraint on the amplitude for the annihilation reaction $K\bar{K} \to N\bar{N}$ in terms of the amplitudes for elastic scattering $KN \to KN$ and/or $\bar{K}N \to \bar{K}N$. It really is rather remarkable that these purely elastic channels contain information to constrain what appears to be a completely different physical process, namely particle annihilation and pair creation. Moreover, the assumption that below the $N\bar{N}$ threshold the t channel is dominated by low mass meson exchanges permits the coupling strength of these mesons to $K\bar{K} / N\bar{N}$ to be found purely from KN and $\bar{K}N$ scattering data, see Bradford and Martin (1979) for details.

References

E.C.Titchmarsh, “Introduction to the Theory of Fourier Integrals” (1948), Oxford, Clarendon Press, 2nd ed.

R.J.Eden, P.V.Landshoff, D.I.Olive and J.C.Polkinghorne (1966) “The Analytic S-Matrix”, Cambridge University Press.

R.A.W.Bradford and B.R.Martin (1979) "t-Channel Couplings in Kaon-Nucleon Scattering", Zeitschrift fur Physik C, Particles and Fields, **1**, 357-368.

L.Klein (editor), “Dispersion Realations and the Abstract Approach to Field Theory” (1961), Gordon and Breach.

A.D.Martin (1966), Nuovo Cimento **42**, 930. Check Ref – get title