# Model of Coupling-Constant Renormalization*

KENNETH G. WILSON
*Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14850*†
*and*
*Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305*
(Received 7 April 1970)

The charged scalar theory of $\pi$ mesons interacting with a fixed nucleon source is truncated as follows: $\pi$ mesons are permitted to exist only in a set of discrete states $\psi_m(k)$ such that $k$ is of order $\Lambda^m$ in the state $\psi_m(k)$; $\Lambda$ is an arbitrary constant above $4\times10^6$. Also, two mesons of the same charge cannot occupy the same state. The resulting Hamiltonian can be solved by a perturbation expansion in $\Lambda^{-1}$ provided there are only a finite number $M$ of states $\psi_m$. When $M\to\infty$, the renormalized coupling constant and ground-state energy diverge in perturbation theory (in the coupling constant). If the unrenormalized coupling constant is allowed to go to infinity as $M\to\infty$, it is proven that the renormalized theory exists (without ghost states) for any value of the renormalized coupling constant. The proof uses the perturbation analysis in $\Lambda^{-1}$ carried to all orders. This analysis leads to the definition of a transformation $T$ which eliminates one meson degree of freedom from any given Hamiltonian, replacing it by an effective Hamiltonian with one less degree of freedom. The effective Hamiltonian gives exactly all energy levels of the original Hamiltonian except those with mesons in the removed degree of freedom. The renormalizability of the theory is proven using topological properties of $T$. In particular, there is a subtransformation $T_A$ with a nontrivial fixed point $P_c$ whose properties determine the principal features of the renormalized theory. The idea of the fixed point is a generalziation of the Gell-Mann–Low eigenvalue condition for the bare coupling constant of quantum electrodynamics.

## I. INTRODUCTION

THE problem of renormalization has been remarkably unimportant in the study of pure strong interactions (i.e., strong interactions without radiative or weak corrections). The ideas developed since 1954—dispersion relations, Regge poles, current algebra, and pole dominance—all can be formulated and applied without encountering any of the divergences that occur in unrenormalized perturbation theory. As a result one gets the impression that renormalization is no more than a technical modification which one makes on closed-loop Feynman graphs when very accurate perturbation formulas are needed, as for the electron magnetic moment. This impression has encouraged the idea that Lagrangian models of current algebra, such as field algebra and the quark model, can be analyzed for their equal-time commutators as if renormalization were unnecessary.[1]

An entirely opposite picture results from exactly soluble models of field theories with interaction. There are two known model theories which require wave-function or charge renormalization, namely, the Lee model[2] and the Thirring model.[3] It is well known that the renormalized Lee model has a ghost state. The Thirring model involves the Fermi interaction for a zero-mass spinor field in one space and one time dimension. The model has a solution after renormalization, but the solution has radically different behavior at short distances from what one would expect from a canonical Lagrangian picture. The renormalized spinor field does not satisfy canonical commutation relations.[3] More generally, the renormalized theory is scale invariant, as one would have predicted from the Lagrangian (there are no dimensional parameters in the Thirring model, the only parameter being a dimensionless coupling constant). However, the renormalized fields (but not the conserved currents) have different scaling properties from those one predicts from the canonical commutation rules. The dimension of the spinor field (which determines its scaling properties) depends on the coupling constant and can vary from $\frac{1}{2}$ to $\infty$.[4]

The only known relativistic theories where renormalization does not affect the short-distance behavior appreciably are the "superrenormalizable" theories which may require mass renormalization but do not require infinite coupling-constant or wave-function renormalization in perturbation theory.[5] In these theories the short-distance behavior is close to the free-field behavior. Unfortunately, there are no acceptable four-dimensional superrenormalizable theories.

In a recent paper, it was proposed that there would be nontrivial renormalization effects in strong interactions.[6] It was postulated that these effects would have the same form as in the Thirring model, namely, scale invariance would be valid at short distances but the dimensions of local fields would be different from any free-field model (except for the currents of current algebra whose dimensions are fixed by the algebra).

* Work supported by the U. S. Atomic Energy Commission.
† Permanent address.
[1] This idea persists despite the rseults of S. Adler and Wu-Ki Tung [Phys. Rev. Letters **22**, 978 (1969)] and R. Jackiw and G. Preparata [*ibid.* **22**, 975 (1969); **22**, 1162 (1969)], who show that equal-time commutators are affected in perturbation theory by renormalization.
[2] T. D. Lee, Phys. Rev. **95**, 1329 (1954).
[3] W. Thirring, Ann. Phys. (N. Y.) **3**, 91 (1958); K. Johnson, Nuovo Cimento **20**, 773 (1961).
[4] The anomalous scaling properties of the Thirring model are implicit in Johnson's solution (Ref. 3); see also Ref. 6.
[5] The two-dimensional $\phi^4$ theory analyzed by Jaffe and Glimm is in this class. For references, see A. Jaffe, Rev. Mod. Phys. **41**, 576 (1969).
[6] K. Wilson, Phys. Rev. **179**, 1499 (1969).

It was shown that renormalization effects could account for a universal $\Delta I=\frac{1}{2}$ rule in weak interactions and could determine the convergence or divergence of some of the Weinberg sum rules.

The fact that the $\Delta I=\frac{1}{2}$ rule might be explained by renormalization effects means that renormalization can be of great practical importance. One would like to understand renormalization better. The Lee model and the Thirring model fall far short of providing the depth of understanding required. The reason is that both models have very special features and the renormalization of these models may simply reflect these special features. The Lee model is special because of the decoupling of the $N$-$\theta$ channel from the many-particle channels. This decoupling is the simplification that makes solution of the Lee model possible. The Thirring model is special for many reasons, but in particular the electromagnetic current of the Thirring model satisfies a free-field equation which is the starting point for solving the model. Also, there is no coupling-constant renormalization in the Thirring model. If there had been coupling-constant renormalization in the Thirring model, it might have shown the same diseases as the Lee model which does involve coupling-constant renormalization.

The purpose of this paper is to define and solve a new model of coupling-constant renormalization. The new model is a cousin of the Lee model but its renormalization is very different from that of the Lee model. The new model is a derivative of the charged scalar theory of pions coupled to a fixed nucleon source. The model Hamiltonian is obtained essentially by projecting the Hamiltonian of the charged scalar theory onto a specially constructed subspace of the original Hilbert space. The result of renormalizing the model is that the renormalized theory exists without ghosts; the renormalized coupling constant is arbitrary but the unrenormalized coupling constant is infinite.

The model of this paper cannot be solved in closed form. To make it soluble by series expansions, a large parameter $\Lambda$ is introduced artificially into the model; the model is then solved by an expansion in $\Lambda^{-1}$. $\Lambda$ is introduced by restricting the $\pi$ mesons of the model to be in one of a discrete set of wave functions $\psi_m(k)$, where the mean momentum of $\psi_m(k)$ is $\Lambda^m$ (in units of the pion mass). Thus instead of the pion energy being continuously variable from 1 to $\infty$, it is restricted to the discrete values 1, $\Lambda$, $\Lambda^2$, etc. This means the Hamiltonian has some terms of order 1, some terms of order $\Lambda$, etc., so one can do perturbation theory when $\Lambda$ is large. This idea was explained in an earlier paper[7] where a more complicated version of the model was proposed.

Because the model cannot be solved in closed form, the renormalization analysis is much more complex than that for either the Lee model or the Thirring model. The analysis is further complicated because one cannot simply study the lowest-order term in the $\Lambda^{-1}$ expansion. To prove the renormalizability of the theory, one must show that the expansion in $\Lambda^{-1}$ of the renormalized theory is finite to all orders and that the sum of the series converges. To prove this, a rigorous analysis of the model is given using formal techniques of analysis in Hilbert space plus some topological methods. The formal analysis is possible because the model is specially constructed to involve only bounded operators. To ensure that no unbounded operators occur, the number of $\pi$ mesons per state $\psi_m$ is limited to one of each charge, and the total number of states $\psi_m$ is cut off at $m=M$. One investigates the limit for $M\to\infty$, but for any finite $M$ one has bounded operators.

The author recommends that the papers of Lee[2] (on the Lee model) and Johnson[3] (on the Thirring model) be read before attacking the present paper. They provide some background on exact solutions of renormalizable theories and are very much simpler to read.

There are three interesting features in the model of this paper. The first is simply that a finite renormalized theory exists. Actually, all that is proved is that the renormalized energy levels exists. Because there are no continuum (momentum) states open to pions, there is no scattering in the model; all energy levels are discrete and hence calculating the energy levels is the most important problem in the model. The theory is found to be free of ghosts. No matrix elements of operators other than the Hamiltonian are discussed. In particular, the nucleon isospin operators are not examined, which means we cannot compute the renormalized coupling constant as conventionally defined. The reason these operators are not considered is that the analysis that would be required exceeds the author's patience.

The second feature of the model is that scale invariance is preserved in the renormalized theory for energies large compared to the pion mass. The unrenormalized Hamiltonian of the full charged scalar theory is scale invariant in the limit of zero pion mass. This invariance is preserved in the unrenormalized Hamiltonian of the model except that it is a discrete invariance: Only scale transformations which take wave functions $\psi_m(k)$ into wave functions $\psi_{m+l}(k)$ occur in the model. The renormalized energy levels exhibit scale invariance when the energies are large, but the scaling law is different from what one predicts from the unrenormalized Hamiltonian. To be precise, the unrenormalized Hamiltonian $H_0$ goes into $\Lambda^{-1}H_0$ when $\psi_m\to\psi_{m+1}$, apart from terms of order 1, but the renormalized Hamiltonian $H_R$ goes into $\Lambda^{-1}\beta H_R$, where $\beta$ is a constant (about $\frac{1}{2}$). Thus the model of this paper supports the hypothesis that renormalization can preserve scale invariance at large energies but will change the scaling laws of operators.

[7] K. Wilson, Phys. Rev. 140, B445 (1965).

The third feature of the model, and probably the most important, is that in order to prove the renormalizability of the model, it is necessary to define and study a topological transformation $T$ acting on a space $S$ of cutoff Hamiltonians. The space $S$ contains the unrenormalized cutoff Hamiltonians for any cutoff $M$. However, it also contains cutoff Hamiltonians involving arbitrarily complicated interactions involving products of arbitrarily many meson creation and destruction operators. In other words, the space $S$ includes nonrenormalizable interactions of arbitrarily complicated structure. The transformation $T$ takes a Hamiltonian with cutoff $M$ into a Hamiltonian with cutoff $M-1$ without changing the physics of these Hamiltonians. To be precise, the original Hamiltonian and the transformed Hamiltonian have exactly the same energy levels except for those energy levels with mesons explicitly present in the state $\psi_M$; such levels are not present in the transformed Hamiltonian. The transformation defines how the coupling constants of all possible interactions must change as the cutoff $M$ changes in order to keep the energy levels of the theory fixed. Having very many coupling constants all changing as the cutoff changes is analogous to having an infinite number of counter terms in a renormalization analysis in ordinary perturbation theory. One has an infinite number of counter terms when one tries to renormalize a nonrenormalizable theory. This is customarily regarded as a disaster, for one presumes that for every infinite counter term there is an arbitrary finite counter term, leading to an infinite number of parameters. This disaster does not occur in the model. The reason is that strict bounds on the coupling constants will be included in the definition of $S$, and one cannot introduce extra free parameters without violating these bounds. What actually happens is that the possible renormalizable theories of the model are described by effective cutoff Hamiltonians obtained by applying $T$ an infinite number of times to the original unrenormalized uncutoff Hamiltonian. This means that the renormalized Hamiltonians must lie in a subspace $R_S$ of $S$, where $R_S$ is the limit of the subspaces $T^m(S)$ for $m\to\infty$. The space $R_S$ is found to be a three-dimensional space for given cutoff $M$. Hence there are only three adjustable parameters in the renormalized Hamiltonian: a scale factor, an additive constant, and the renormalized coupling constant (suitably defined).

If one is interested only in the first two features of the model one can probably skim the hard parts (Sec. V and Appendix B). One would read these sections in detail only to check for mistakes. However, to understand the transformation $T$ one must study the whole paper in detail; it is hard to have a clear understanding of the role of the transformation $T$ without studying the spaces $T^m(S)$; one must see how these spaces shrink with $m$ to the limiting space $R_S$, and one must understand in practice the relevance of these spaces to the renormalization problem. At present the only way to get the necessary practice is to work through the model of this paper.

Gell-Mann and Low have given a general discussion of nonperturbative renormalization theory, using quantum electrodynamics as an example.[8] The relation of their work to the type of model considered here is discussed in Sec. VII. The idea of a transformation $T$ in which an infinite set of coupling constants are transformed as the cutoff $M$ is reduced is a generalization of Gell-Mann and Low's idea of a cutoff-dependent electromagnetic coupling $e(\Lambda)$.

In the author's previous paper on model Hamiltonians,[7] a more complicated model was discussed, in which $\pi$ mesons were allowed to have any momentum in the intervals $0<k<k_0$, $\frac{1}{2}\Lambda<k<\Lambda$, $\frac{1}{2}\Lambda^2<k<\Lambda^2$, etc., where $k_0$ was a constant. This meant the meson creation and destruction operators were continuum creation and destruction operators, which are hardly suitable for rigorous analyses. The $\Lambda^{-1}$ expansion was proposed but only carried out in lowest order. Even the lowest-order calculation was complicated by the fact that the unperturbed Hamiltonians were themselves insoluble field-theoretic Hamiltonians. One had to guess the qualitative structure of their solution. Furthermore, as the cutoff $M$ went to infinity the coupling constant in the unperturbed Hamiltonian had to become large, resulting in closely spaced isobar states, which interfered with the perturbation calculation in $\Lambda^{-1}$. None of these difficulties are present in the model of this paper. The meson creation and destruction operators of this paper are defined to be discrete and bounded. The unperturbed Hamiltonians are finite dimensional and diagonalizable in closed form (cf. Table I). The energy-level spacing of the unperturbed Hamiltonian does not become small for large coupling—the isobars in the previous theory involved many mesons in a single quantum state and this is forbidden in the present model. This means the present model lacks much of the physics of the full charged scalar theory, but it still illustrates the renormalization problem, which is its only purpose.

This paper divides into three stages. The first stage consists of Secs. II–IV. In Sec. II, the Hamiltonian of the model is defined. In Sec. III the perturbation expansion in $\Lambda^{-1}$ is formulated for the cutoff Hamiltonian and some properties of the expansion are worked out in low orders. In Sec. IV a perturbation formula is defined which allows the $\Lambda^{-1}$ expansion to be defined to all orders in a convenient form. The second stage consists of Secs. V and VI. In Sec. V the transformation $T$ is defined. Its principal properties are stated (Theorems 1–4; the proofs of these theorems are in Appendix B). Then the topological analysis required to prove renormalizability is carried through. Finally, the renormalized Hamiltonian is defined for any given

[8] M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954). See also M. Baker and K. Johnson, *ibid.* 183, 1292 (1969).

renormalized coupling constant. In Sec. VI scale transformations are defined, and the scaling properties of the renormalized energy levels are computed. The third stage consists of Sec. VII, where it is shown that the transformation $T$ is more than a technical device to prove the existence of the renormalized theory. Specifically it is shown that the renormalized theories are not the unique solution of any uncutoff Hamiltonian; instead, the transformation $T$ is involved in the definition of the renormalized theory, and this definition is most simply stated in terms of one of the fixed points of the transformation. We also relate the renormalization program of this paper to conventional renormalization theory and especially to the Gell-Mann–Low analysis.

## II. MODEL HAMILTONIAN

The unrenormalized Hamiltonian of the model is as follows:

$$H=\sum_{m=0}^{\infty}\Lambda^m\{(a_m^\dagger a_m+b_m^\dagger b_m-1)+g_0(a_m+b_m^\dagger)\tau^+ +g_0(a_m^\dagger+b_m)\tau^-\}, \quad (2.1)$$

where $g_0$ and $\Lambda$ are constants, $\tau^+$ and $\tau^-$ are the isospin raising and lowering operators for the nucleon, and the operators $a_m^\dagger$ and $b_m^\dagger$ are $\pi^+$ and $\pi^-$ creation operators, respectively, for the state $\psi_m$. The subtraction $-1$ is included for irrelevant reasons. The constant $\Lambda$ must be large ($>4\times10^6$ in the rigorous analysis). To prevent two $\pi^+$ or two $\pi^-$ from occupying the same state, the operators $a_m$, $a_m^\dagger$, $b_m$, and $b_m^\dagger$ are assigned the commutation relations of a set of Pauli spin operators:

$$\{a_m,a_m^\dagger\}=\{b_m,b_m^\dagger\}=1, \quad (2.2)$$

$$a_m{}^2=(a_m^\dagger)^2=b_m{}^2=(b_m^\dagger)^2=0, \quad (2.3)$$

$$[a_m,b_m]=[a_m,b_m^\dagger]=[a_m,a_n]=0, \text{ etc.} \quad (m\neq n), \quad (2.4)$$

where $[\ ]$ is a commutator and $\{\ \}$ is an anticommutator. The Hilbert space on which $H$ acts is a product space. The components of the product are, first, the two-dimensional nucleon space with the bare proton state $|p\rangle$ and bare neutron state $|n\rangle$ as a basis. Second, for each wave function $\psi_m$ there is a component space of four dimensions. A basis for each such component consists of a vacuum state, a $\pi^+$ state, a $\pi^-$ state, and a $\pi^+\pi^-$ state, each meson being in the state $\psi_m$.

The model Hamiltonian can be arrived at starting from the full Hamiltonian of the charged scalar fixed-source theory[9] if one replaces the fixed momentum creation operators $a_{\mathbf{k}}^\dagger$ and $b_{\mathbf{k}}^\dagger$ of the mesons by

$$a_{\mathbf{k}}^\dagger\rightarrow\sum_m a_m^\dagger\psi_m(\mathbf{k}), \quad (2.5)$$

$$b_{\mathbf{k}}^\dagger\rightarrow\sum_m b_m^\dagger\psi_m(\mathbf{k}). \quad (2.6)$$

After these substitutions are inserted in the full Hamiltonian, one must drop any off-diagonal products such as $a_n^\dagger a_m$ $(n\neq m)$ and replace integrals such as $\int_{\mathbf{k}}\omega_{\mathbf{k}}|\psi_m(\mathbf{k})|^2$ or $\int_{\mathbf{k}}(2\omega_{\mathbf{k}})^{-1/2}\psi_m(\mathbf{k})$ [where $\omega_{\mathbf{k}}$ is $(1+\mathbf{k}^2)^{1/2}$] by order-of-magnitude estimates, assuming that the functions $\psi_m(\mathbf{k})$ are normalized to unity and vanish unless $\mathbf{k}\sim\Lambda^m$. There is no need for the model Hamiltonian to have any connection with the fixed-source theory, because the model will be studied on its own merits. The connection with the fixed-source theory is used only to provide a language to describe the operators $a_m$, etc. Likewise, the wave functions $\psi_m(\mathbf{k})$ play no role in the analysis of the model; their only purpose is to give an intuitive meaning to the operators $a_m$, etc.

One can cut off the Hamiltonian by restricting the sum over $m$ to a finite range, say $0\leq m\leq M$. Then the Hamiltonian becomes a finite bounded matrix; in this case it is diagonalizable without renormalization. The problem of renormalization arises when one tries to let $M\rightarrow\infty$. Then one has an infinite number of degrees of freedom, which is well known to be a source of difficulties.[10] To compound the situation, the scale of energy associated with the $m$th degree of freedom increases as $\Lambda^m$, so that the most important degrees of freedom are those with $m\sim M$ instead of small $m$. Clearly one has difficulties in the limit $M\rightarrow\infty$ regardless of what happens in perturbation theory, but it is still worth showing that in perturbation theory one has a problem specifically with coupling-constant renormalization. Let $|P\rangle$ and $|N\rangle$ be the normalized physical proton and neutron states, i.e., the ground states of $H$. The renormalized coupling constant is

$$g_R=g_0\langle P|\tau^+|N\rangle, \quad (2.7)$$

using the definition analogous to that used in the full charged scalar theory. The matrix element $\langle P|\tau^+|N\rangle$ can be computed to second order in $g_0$ by straightforward perturbation theory. If the cutoff $M$ is finite, then

$$g_R=g_0-g_0{}^3(M+1)+O(g_0{}^5). \quad (2.8)$$

The cutoff momentum $k_M$ is of order $\Lambda^M$ so $M$ is proportional to $\ln k_M$; hence $g_R$ is logarithmically divergent as in the full charged scalar theory. The divergence for $M\rightarrow\infty$ is directly due to there being an infinite number of degrees of freedom in the no-cutoff limit.

The structure of the energy-level spectrum of the cutoff Hamiltonian can be seen by a qualitative analysis. It is convenient to call a meson in a state $\psi_m(k)$ an "$m$-meson." Let the cutoff Hamiltonian be

[9] The full Hamiltonian is given by Eq. (1) of Ref. 7.

[10] Cf. the paper of A. Wightman, in *High Energy Electromagnetic Interactions and Field Theory*, edited by M. Lévy (Gordon and Breach, New York, 1967), especially pp. 245–262 and references cited therein.

denoted $H_M$. It has the structure

$$H_M=\sum_{m=0}^{M}\Lambda^m O_m, \tag{2.9}$$

where $O_m$ is independent of $\Lambda$ and involves only $m$-meson operators and the nucleon operators $\tau^+$ and $\tau^-$. The smallest part of $H_M$ is $O_0$. This is the only part of $H_M$ involving 0-mesons, and for $\Lambda$ large, $O_0$ is a perturbation on the rest of the Hamiltonian. The remainder of the Hamiltonian has energies of order $\Lambda$ or larger and thus should have energy-level spacings of order $\Lambda$; each level is four-fold degenerate (at least) because each level is independent of the presence or absence of 0-mesons. Adding $O_0$ splits these levels, with the splitting being of order 1. Next one can discuss the effect of the term $\Lambda O_1$; clearly this should lead to a gross spacing of order $\Lambda$, neglecting fine structure due to $O_0$. But $\Lambda O_1$ can itself be regarded as a perturbation; there exists (neglecting $\Lambda O_1$ and $O_0$) a spacing of order $\Lambda^2$, then a spacing of order $\Lambda^3$, etc.

The problem of renormalization is the problem of computing the ground state and those excited states which have a finite energy above the ground state in the limit $M\to\infty$. This means calculating states with an energy of order $\Lambda^m$ above the ground state, for any $m$, but with $m$ held when $M\to\infty$. In practice one calculates only energy differences between the ground state and various excited states. The ground-state energy itself diverges for $M\to\infty$. An energy difference of order $\Lambda^m$ is much smaller than the basic energy scale $\Lambda^M$, when $M$ is large, so a very precise calculation is required to give these energy differences accurately. This fact plus the fact that the model cannot be solved exactly, and must be solved as a perturbation expansion in $\Lambda^{-1}$, is the reason this paper is so long.

The model Hamiltonian is invariant to three symmetries: charge conservation, charge conjugation, and time reversal. The charge $Q$ is

$$Q=\sum_m (a_m^\dagger a_m-b_m^\dagger b_m)+\tfrac{1}{2}(\tau_z+1), \tag{2.10}$$

where $\frac{1}{2}\tau_z$ is the $z$ component of the nucleon isospin; $Q$ commutes with $H$. The charge conjugation transformation interchanges $\pi^+$ with $\pi^-$ and $p$ with $n$. Let $U_c$ be the unitary transformation giving these interchanges; then

$$U_c^\dagger a_m U_c=b_m, \tag{2.11}$$

$$U_c^\dagger b_m U_c=a_m, \tag{2.12}$$

$$U_c^\dagger \tau^+ U_c=\tau^-, \tag{2.13}$$

$$U_c^\dagger H U_c=H. \tag{2.14}$$

The time-reversal transformation is an antilinear unitary transformation $U_T$ with the properties

$$U_T^\dagger a_m U_T=a_m^*, \tag{2.15}$$

$$U_T^\dagger b_m U_T=b_m^*, \tag{2.16}$$

$$U_T^\dagger \tau^+ U_T=(\tau^+)^*, \tag{2.17}$$

$$U_T^\dagger H U_T=H^*. \tag{2.18}$$

## III. PRELIMINARY ANALYSIS OF MODEL HAMILTONIAN

In order to solve the renormalization problem, one must first be able to solve the cutoff Hamiltonian for arbitrarily large cutoff $M$. In this section, we give a preliminary discussion of the solution of the cutoff Hamiltonian for large $M$.[11] The constant $\Lambda$ is also large, but held fixed, and $M$ can be arbitrarily large even compared to $\Lambda$. The cutoff Hamiltonian naturally separates into an unperturbed Hamiltonian and a perturbation:

$$H_M=H_{0M}+H_{IM}, \tag{3.1}$$

where

$$H_{0M}=\Lambda^M O_M, \tag{3.2}$$

$$H_{IM}=\sum_{m=0}^{M-1}\Lambda^m O_m, \tag{3.3}$$

and

$$O_m=a_m^\dagger a_m+b_m^\dagger b_m-1+g_0(a_m+b_m^\dagger)\tau^+ +g_0(a_m^\dagger+b_m)\tau^-. \tag{3.4}$$

The operator $O_M$ is easily diagonalized. One can ignore the mesons in states other than $\psi_M$, in which case $O_M$ acts on the eight-dimensional Hilbert space involving the nucleon and mesons in the state $\psi_M$. Owing to charge conservation, the matrix for $O_M$ separates into submatrices of size $3\times3$ at most. The eigenstates of $O_M$ are given in Table I [the variables $(m,g)$ of Table I must be replaced by $(1,g_0)$]. It has two degenerate ground states: a state $|P\rangle$ of charge 1 and a state $|N\rangle$ of charge 0. The ground state becomes highly degenerate when mesons in other states $\psi_m$ are considered, since one can add such mesons to the states $|P\rangle$ and $|N\rangle$ without changing the eigenvalue of $O_M$.

The Hamiltonian $H_{0M}$ has an energy-level spacing of order $\Lambda^M$ ($g_0\Lambda^M$ if $g_0$ is large), while $H_{IM}$ is at most of order $\Lambda^{M-1}$ ($g_0\Lambda^{M-1}$ for $g_0$ large). Hence one is allowed to treat $H_{IM}$ as a perturbation when $\Lambda$ is large, for any value of $g_0$. However, one must carry the perturbation

TABLE I. Eigenstates of the Hamiltonian $m(a^\dagger a+b^\dagger b-1)+g(a+b^\dagger)\tau^++g(a^\dagger+b)\tau^-$, where $a^\dagger$ creates $\pi^+$, $b^\dagger$ creates $\pi^-$, $|p\rangle$ and $|n\rangle$ are nucleon states, and $\mu=m(m^2+2g^2)^{-1/2}$, $\gamma=g(m^2+2g^2)^{-1/2}$. The other four eigenstates are obtained by charge conjugation ($p\leftrightarrow n$, $\pi^+\leftrightarrow\pi^-$).

| Eigenvalue | Eigenstate |
|---|---|
| $-(m^2+2g^2)^{1/2}$ | $\frac{1}{2}(1+\mu)\lvert p\rangle-\gamma\lvert n\pi^+\rangle+\frac{1}{2}(1-\mu)\lvert p\pi^+\pi^-\rangle$ |
| 0 | $\lvert p\pi^+\rangle$ |
| 0 | $\gamma\lvert p\rangle+\mu\lvert n\pi^+\rangle-\gamma\lvert p\pi^+\pi^-\rangle$ |
| $(m^2+2g^2)^{1/2}$ | $\frac{1}{2}(1-\mu)\lvert p\rangle+\gamma\lvert n\pi^+\rangle+\frac{1}{2}(1+\mu)\lvert p\pi^+\pi^-\rangle$ |

[11] See also Sec. III of Ref. 7.

expansion out to order $M$ at least, because one ultimately is interested in energy-level spacings which may be of order 1. In the lowest order of degenerate perturbation theory, the ground state of $H$ and excited states at energy $\Lambda^{M-1}$ or less above the ground state are given by an effective Hamiltonian

$$H_{\text{eff}}=E_{0M}+PH_{IM}P, \tag{3.5}$$

where $E_{0M}$ is the ground-state energy of $H_{0M}$ and $P$ is a projection operator on the ground states of $H_{0M}$. $H_{\text{eff}}$ acts on a product space whose components are the two-dimensional space with basis $|P\rangle$ and $|N\rangle$ and the meson space for the states $\psi_m$, $0\leq m\leq M-1$. One can introduce isospin raising and lowering operators $\tau_R^{\pm}$ for $|P\rangle$ and $|N\rangle$; then $H_{\text{eff}}$ involves a set of operators $[\tau_R^{\pm}, a_m\ (0\leq m\leq M-1)$, etc.], which are equivalent to the operators of $H_{M-1}$. The only way $P$ affects the operator $H_{IM}$ is through the nucleon operators $\tau^+$ and $\tau^-$; the meson operators in $H_{IM}$ are unaffected. To express $H_{\text{eff}}$ in terms of $\tau_R^{\pm}$, one must express $P\tau^+P$ and $P\tau^-P$ in terms of $\tau_R^{\pm}$ and the meson operators. The operator $P\tau^+P$ affects only the states $|P\rangle$ and $|N\rangle$ not the meson states, and because it increases the charge by one unit, $P\tau^+P$ must be proportional to $\tau_R^+$. The proportionality constant $Z$ is found from Table I to be (using the constants of Table I)

$$Z=\langle P|\tau^+|N\rangle=(m^2+g^2)(m^2+2g^2)^{-1}. \tag{3.6}$$

With $m=1$ and $g=g_0$ this is

$$Z(g_0)=(1+g_0^2)(1+2g_0^2)^{-1}. \tag{3.7}$$

Likewise $P\tau^-P$ is $Z(g_0)\tau_R^-$. Hence $PH_{IM}P$ has the same form as $H_M$ itself except that $M$ is replaced by $M-1$ and $g_0\tau^{\pm}$ is replaced (in Eq. 3.4) by $g_{M-1}\tau_R^{\pm}$, with

$$g_{M-1}=h(g_0), \tag{3.8}$$

$$h(g)=g(1+g^2)(1+2g^2)^{-1}. \tag{3.9}$$

When degenerate perturbation theory is carried to higher orders, one still computes an effective Hamiltonian $H_{\text{eff}}$ which acts on the space of ground states of $H_{0M}$. The effective Hamiltonian is no longer just $PH_{IM}P$ but contains higher-order terms in $H_{IM}$, for example, the second-order term is $PH_{IM}(1-P)\times(E_{0M}-H_{0M})^{-1}H_{IM}P$. The term of $n$th order involves products of $n$ interaction Hamiltonians and $n-1$ energy denominators. It is useful to discuss in a schematic way the types of terms generated in the higher-order calculation. Let $x_m$ stand for an operator of the form $a_m^\dagger a_m+b_m^\dagger b_m-1$, $a_m+b_m^\dagger$, or $a_m^\dagger+b_m$. Let $\tau$ stand for any nucleon operator and $\tau_R$ for any operator acting on $|P\rangle$ and $|N\rangle$. Let $x_m^2$ stand for operators made of any product of operators of type $x_m$. One can easily make a table of the type of operators that can occur in $H_{\text{eff}}$ for a given order in $\Lambda$, remembering that $H_{\text{eff}}$ involves $H_{IM}$ times products of $(E_{0M}-H_{0M})^{-1}H_{IM}$, the whole product being projected with $P$. The results are shown in Table II.

TABLE II. Breakdown of $H_{\text{eff}}$ by type of operator for each order in $\Lambda$. The symbols $x_m$, $\tau_R$, and $(x_m)^2$ are explained in the text. Any operator listed for a power $\Lambda^m$ can occur for lower powers of $\Lambda$ also.

| Order in $\Lambda$ | Types of operators |
|---|---|
| $\Lambda^M$ | const |
| $\Lambda^{M-1}$ | $x_{M-1}\tau_R$ |
| $\Lambda^{M-2}$ | $x_{M-2}\tau_R$, $(x_{M-1})^2\tau_R$ |
| $\Lambda^{M-3}$ | $x_{M-3}\tau_R$, $x_{M-2}x_{M-1}\tau_R$ |
| $\Lambda^{M-4}$ | $x_{M-4}\tau_R$, $x_{M-3}x_{M-1}\tau_R$, $(x_{M-2})^2\tau_R$, $x_{M-2}(x_{M-1})^2\tau_R$ |

The formulas for the higher-order terms of the degenerate perturbation calculation are too complicated to quote explicitly. Fortunately they are not needed; it is sufficient to have upper bounds for each type of term and these can be obtained. Table II gives the order in $\Lambda$ of each term and all that remains is to obtain numerical upper bounds. This will be done in Sec. V.

The Hamiltonian $H_{\text{eff}}$ has a basic energy scale $\sim\Lambda^{M-1}$ which is still much larger (for large $M$) than the energy scales of interest. $H_{\text{eff}}$ can again be analyzed by perturbation theory. One writes

$$H_{\text{eff}}=H_{0\,\text{eff}}+H_{I\,\text{eff}}. \tag{3.10}$$

The unperturbed Hamiltonian is

$$H_{0\,\text{eff}}=\Lambda^{M-1}O_{M-1}(g_{M-1}), \tag{3.11}$$

where $O_{M-1}(g_{M-1})$ is the same as $O_{M-1}$, except that $g_0$ is replaced by $g_{M-1}$, and $\tau^{\pm}$ by $\tau_R^{\pm}$. All other terms in $H_{\text{eff}}$ form the perturbation $H_{I\,\text{eff}}$, which is at most of order $\Lambda^{M-2}$. The eigenstates of $O_{M-1}(g_{M-1})$ can be determined from Table I; like $O_M$ it has two degenerate ground states $|P'\rangle$ and $|N'\rangle$ if mesons in states other than $\psi_{M-1}$ are ignored. One can use degenerate perturbation theory starting from the states $|P'\rangle$ and $|N'\rangle$ to determine the eigenstates of $H_{\text{eff}}$ of energy $\Lambda^{M-2}$ or less above the ground state. Again one must calculate the perturbation analysis to many orders, in order to keep terms with energies of order 1 or larger. The result is a second effective Hamiltonian $H_{\text{eff}}'$ involving meson operators $a_m$, etc., for $m\leq M-2$ and isospin operators $\tau_R^{\pm\prime}$ connecting the states $|P'\rangle$ and $|N'\rangle$.

One can determine the type of operators that occur in $H_{\text{eff}}'$ for each order in $\Lambda$. The basic operators are operators acting on $|P'\rangle$ and $|N'\rangle$, denoted $\tau_R'$, and meson operators of type $x_m$ for $m\leq M-2$. The results are shown in Table III. In constructing Table III, one uses the fact that operators of the form $(x_{M-1})\tau_R$ and $(x_{M-1})^2\tau_R$ in $H_{I\,\text{eff}}$ are reduced to the form $\tau_R'$ in $H_{\text{eff}}'$. Furthermore, the symmetries of the theory ensure that an operator of the form $\tau_R'$ in $H_{\text{eff}}$ not multiplied by a meson operator can only be a constant. The important result illustrated by Table III is the following: To compute $H_{\text{eff}}'$, one must compute many orders in a perturbation treatment of $H_{I\,\text{eff}}$. $H_{I\,\text{eff}}$ itself divides into

TABLE III. Breakdown of $H_{\text{eff}}'$ by type of operator for each order in $\Lambda$ (cf. Table II).

| Order in $\Lambda$ | Types of operators |
|---|---|
| $\Lambda^M$, $\Lambda^{M-1}$ | const |
| $\Lambda^{M-2}$ | $x_{M-2}\tau_R'$ |
| $\Lambda^{M-3}$ | $x_{M-3}\tau_R'$, $(x_{M-2})^2\tau_R'$ |
| $\Lambda^{M-4}$ | $x_{M-4}\tau_R'$, $x_{M-3}x_{M-2}\tau_R'$ |

two parts. The simple part of $H_{I\,\text{eff}}$ are the terms coming from $PH_{IM}P$; these terms have the structure $x_{M-2}\tau_R$, $x_{M-3}\tau_R$, etc., and depend only on the single constant $g_{M-1}$. The complex part of $H_{I\,\text{eff}}$ comes from the higher-order terms in $H_{IM}$, and includes all terms of type $(x_{M-2})^2\tau_R$, $(x_{M-2})(x_{M-3})\tau_R$, etc. In computing $H_{\text{eff}}'$, even the simple part of $H_{I\,\text{eff}}$ generates all types of terms in $H_{\text{eff}}'$, through terms of order $(H_{I\,\text{eff}})^2$, $(H_{I\,\text{eff}})^3$, etc. The important fact is for a given term in $H_{\text{eff}}'$, say $(x_{M-2})^2\tau_R'$, its coefficient comes predominately from the simple part of $H_{I\,\text{eff}}$, and hence the coefficient is primarily determined by the constant $g_{M-1}$. $H_{I\,\text{eff}}$ also has an $(x_{M-2})^2\tau_R$ term but this affects the coefficient of $(x_{M-2})^2\tau_R'$ only in order $\Lambda^{M-4}$, whereas the dominant part of the coefficient is of order $\Lambda^{M-3}$. Because of this result one can give bounds on the complex terms like $(x_{M-2})^2\tau_R'$ in $H_{\text{eff}}'$ which depend on $g_{M-1}$ only and do not involve the size of the corresponding term in $H_{\text{eff}}$. These bounds are of crucial importance for the rigorous analysis; they ensure that the complex interactions cannot increase without bound as one repeats the perturbation analysis many times. Furthermore, it means that Table III has the same form it would have had if one had started with the cutoff $M-1$, and obtained $H_{\text{eff}}'$ by solving $H_{M-1}$. The only exception is the constant in Table III of order $\Lambda^M$.

One can repeat the perturbation analysis many times, generating a sequence of Hamiltonians which will be denoted $H_N(M)$. The Hamiltonian $H_{M-1}(M)$ is $H_{\text{eff}}$, $H_{M-2}(M)$ is $H_{\text{eff}}'$. In general, $H_N(M)$ is the effective Hamiltonian after $M-N$ perturbation calculations; $H_N(M)$ involves the meson operators $a_m$, etc., for $m\leq N$, and isospin operators analogous to $\tau_R{}^{\pm}$ or $\tau_R{}^{\pm\prime}$. The operator $H_N(M)$ gives the energy levels of $H$ with energies of order $\Lambda^N$ or less above the ground state. For each operator $H_N(M)$ one can give a classification table analogous to Tables II and III; the result is Table II with $M$ replaced by $N$, except for constant terms. The unperturbed part of $H_N(M)$ would appear to be just $\Lambda^N O_N(g_N)$, where

$$g_N=h(g_{N+1}). \tag{3.12}$$

This is what one gets if the unperturbed Hamiltonian is defined as the term of order $\Lambda^N$ in $H_N(M)$. However, to ensure that the perturbation is small even when $M-N$ is much larger than $\Lambda$, the unperturbed part of $H_N(M)$ will be defined to include other terms of the form $(a_N+b_N{}^\dagger)\tau^+$, or $(a_N{}^\dagger+b_N)\tau^-$, regardless of their order in $\Lambda$. The unperturbed Hamiltonian still has the form $\Lambda^N O_N(g_N)$, but $g_N$ differs from $h(g_{N+1})$ in order $\Lambda^{-1}$. Since one has to compute a whole sequence of constants $g$ ($N=M-1$, $M-2$, etc.), the small differences between $g_N$ and $h(g_{N+1})$ for each $N$ can build up to a macroscopic effect when $M-N$ is large.

To compute an eigenstate of energy $\Lambda^m$ above the ground state of $H_M$, one must take the effective Hamiltonian $H_m(M)$ and solve for states corresponding to excited states of the unperturbed part of $H_m(M)$. One could set up a perturbation method for computing these states. It will not be necessary for the purposes of this paper to discuss these states in detail, so the perturbation method will not be developed here.

## IV. PERTURBATION FORMULA

There are various standard formulas for the effective Hamiltonian that results when a perturbation $H_I$ is treated to all orders. They all have drawbacks, so a suitable formula will be derived here. The formula obtained below has two properties: The effective Hamiltonian is Hermitian, and involves only unperturbed energies in energy denominators. The second property is useful because the unperturbed energies are known explicitly. The first property is obviously useful, and is not true of many standard formulas.

Let $H=H_0+H_I$ and let $P$ be the projection operator on the ground states of $H_0$. Let $|\psi\rangle$ be any eigenstate of $H$ with an energy $E$ close to the ground-state energy $E_0$ of $H_0$. It is convenient to have an operator $R$ which gives the part of $|\psi\rangle$ outside the space projected by $P$ in terms of the part of $|\psi\rangle$ inside the space. That is,

$$(1-P)|\psi\rangle=RP|\psi\rangle. \tag{4.1}$$

Such an operator can be defined as follows. The eigenvalue equation has two parts:

$$E(1-P)|\psi\rangle=(1-P)H(1-P)|\psi\rangle + (1-P)H_IP|\psi\rangle, \tag{4.2}$$

$$EP|\psi\rangle=PH_I(1-P)|\psi\rangle+PHP|\psi\rangle. \tag{4.3}$$

If an operator $R$ satisfying Eq. (4.1) exists, one can multiply the second equation [Eq. (4.3)] by $R$ and subtract from the first, giving

$$0=\{(1-P)HR+(1-P)H_I -RPH_IR-RPH\}P|\psi\rangle. \tag{4.4}$$

Equation (4.4) will certainly be satisfied if we demand that

$$(1-P)HR+(1-P)H_IP-RPH_IR-RPHP=0. \tag{4.5}$$

This equation can be cast in a form suitable for iteration in $H_I$. From the original definition of $R$, it should take states within the subspace projected by $P$ into states orthogonal to this subspace; we can also require that $R$ gives zero acting on states outside the subspace. This

means that

$$R=RP, \tag{4.6}$$

$$R=(1-P)R. \tag{4.7}$$

Assuming this, and using the fact that $PH_0P=E_0P$, one can rewrite Eq. (4.5) as

$$(E_0-H_0)R=(1-P)H_IP+(1-P)H_IR -RH_IP-RH_IR, \tag{4.8}$$

or

$$R=(E_0-H_0)^{-1}(1-P-R)H_I(P+R). \tag{4.9}$$

This equation can be solved iteratively to give $R$ as a power series in $H_I$. It is easily seen that the expansion satisfies the assumptions of Eqs. (4.6) and (4.7).

The argument so far does not prove that any operator $R$ satisfying Eq. (4.9) will also satisfy Eq. (4.1), but this will be established later if $H_I$ is sufficiently small.

One can now write Eq. (4.3) as

$$EP|\psi\rangle=\{PH_0P+PH_IP+PH_IR\}P|\psi\rangle. \tag{4.10}$$

One could therefore define $H_{\text{eff}}$ to be $H_0+PH_IP+PH_IR$, except that $PH_IR$ is not Hermitian. The reason for this is that although two eigenstates $|\psi_1\rangle$ and $|\psi_2\rangle$ with distinct eigenvalues are orthogonal, the corresponding projected states $P|\psi_1\rangle$ and $P|\psi_2\rangle$ will probably not be orthogonal, and therefore cannot be distinct eigenstates of a Hermitian operator. To remedy the situation, one notes that

$$\langle\psi_1|\psi_2\rangle=\langle\psi_1|P|\psi_2\rangle+\langle\psi_1|R^\dagger R|\psi_2\rangle. \tag{4.11}$$

This suggests replacing the projected states $P|\psi_1\rangle$ and $P|\psi_2\rangle$ by the states $(1+R^\dagger R)^{1/2}P|\psi_1\rangle$ and $(1+R^\dagger R)^{1/2}P|\psi_2\rangle$, which are still states in the subspace projected by $P$ but have the same scalar product as $|\psi_1\rangle$ and $|\psi_2\rangle$. The operator $(1+R^\dagger R)^{1/2}$ is well defined as a power series in $R^\dagger R$ when $H_I$ is small. To obtain $H_{\text{eff}}$, write the eigenvalue equation as

$$(E-H)(P+R)P|\psi\rangle=0 \tag{4.12}$$

and multiply by $(P+R^\dagger)$:

$$E(P+R^\dagger)(P+R)P|\psi\rangle=(P+R^\dagger)H(P+R)P|\psi\rangle. \tag{4.13}$$

Now

$$(P+R^\dagger)(P+R)=P+R^\dagger R=(1+R^\dagger R)P \tag{4.14}$$

using Eqs. (4.6) and (4.7). Hence, multiplying Eq. (4.13) by $(1+R^\dagger R)^{-1/2}$ gives

$$E|\phi\rangle=H_{\text{eff}}|\phi\rangle, \tag{4.15}$$

where

$$|\phi\rangle=(1+R^\dagger R)^{1/2}P|\psi\rangle, \tag{4.16}$$

$$H_{\text{eff}}=(1+R^\dagger R)^{-1/2}(P+R^\dagger) \times H(P+R)(1+R^\dagger R)^{-1/2}. \tag{4.17}$$

The formula for $H_{\text{eff}}$ is evidently Hermitian.

The above argument is not rigorous, so it must now be proven that the eigenvalues of $H_{\text{eff}}$ are the eigenvalues of $H$ near $E_0$, and that eigenstates $|\phi\rangle$ of $H_{\text{eff}}$ become eigenstates $|\psi\rangle$ of $H$ through the formula

$$|\psi\rangle=(P+R)(1+R^\dagger R)^{-1/2}|\phi\rangle. \tag{4.18}$$

Assume that $R$ is defined by Eq. (4.9) solved by iteration assuming $H_I$ is small. It is shown in Appendix A that the iteration converges if $H_I$ is sufficiently small. The solution satisfies Eqs. (4.6) and (4.7). From these and Eq. (4.9) one obtains

$$(1-P-R)H(P+R)=0, \tag{4.19}$$

which is essentially Eq. (4.5). Also,

$$(1-P-R)(P+R)=0. \tag{4.20}$$

This is because

$$(1-P-R)(P+R)=(1-P)(1-R)(1+R)P =(1-P)R^2P \tag{4.21}$$

and

$$R^2=RP(1-P)R=0. \tag{4.22}$$

Let $|\phi\rangle$ be an eigenstate of $H_{\text{eff}}$ in the subspace projected by $P$, and let $E$ be its eigenvalue. Define $|\psi\rangle$ by Eq. (4.18).

One can write

$$(1+R^\dagger R)^{1/2}(E-H_{\text{eff}})P|\phi\rangle=0. \tag{4.23}$$

Using Eqs. (4.17), (4.18), and (4.14), Eq. (4.23) may be rewritten

$$(P+R^\dagger)(E-H)|\psi\rangle=0. \tag{4.24}$$

This equation cannot be used to infer that $(E-H)|\psi\rangle=0$ because $P+R^\dagger$ projects onto a subspace and does not have an inverse. However, from Eqs. (4.19) and (4.20) one can obtain

$$(1-P-R)(E-H)|\psi\rangle=(1-P-R)(E-H)(P+R) \times\{(1+R^\dagger R)^{-1/2}|\phi\rangle\}=0. \tag{4.25}$$

Adding Eqs. (4.24) and (4.25) gives

$$(1+R^\dagger-R)(E-H)|\psi\rangle=0. \tag{4.26}$$

It is shown in Appendix A that $(1+R^\dagger-R)$ has an inverse (for sufficiently small $H_I$) so this equation does imply that $|\psi\rangle$ is an eigenstate of $H$ with eigenvalue $E$. The Hamiltonian $H_{\text{eff}}$ has matrix elements equal to zero except within the subspace projected by $P$. Within this subspace $H_{\text{eff}}$ has $d$ orthogonal eigenstates, where $d$ is the dimension of the subspace. These eigenstates define [through Eq. (4.18)] $d$ orthogonal eigenstates of $H$ (orthogonality is easily verified). The energies of these eigenstates are close to $E_0$ because $H_{\text{eff}}$ is approximately $PHP$ when $H_I$ is small so that $R$ is small.

An alternative form of $H_{\text{eff}}$ is obtained as follows. Write

$$H_{\text{eff}}=E_0P+(1+R^\dagger R)^{-1/2}(P+R^\dagger) \times(H_I+H_0-E_0)(P+R)(1+R^\dagger R)^{-1/2}. \tag{4.27}$$

Using $(H_0-E_0)P=0$ and Eq. (4.9), one can rewrite this as

$$H_{\text{eff}}=E_0P+(1+R^\dagger R)^{-1/2}(P+R^\dagger)\{H_I(P+R) -(1-P-R)H_I(P+R)\}(1+R^\dagger R)^{-1/2}. \quad (4.28)$$

Using Eq. (4.14), this simplifies to

$$H_{\text{eff}}=E_0P+P(1+R^\dagger R)^{1/2} \times H_I(P+R)(1+R^\dagger R)^{-1/2}. \quad (4.29)$$

This formula is not manifestly Hermitian, but $H_{\text{eff}}$ is still Hermitian since it is still defined by Eq. (4.17).

## V. EXACT PERTURBATION ANALYSIS OF MODEL

The outline of a method of solving the cutoff-model Hamiltonian $H_M$ was given in Sec. III. One uses the definition of $H_{\text{eff}}$ given in Sec. IV in each degenerate perturbation calculation. The result is that starting from $H_M$, for any $M$, one defines a sequence of effective Hamiltonians denoted $H_N(M)$ involving meson operators $a_m$, $a_m^\dagger$, $b_m$, $b_m^\dagger$ for $0\leq m\leq N$ and isospin operators which will be denoted $\tau^\pm$ regardless of what states they act on ($|p\rangle$, $|n\rangle$, or $|P\rangle$, $|N\rangle$, or $|P'\rangle$, $|N'\rangle$, etc.). The effective Hamiltonians involve very complicated interactions, not just the $O_m$ terms of the original model. From the analysis of Sec. III, one can expect to get upper bounds on these terms such that a $\Lambda^N O_N$ term is the dominant term in $H_N(M)$, provided that an appropriate coupling constant replaces $g_0$ in $O_N$. The Hamiltonians $H_N(M)$ give the energies of the ground state of $H_M$ and the excited states of $H_M$ in which only the first $N$ degrees of freedom are excited. If the energy levels are counted from the lowest level up, the ground state being number one, then $H_N(M)$ describes the first $2^{2N+3}$ levels of $H_M$.

The limit of no cutoff, that is, the $M\to\infty$ limit, can be studied by studying the limits of $H_N(M)$ for fixed $N$, as $M\to\infty$. This means one is studying a fixed number of energy levels as $M$ increases. It will be proven in this section that the limit of $H_N(M)$ for $M\to\infty$ exists provided that one makes the renormalizations one expects from ordinary perturbation theory. This means that before letting $M\to\infty$ one must first subtract a constant $E_M$ from $H_N(M)$ and allow the bare coupling constant $g_{0M}$ to vary with $M$. The variation will be such that $g_{0M}\to\infty$ as $M\to\infty$, i.e., the interaction term in $H_M$ swamps the free meson energy in the limit $M\to\infty$. The proof requires that $\Lambda$ be larger than $4\times10^6$. The limit may exist for smaller $\Lambda$ but in this case the upper bounds used in the proof no longer apply.

The Hamiltonians of this paper involve only bounded operators: the operators $a_m$, $a_m^\dagger$, $\tau^+$, $\tau^-$, etc. All have operator bounds of order 1. Anyone with experience in rigorous quantum mechanics knows the joys of having only bounded operators. This ensures that terms that look small by a power of $\Lambda$ will indeed be small if $\Lambda$ is large enough; for finite $M$ the perturbation expansions in $\Lambda^{-1}$ will be easily proven to converge and one can concentrate on the problems of the $M\to\infty$ limit.

The analysis of the limit for $M\to\infty$ is still very complex; it will be presented here in a formal and not well-motivated manner. Before presenting the procession of theorems and definitions, the basic problem involved will be sketched briefly. The essential problem is to have a bound on the difference $\|H_N(M,g_{0M}) -E_M-H_N(L,g_{0L})+E_L\|$, where $\|\ \|$ is the ordinary operator bound, and the dependence of $H_N(M)$ on $g_{0M}$ has been noted explicitly. One must be able to show that this bound goes to zero as $M$ and $L$ go to $\infty$, provided that the sequences $\{g_{0M}\}$ and $\{E_M\}$ have been chosen appropriately. The crucial step in establishing such a bound will be to show that the difference $H_N(M)-H_N(L)-E$ is arbitrarily small when $M$ and $L$ are large, provided that $E$ is properly chosen and that the terms of order $\Lambda^N O_N$ which dominate $H_N(M)$ and $H_N(L)$ have identical effective coupling constants (see Theorem 10). This condition will force one to have different bare coupling constants; $g_{0M}\neq g_{0L}$. As a preliminary to proving this theorem it will be proved (Theorem 1) that $H_N(M)$ is dominated by a term of the structure $\Lambda^N O_N$ with an appropriate effective coupling constant in $O_N$. This proof is necessary because otherwise one might worry that terms nominally of order $\Lambda^{N-1}$ or less would be multiplied by powers of $M$, which would dominate the $\Lambda^N$ term when $M\gg\Lambda$.

In order to clarify the calculation of bounds, some topological language will be used. A space $S$ of Hamiltonians will be defined which includes the effective Hamiltonians $H_N(M)$ as special cases. The perturbation analysis which defines $H_{N-1}(M)$ given $H_N(M)$ defines a transformation $T$ on the space $S$. The space $S$ will be defined so that $T(S)$ is contained in $S$. A metric will be defined on $S$, and convergence questions discussed in terms of this metric. The Hamiltonians $H_N(M)$, considered for all possible values of $g_0$, define "curves" in $S$.

The exact and rigorous analysis of the renormalization problem begins here. The first step is to define the space $S$ of Hamiltonians. It is convenient to adopt a specific way of representing the Hamiltonians that will be included in $S$. Let $H$ be any Hamiltonian involving the meson degrees of freedom 0-$N$ plus nucleon operators, for example, $H_N(M)$ for some $M$. It will be convenient to renumber the meson operators, making the switches $a_0$, $b_0\leftrightarrow a_N$, $b_N$, $a_1$, $b_1\leftrightarrow a_{N-1}$, $b_{N-1}$, etc. In the new numbering, $a_m^\dagger$ creates a meson in the state $\psi_{N-m}$. This is to be true for all $N$, so the state associated with $a_m^\dagger$ is different for different $N$. It is also convenient to separate an additive and a multiplicative factor from $H$, writing

$$H=J\mathcal{H}+\mathcal{E}, \quad (5.1)$$

where $J$ and $\mathcal{E}$ are constants. A normalization condition will be imposed on $\mathfrak{IC}$, determining $J$, but the separation of $\mathcal{E}$ from $J\mathfrak{IC}$ will be left indeterminate. (The transformation $T$ will be defined to determine $J$, $\mathfrak{IC}$, and $\mathcal{E}$ separately.) One now lets $\mathfrak{IC}$ have the following structure:

$$\mathfrak{IC}=mV_{01}+\sqrt{2}gV_{02}\tau^{+}+\sqrt{2}gV_{03}\tau^{-}+\sum_{k=1}^{N}\mathbf{V}_k\cdot\mathbf{A}_{k-1}+\sum_{k=0}^{N}C_k, \quad (5.2)$$

where $\mathbf{V}_k$ is a vector $(V_{k1},V_{k2},V_{k3})$ and

$$\mathbf{V}_k=\sum_{m=k}^{N}\mathbf{T}_m, \quad (5.3)$$

$$T_{m1}=\Lambda^{-m}(a_m^{\dagger}a_m+b_m^{\dagger}b_m-1), \quad (5.4)$$

$$T_{m2}=(1/\sqrt{2})\Lambda^{-m}(a_m+b_m^{\dagger}), \quad (5.5)$$

$$T_{m3}=(1/\sqrt{2})\Lambda^{-m}(a_m^{\dagger}+b_m), \quad (5.6)$$

and $A_{k1}$, $A_{k2}$, $A_{k3}$, and $C_k$ are operators which depend only on $\tau^{\pm}$ and the meson operators numbered from 0 to $k$. The vector notation $\mathbf{A}_k$, $\mathbf{S}_k$, etc. is used purely for convenience. The constants $m$ and $g$ will be required to satisfy a normalization condition:

$$m^2+2g^2=1. \quad (5.7)$$

To ensure this normalization condition, $m$ and $g$ will be represented as

$$m=\cos\theta, \quad (5.8)$$

$$g=(1/\sqrt{2})\sin\theta. \quad (5.9)$$

The set of parameters $J$, $\mathcal{E}$, $N$, and $\theta$, and the operators $\mathbf{A}_k$ and $C_k$, will be called the "decomposition" of $H$. The representation is highly redundant; for example, $C_N$ is by itself totally arbitrary. The reason for using this redundant representation is the following. One can see from Table II that the operators $a_k$, $b_k$ for large $k$ (new numbering) appear in any effective Hamiltonian $H_N(M)$ predominantly in terms such as $\mathbf{V}_1\cdot\mathbf{A}_0$ or $\mathbf{V}_2\cdot\mathbf{A}_1$. Terms which must go into $C_k$ (the $x_k^2$ terms of Table II) have much smaller coefficients. Hence, by making the separation, one can put stringent upper bounds on the operators $C_k$.

The space $S$ will be defined in two steps, the first step being to define a subsidiary space $S_A$.

*Definition.* A point $P_A\in S_A$ consists of an angle $\theta$ and an infinite set of operators $\mathbf{A}_k$ and $C_k$ $(0\leq k<\infty)$. The angle $\theta$ is restricted to the range $0\leq\theta\leq\frac{1}{2}\pi$. The operators $\mathbf{A}_k$ and $C_k$ can depend only on the nucleon isospin operators $\tau^{\pm}$ and the meson operators $a_m$, $a_m^{\dagger}$, $b_m$, and $b_m^{\dagger}$ for $0\leq m\leq k$. The dependence on these operators is arbitrary except as follows. The operators $\mathbf{A}_k$ and $C_k$ must satisfy the following operator bounds:

$$\|A_{k1}\|\leq 200mg^2\Lambda^{-k-1}, \quad (5.10)$$

$$\|A_{k2}\|\leq 200\sqrt{2}g^3\Lambda^{-k-1}, \quad (5.11)$$

$$\|A_{k3}\|\leq 200\sqrt{2}g^3\Lambda^{-k-1}, \quad (5.12)$$

$$\|C_k\|\leq 200g^2\Lambda^{-2k-1}, \quad (5.13)$$

where $m=\cos\theta$, $g=(1/\sqrt{2})\sin\theta$. Secondly, the operators $\mathbf{A}_k$ and $C_k$ must satisfy symmetry requirements: $A_{k1}$ and $C_k$ must carry charge 0 while $A_{k2}$ creates one unit of charge and $A_{k3}$ destroys a unit of charge. Under charge conjugation, $A_{k1}\rightarrow A_{k1}$, $A_{k2}\leftrightarrow A_{k3}$, and $C_k\rightarrow C_k$. Under time reversal, $\mathbf{A}_k\rightarrow\mathbf{A}_k^{*}$ and $C_k\rightarrow C_k^{*}$. Also $A_{k1}$ and $C_k$ must be Hermitian, while $A_{k3}=A_{k2}^{+}$. These requirements ensure that $\mathfrak{IC}$ [defined by Eq. (5.2)] is Hermitian and invariant to the symmetries. The parameter $\theta$ and the operators $\mathbf{A}_k$ and $C_k$ will be called the decomposition of $P_A$.

The powers of $\Lambda$ in these bounds are what one would expect from Table II; the coefficients are hindsight bounds. It is convenient for the following analysis to insist that an infinite set of $\mathbf{A}_k$ and $C_k$ be specified even if a particular Hamiltonian involves only a finite subset of them. The superfluous $\mathbf{A}_k$ and $C_k$ can be chosen arbitrarily subject to the restrictions of the definition of $S_A$.

The space $S$ is defined as follows.

*Definition.* A point $P\in S$ consists of three constants $J$, $\mathcal{E}$, and $N$, and a point $P_A\in S_A$. The four objects $J$, $\mathcal{E}$, $N$, and $P_A$ will be called the decomposition of $P$. $N$ must be an integer, $J$ must be positive, but $\mathcal{E}$ is arbitrary. There are no upper bounds on $J$, $|\mathcal{E}|$, or $N$.

Next the transformation $T$ acting on $S$ will be defined. Many details of the definition are handled in Appendix B and only an outline is given here. Any Hamiltonian $H$ in $S$ has a dominant term of the form

$$H_0=\mathcal{E}+J\{m(a_0^{\dagger}a_0+b_0^{\dagger}b_0-1)+g(a_0+b_0^{\dagger})\tau^{+}+g(a_0^{\dagger}+b_0)\tau^{-}\}. \quad (5.14)$$

The remaining terms in $H$ form a perturbation $H_I$:

$$H_I=H-H_0. \quad (5.15)$$

From the definitions (5.1) and (5.2) and the bounds (5.10)–(5.13), $H_I$ is of order $J\Lambda^{-1}$ or less and therefore can be treated as a perturbation relative to $H_0$. In particular, one can use the formulas of Sec. IV to define a new Hamiltonian $H_{\text{eff}}$ whose eigenvalues are the eigenvalues of $H$ near the ground-state energy of $H_0$.

Suppose $H$ has a decomposition $(J,\mathcal{E},N,P_A)$ (with $P_A$ in $S_A$). The Hamiltonian $H_{\text{eff}}$ can also be decomposed in the form $(J',\mathcal{E}',N',P_A')$ with $P_A'$ in $S_A$, that is, $H_{\text{eff}}$ can be written in the form defined in Eqs. (5.1)–(5.9). [The resulting operators $\mathbf{A}_k'$, etc. satisfy the bounds of Eqs. (5.10)–(5.13); see Theorem 1.] Specific formulas for $J'$, $\mathcal{E}'$, $N'$, and $P_A'$ (i.e., $g'$, $m'$, $\mathbf{A}_k'$, and $C_k'$) are obtained in Appendix B [cf. Eqs.

(B20)–(B24)]. The general form of these formulas is as follows:

$$N' = N - 1, \tag{5.16}$$

$$J' = \Lambda^{-1} J T_B(P_A), \tag{5.17}$$

and

$$\mathcal{E}' = \mathcal{E} + J T_C(P_A), \tag{5.18}$$

$$P_A' = T_A(P_A), \tag{5.19}$$

where $T_B(P_A)$ and $T_C(P_A)$ are functions depending on $P_A$ but not on $N$, $J$, or $\mathcal{E}$, and $T_A$ is a transformation on the space $S_A$, independent of $N$, $J$, or $\mathcal{E}$. It is clear that $J$ and $\mathcal{E}$ will be multiplicative and additive factors in $H_{\text{eff}}$ and thus do not affect $T_A$, $T_B$, or $T_C$. It is less obvious that $T_A$, $T_B$, and $T_C$ can be defined to be independent of $N$; this result is proven in Appendix B. Equation (5.16)–(5.19) define the transformation $T$.

The reason for defining the subsidiary space $S_A$ is that the transformation $T_A$ acts on this space, and it is convenient to do much of the topological analysis on the transformation $T_A$ rather than on $T$ itself. The space $S_A$ is a continuous closed space; in particular, it does not involve the discrete variable $N$.

The unrenormalized cutoff Hamiltonians $H_M$ are all in $S$. The decomposition of $H_M$ can be defined to be

$$J = \Lambda^M (1+2g_0^2)^{1/2}, \tag{5.20}$$

$$\mathcal{E} = 0, \tag{5.21}$$

$$\theta = \tan^{-1}(\sqrt{2} g_0), \tag{5.22}$$

$$m = (1+2g_0^2)^{-1/2}, \quad g = g_0(1+2g_0^2)^{-1/2}, \tag{5.23}$$

$$\mathbf{A}_k = C_k = 0 \quad (\text{all } k). \tag{5.24}$$

$g_0$ must be positive so that $\theta$ lies between 0 and $\frac{1}{2}\pi$. Note that $m \leq 1$ and $g \leq (1/\sqrt{2})$; this is required by the normalization condition (5.7).

In Appendix B several theorems about the transformation $T_A$ are proven. These theorems will be quoted below and are the basis for the analysis in this section.

*Theorem 1.* If $P_A \in S_A$ then $T_A(P_A)$ is also in $S_A$, i.e.,

$$T_A(S_A) \subset S_A. \tag{5.25}$$

*Theorem 2.* Let $P_A \in S_A$ have a decomposition $(\theta, \mathbf{A}_k, C_k)$, and let $T_A(P_A)$ have the decomposition $\theta'$, $\mathbf{A}_k'$, and $C_k'$. Let $m = \cos\theta$ and $g = (1/\sqrt{2}) \sin\theta$. Then

$$\tan\theta' = \sqrt{2} g''/m'', \tag{5.26}$$

$$T_B(P_A) = (m''^2 + 2g''^2)^{1/2}, \tag{5.27}$$

where

$$|m'' - m| < 0.01 m g^2, \tag{5.28}$$

$$|g'' - g(1-g^2)| < 0.01 g^3. \tag{5.29}$$

Also,

$$|T_C(P_A) + 1| < 0.01, \tag{5.30}$$

$$(1 - 0.51 \sin^2\theta) \tan\theta \leq \tan\theta' \leq (1 - 0.48 \sin^2\theta) \tan\theta. \tag{5.31}$$

*Theorem 3.* Let $P_A$ and $T_A(P_A)$ have the decompositions defined in Theorem 2. Let the component $A_{k1}$ of $\mathbf{A}_k$ vanish for all $k$. Then

$$A_{k1}' = 0 \quad (\text{all } k), \tag{5.32}$$

$$m'' = m, \tag{5.33}$$

where $m''$ is the constant in Theorem 2.

The significance of these theorems is essentially as follows. Theorem 1 ensures that if the decomposition of $P_A$ satisfies the bounds (5.10)–(5.13), then so does the decomposition of $T_A(P_A)$. A consequence of Theorem 1 is that the effective Hamiltonians $H_N(M)$ are in $S$ for any $N$, any $M$, and any value of $g_0$. Theorem 2 gives limits on the values of $T_B(P_A)$, $T_C(P_A)$, and $\theta'$ which depend only on $m$ and $g$, not on $\mathbf{A}_k$ and $C_k$. The constants $m''$ and $g''$ appear in an intermediate stage in the calculation of $H_{\text{eff}}$. To lowest order in $\Lambda^{-1}$, $g''$ is equal to $g(1-g^2)$; this follows from Eq. (3.6) using Eq. (5.7). The bounds in Eqs. (5.28)–(5.30) were originally of order $\Lambda^{-1}$, but were replaced by numerical bounds (valid for $\Lambda > 4 \times 10^6$) for convenience. Theorem 3 shows that $A_{k1}$ will vanish for the effective Hamiltonians $H_N(M)$. It was not obvious (to the author, at least) that this would be so.

Before presenting Theorem 4, a metric must be defined in the space $S_A$. Let $P_A = (\theta, \mathbf{A}_k, C_k)$ and $P_A' = (\theta', \mathbf{A}_k', C_k')$ be two points in $S_A$. It is convenient to define two distances in $S_A$, one being a distance between $\theta$ and $\theta'$, the other a distance between the operators $\{\mathbf{A}_k, C_k\}$ and the operators $\{\mathbf{A}_k', C_k'\}$. It is also convenient to use the notation $|P_A - P_A'|$ for the pair of distances $(d_1, d_2)$.

*Definition.* Let $P_A = (\theta, \mathbf{A}_k, C_k)$ and $P_A' = (\theta', \mathbf{A}_k', C_k')$ be in $S_A$. Then $|P_A - P_A'| = (d_1, d_2)$ with

$$d_1 = 2 |\sin\tfrac{1}{2}(\theta - \theta')|, \tag{5.34}$$

$$d_2 = \max\{\sqrt{2}\Lambda^{k+1} \|A_{ki} - A_{ki}'\|, \Lambda^{2k+1} \|C_k - C_k'\|\}, \tag{5.35}$$

where the maximum is over all possible values of $k$ and $i$.

The distance $d_1$ is more transparent if written in terms of $m$, $g$, $m'$, and $g'$:

$$d_1 = [(m-m')^2 + 2(g-g')^2]^{1/2}. \tag{5.36}$$

No *a priori* rationale for these definitions of $d_1$ and $d_2$ will be given. A certain amount of experimentation was required to determine how to define these distances; the above formulas turned out to be useful. It is clear from Eqs. (5.35) and (5.36) that the metric satisfies the triangle inequality and that $|P_A - P_A'| = (0,0)$ only if $P_A = P_A'$.

*Theorem 4.* Let $P_A$ and $P_A'$ be in $S_A$. Let $|P_A - P_A'| = (d_1, d_2)$ and $|T_A(P_A) - T_A(P_A')| = (d_1', d_2')$. Then

$$0.38 d_1 - 10^{-5} d_2 \leq d_1' \leq 20 d_1 + 10^{-5} d_2, \tag{5.37}$$

$$d_2' \leq 1100 d_1 + 0.06 d_2. \tag{5.38}$$

The coefficients $10^{-5}$ and 0.06 are numerical upper bounds to terms behaving as $\Lambda^{-1}$. These bounds are valid for $\Lambda>4\times10^6$. The first set of bounds force $d_1'$ to be of order $d_1$ unless $d_2\gg d_1$; $d_1'$ cannot be much less or much greater than $d_1$ unless $d_2\gg d_1$. The second bound is a straight upper bound on $d_2'$. In particular, if $d_1=0$, then $d_2'$ is smaller than $d_2$. Hence as long as $\theta=\theta'$, the transformation $T_A$ brings the points $P_A$ and $P_A'$ closer together.

The four theorems stated above are proved in Appendix B. The only assumption is $\Lambda>4\times10^6$. The remainder of the analysis of this section is self-contained. The next stage is a set of topological theorems and definitions. First one defines a set of curves $Q_L$ in the space $S_A$. They are generated by the effective Hamiltonians $H_N(M)$ as a function of the coupling constant $g_0$. The curves turn out to depend only on the difference $L=M-N$, not on $M$ or $N$ separately. It is convenient to parametrize these curves by their $\theta$ coordinate rather than by the unrenormalized coupling constant. The parameter in these curves will be denoted $t$. Let the decomposition of $Q_L(t)$ be written

$$(\theta_L(t),\mathbf{A}_{Lk}(t),C_{Lk}(t)).$$

*Definition.* $Q_0(t)$ is the curve

$$\theta_0(t)=t, \tag{5.39}$$

$$\mathbf{A}_{0k}(t)=C_{0k}(t)=0. \tag{5.40}$$

*Definition.* The curve $Q_L$ is defined iteratively for $L>0$ by the relation

$$Q_L=T_A(Q_{L-1}). \tag{5.41}$$

If one were parametrizing using the unrenormalized coupling constant, one would have had $Q_L(t)=T_A(Q_{L-1}(t))$. With the alternative parametrization, $Q_L(t)$ must still be the transform of some point on $Q_{L-1}$. This point can be denoted $Q_{L-1}(F_L(t))$:

$$Q_L(t)=T_A(Q_{L-1}[F_L(t)]). \tag{5.42}$$

*Definition.* The parametrization $Q_L(t)$ of $Q_L$ is to be chosen so that

$$\theta_L(t)=t \quad (\text{all } L). \tag{5.43}$$

In practice this definition defines the function $F_L(t)$. We shall also be interested in the inverse function $f_L(t)$ to $F_L(t)$. This function satisfies

$$Q_L(f_L(t))=T_A(Q_{L-1}(t)). \tag{5.44}$$

Since the $\theta$ coordinate of $Q_L(f_L(t))$ is $f_L(t)$, one has

$$f_L(t)=\theta \text{ coordinate of } T_A(Q_{L-1}(t)). \tag{5.45}$$

The next theorem gives several properties of $Q_L(t)$, $f_L(t)$, and $F_L(t)$. These properties will be established simultaneously in a proof by induction.

*Theorem 5.*

(a) $Q_L(t)$ is a single-valued function of $t$ defined for $0\leq t\leq\frac{1}{2}\pi$.

(b) $f_L(t)$ is a continuous single-valued function of $t$ defined for $0\leq t\leq\frac{1}{2}\pi$ satisfying

$$f_L(0)=0, \tag{5.46}$$

$$f_L(\tfrac{1}{2}\pi)=\tfrac{1}{2}\pi, \tag{5.47}$$

$$0<f_L(t)<t \quad \text{for} \quad 0<t<\tfrac{1}{2}\pi. \tag{5.48}$$

(c) $F_L(t)$ is a continuous single-valued function of $t$ defined for $0\leq t\leq\frac{1}{2}\pi$ satisfying

$$F_L(0)=0, \tag{5.49}$$

$$F_L(\tfrac{1}{2}\pi)=\tfrac{1}{2}\pi, \tag{5.50}$$

$$t<F_L(t)<\tfrac{1}{2}\pi \quad \text{for} \quad 0<t<\tfrac{1}{2}\pi. \tag{5.51}$$

(d) Consider any pair of numbers $t$ and $t'$ in the range 0 to $\frac{1}{2}\pi$. Let $|Q_L(t)-Q_L(t')|=(d_1,d_2)$. Then

$$d_2\leq4000d_1, \tag{5.52}$$

$$|f_L(t)-f_L(t')|\leq40|t-t'|, \tag{5.53}$$

$$|F_L(t)-F_L(t')|\leq40|t-t'|. \tag{5.54}$$

Part (a) is the crucial part of the theorem. It states that the curve $Q_L$, projected on the $\theta$ axis, covers the full range $0\leq\theta\leq\frac{1}{2}\pi$ once and only once. If, for example, the curve $Q_L$ covered only part of this range, the theory would not be renormalizable. This point will be discussed later.[12]

*Proof of Theorem 5.* The property (a) and Eq. (5.52) hold for $L=0$. That is, $Q_0(t)$ satisfies (a) from its definition, and $|Q_0(t)-Q_0(t')|=(d_1,0)$ for all $t$ and $t'$ and thus it satisfies (5.52). Suppose property (a) and Eq. (5.52) are true of $Q_{L-1}(t)$. We prove (a)–(d) for $Q_L$, $F_L$, and $f_L$. Equations (5.46)–(5.48) are consequences of the inequalities (5.31) [remember that the $\theta$ coordinate of $Q_{L-1}(t)$ is $t$]. Now let $t''$ and $t'''$ be two parameters in the range 0 to $\frac{1}{2}\pi$. Let

$$|Q_{L-1}(t'')-Q_{L-1}(t''')|=(d_1,d_2),$$

and let

$$|T_A(Q_{L-1}(t''))-T_A(Q_{L-1}(t'''))|=(d_1',d_2').$$

These distances must satisfy the inequalities of Theorem 4, and $d_2$ satisfies Eq. (5.52) by assumption. These equations can be combined to give inequalities not involving $d_2$:

$$0.34d_1\leq d_1'\leq21d_1, \tag{5.55}$$

$$d_2'\leq1340d_1. \tag{5.56}$$

Note that

$$d_1=2|\sin\tfrac{1}{2}(t''-t''')|, \tag{5.57}$$

$$d_1'=2|\sin\tfrac{1}{2}[f_L(t'')-f_L(t''')]|. \tag{5.58}$$

[12] See Sec. VII C of the present paper.

Because $t''$, $t'''$, $f_L(t'')$, and $f_L(t''')$ all lie between 0 and $\frac{1}{2}\pi$, the arguments of the sines lie between $-\frac{1}{4}\pi$ and $\frac{1}{4}\pi$. For angles $\phi$ in this range

$$(2\sqrt{2}/\pi)|\phi| \leq |\sin\phi| \leq |\phi| . \tag{5.59}$$

One deduces from Eqs. (5.55) and (5.57)–(5.59) that

$$|f_L(t'')-f_L(t''')| \leq (21\pi/2\sqrt{2})|t''-t'''| . \tag{5.60}$$

This proves that $f_L(t)$ is continuous; it also proves Eq. (5.53). Since $f_L(t)$ is continuous and satisfies Eqs. (5.46) and (5.47), there must be at least one root $t'$ to the equation $t=f_L(t')$ for any $t$ between 0 and $\frac{1}{2}\pi$. This equation cannot have two roots $t''$ and $t'''$, for if $t=f_L(t'')=f_L(t''')$ then $d_1'=0$; by Eq. (5.55), $d_1$ must also be zero which means that $t''=t'''$. Finally, if $t=f_L(t'')$ then $t\leq t''\leq\frac{1}{2}\pi$ [using Eq. (5.48)]. Hence $F_L(t)$ [the inverse function of $f_L(t)$] satisfies (*c*). Now let $t$ and $t'$ be arbitrary parameters in the range 0 to $\frac{1}{2}\pi$. Let $t''=F_L(t)$ and $t'''=F_L(t')$. Then $t=f_L(t'')$, $t'=f_L(t''')$. Using Eqs. (5.55) and (5.57)–(5.59), one gets

$$|F_L(t)-F_L(t')| < \pi(0.68\sqrt{2})^{-1}|t-t'| , \tag{5.61}$$

which proves Eq. (5.54). Furthermore, the inequalities (5.55) and (5.56) give $d_2'\leq 4000 d_1'$, which proves Eq. (5.52). Finally, (a) is a consequence of (c), using Eq. (5.42) and the continuity of $T_A$ (Theorem 4).

The next problem is to discuss the limit of the curve $Q_L$ for $L\to\infty$. Determining the limit of $Q_L(t)$ for $L\to\infty$ with $t$ held fixed is equivalent to determining the limit of the Hamiltonians $H_N(M)$ for $M\to\infty$ holding the effective coupling constant in $H_N(M)$ fixed. It is convenient to introduce subsets $S_L$ of $S_A$ which contain $Q_L$. The set $S_0$ is the set $S_A$ itself; the definition of $S_L$ is as follows.

*Definition.* $S_L$ for $L>0$ is the set

$$S_L = T_A(S_{L-1}) . \tag{5.62}$$

$S_L$ consists of all points in $S_A$ which can be obtained by applying the transformation $T_A$ $L$ times to some point in $S_A$. Evidently all points in $S_L$ also are in $S_{L-1}$:

*Theorem 6.*

$$S_L \subset S_{L-1} \quad \text{for} \quad L\geq 1. \tag{5.63}$$

The following theorem gives an upper bound on the "cross-sectional size" of $S_L$ for given angle $\theta$.

*Theorem 7.* Let $P_A$ and $P_A'$ be any pair of points in $S_L$. Let $|P_A-P_A'| = (d_1,d_2)$. Then

$$d_2 \leq 4000 d_1 + 300\times(0.2)^L. \tag{5.64}$$

The cross section is the maximum value of $d_2$ for $d_1=0$. Theorem 7 states that the cross section goes to zero as $L\to\infty$; the spaces $S_L$ to a single curve as $L\to\infty$ (see below).

*Proof of Theorem 7.* The proof is by induction. For $L=0$ the theorem is true simply because the bounds (5.10)–(5.13) force $d_2$ to be less than 300 for any pair of points in $S_A$. Suppose the theorem is true for $S_{L-1}$. Let $P_A$ and $P_A'$ be two points in $S_L$. Let $|P_A-P_A'|$ be $(d_1',d_2')$. There must exist (by definition of $S_L$) two points $P_B$ and $P_B'$ in $S_{L-1}$ with $P_A=T_A(P_B)$, $P_A'=T_A(P_B')$. Let $|P_B-P_B'|$ be $(d_1,d_2)$. Then the distances $d_1$, $d_2$, $d_1'$, and $d_2'$ satisfy Eqs. (5.37) and (5.38). Also $d_2$ satisfies the inequality (5.64) with $L-1$ substituted for $L$. Combining these inequalities gives

$$0.34 d_1 - 0.003\times(0.2)^{L-1} \leq d_1', \tag{5.65}$$

$$d_2' \leq 1340 d_1 + 18\times(0.2)^{L-1}. \tag{5.66}$$

These inequalities can be combined to give[13]

$$d_2' \leq 4000 d_1' + 300\times(0.2)^L. \quad \text{Q.E.D.} \tag{5.67}$$

The next three theorems will be used to show that the curves $Q_L(t)$ have a limit curve $R(t)$ for $L\to\infty$. The curve $R$ is the limit of the subsets $S_L$ for $L\to\infty$. The curve $R$ has the property $T_A(R)=R$: It is an invariant subspace of the transformation $T_A$.

*Theorem 8.* Let $\{P_L\}$ be a sequence of points with $P_L\in S_L$. Denote the $\theta$ coordinate of $P_L$ by $\theta_L$. Assume that $\theta_L$ approaches a limit $\theta$ for $L\to\infty$. Define $P_L'$ to be $P_L'=T_A(P_L)$. Denote the $\theta$ coordinate of $P_L'$ by $\theta_L'$. Then

(a) $\lim_{L\to\infty} P_L$ exists (call this limit $R$);

(b) $\lim_{L\to\infty}\theta_L'=\theta'$ exists;

(c) $\lim_{L\to\infty} P_L' = T_A(R)$.

*Proof of Theorem 8.* Let $L$ be large and $K$ be even larger. Because $S_K\subset S_L$, both $P_L$ and $P_K$ are in $S_L$. Let $|P_K-P_L| = (d_1,d_2)$. Then

$$d_1 = 2|\sin\tfrac{1}{2}(\theta_K-\theta_L)| \tag{5.68}$$

and by Theorem 7

$$d_2 \leq 4000 d_1 + 300\times(0.2)^L. \tag{5.69}$$

One can make both $d_1$ and $d_2$ arbitrarily small by choosing $L$ and $K$ large enough. This is true of $d_1$ by the assumption that $\theta_L$ approaches a limit for $L\to\infty$. It is true of $d_2$ from Eq. (5.69). Hence by the Cauchy criterion the sequence $P_L$ has a limit $R$. That is, if $P_L$ has a decomposition $\theta_L$, $\mathbf{A}_{Lk}$, $C_{Lk}$, then $\theta_L$, $\mathbf{A}_{Lk}$, and $C_{Lk}$ all have limits for $L\to\infty$, and the limits $\theta$, $\mathbf{A}_k$, and $C_k$ define the point $R$.

To prove (b) and (c) consider the distances $|P_L-R| = (d_1',d_2')$ and $|P_L'-T_A(R)| = (d_1'',d_2'')$. Since Eqs. (5.68) and (5.69) hold for any $K$, they hold for the limit $K\to\infty$, giving

$$d_1' = 2|\sin\tfrac{1}{2}(\theta_L-\theta)| , \tag{5.70}$$

$$d_2' \leq 4000 d_1' + 300\times(0.2)^L. \tag{5.71}$$

One can make $d_1'$ and $d_2'$ arbitrarily small by making $L$ large enough. Therefore, because of the inequalities of Theorem 4, one can also make $d_1''$ and $d_2''$ small

[13] Throughout this paper $\leq$ means "not greater than"; there is no implication that equality can be realized.

enough by making $L$ large enough. Hence (c) is true, and (b) is a corollary of (c).

*Theorem 9.* Let $\{P_L\}$ and $\{P_L''\}$ be any two sequences satisfying $P_L \in S_L$ and $P_L'' \in S_L$. Let the $\theta$ coordinates of $P_L$ and $P_L''$ be $\theta_L$ and $\theta_L''$, respectively. Assume that the sequences $\theta_L$ and $\theta_L''$ approach the same limit $\theta$ as $L \to \infty$. Then

$$\lim_{L\to\infty} P_L = \lim_{L\to\infty} P_L''. \tag{5.72}$$

*Proof of Theorem 9.* The proof is simple. Let

$$|P_L - P_L''| = (d_1, d_2).$$

Then since $P_L$ and $P_L''$ are in $S_L$

$$d_1 = 2|\sin\tfrac{1}{2}(\theta_L - \theta_L'')|, \tag{5.73}$$

$$d_2 \leq 4000 d_1 + 300 \times (0.2)^L. \tag{5.74}$$

As $L \to \infty$, $d_1 \to 0$, and hence $d_2 \to 0$ also. Q.E.D.

*Theorem 10.*

(a) $\lim_{L\to\infty} Q_L(t) = R(t)$ exists for all $t$ in the range $0 \leq t \leq \frac{1}{2}\pi$.

(b) $\lim_{L\to\infty} f_L(t) = f(t)$ exists $(0 \leq t \leq \frac{1}{2}\pi)$.

(c) $$T_A(R(t)) = R(f(t)). \tag{5.75}$$

(d) $$f(0) = 0, \tag{5.76}$$

$$f(\tfrac{1}{2}\pi) = \tfrac{1}{2}\pi, \tag{5.77}$$

$$0 < f(t) < t \quad (0 < t < \tfrac{1}{2}\pi). \tag{5.78}$$

(e) $\lim_{L\to\infty} F_L(t) = F(t)$, where $F$ is the inverse function to $f$; also both $F(t)$ and $f(t)$ are continuous single-valued functions of $t$ defined for $0 < t < \frac{1}{2}\pi$.

(f) $$F(0) = 0, \tag{5.79}$$

$$F(\tfrac{1}{2}\pi) = \tfrac{1}{2}\pi, \tag{5.80}$$

$$t < F(t) < \tfrac{1}{2}\pi \quad (0 < t < \tfrac{1}{2}\pi). \tag{5.81}$$

*Proof of Theorem 10.* Part (a) is a consequence of Theorem 8(a). Now let $P_L = Q_L(t)$ be a sequence as in Theorem 8; define $P_L' = T_A(P_L)$ as in Theorem 8. Then $\theta_L'$ is

$$\theta_L' = f_{L+1}(t). \tag{5.82}$$

By Theorem 8(b), $\theta_L'$ has a limit; this is true for any $t$ so the function $f_L(t)$ has a limit $f(t)$ for $L \to \infty$. This proves (b). To prove (c), compare the sequence $\{P_L'\}$ with the sequence $P_L'' = Q_L(f(t))$. These two sequences satisfy the assumptions of Theorem 9. Hence they have the same limit point. By Theorem 8(c), $P_L'$ has the limit $T_A(R(t))$. By Theorem 10(a), $P_L''$ has the limit $R(f(t))$. This proves (c). To prove (d) one uses (c) and the inequality (5.31) [note that the $\theta$ coordinate of $R(t)$ is $t$ since the $\theta$ coordinate of $Q_L(t)$ is $t$ for all $L$].

To prove (e), let $t$ be arbitrary in the range $0 \leq t \leq \frac{1}{2}\pi$ and define the sequence $t_L = F_L(t)$. Let $L$ and $K (K > L)$ be large. Then $t = f_L(t_L) = f_K(t_K)$. Therefore

$$0 = f_L(t_L) - f_K(t_K) = [f_L(t_L) - f_K(t_L)] + [f_K(t_L) - f_K(t_K)]. \tag{5.83}$$

Therefore

$$|f_K(t_L) - f_K(t_K)| = |f_L(t_L) - f_K(t_L)|. \tag{5.84}$$

Now use Theorem 5(d):

$$|t_L - t_K| = |F_K(f_K(t_L)) - F_K(f_K(t_K))| \leq 40|f_K(t_L) - f_K(t_K)| = 40|f_L(t_L) - f_K(t_L)|. \tag{5.85}$$

The function $f_L(t)$ approaches $f(t)$ for $L \to \infty$ on the closed interval $0 \leq t \leq \frac{1}{2}\pi$. Hence this limit is uniform in $t$. Hence, $|f_L(t_L) - f_K(t_L)|$ is arbitrarily small for sufficiently large $L$ and $K$ irregardless of the value of $t_L$. This means that $|t_L - t_K| \to 0$ as $L$ and $K$ approach $\infty$; hence the sequence $t_L$ has a limit for $L \to \infty$. This is true for any $t$ so $F_L(t)$ has a limit $F(t)$. Since $F_L(t)$ is the inverse to $f_L(t)$, and since both $F_L$ and $f_L$ are continuous uniformly in $L$ by Theorem 5(d), $F(t)$ is the inverse to $f(t)$ and both are continuous. Also, since $F_L$ and $f_L$ are single valued, so are $F$ and $f$. Finally, (f) is a consequence of (d) and (e).

Armed with Theorems 1–10, one can now attack the renormalization problem. One starts with a sequence of unrenormalized cutoff Hamiltonians $H_M$. The bare coupling constant $g_0$ is permitted to vary with $M$ and is denoted $g_{0M}$. In addition, $H_M$ is permitted to have an additive constant $\mathcal{E}_{0M}$ also varying with $M$. The renormalization problem is to choose the sequences $g_{0M}$ and $\mathcal{E}_{0M}$ so that $H_M$ has a finite limit for $M \to \infty$. Since the number of degrees of freedom changes as $M \to \infty$, one has to specify what one means by the limit. To be precise, we demand that each energy level, counting in order of increasing energy, has a finite limit. This is equivalent to demanding that the energy levels of the effective Hamiltonians $H_N(M)$ have limits as $M \to \infty$ keeping $N$ fixed, since the effective Hamiltonians $H_N(M)$ describe the first $2^{2N+3}$ energy levels of $H_M$. The limit of $H_N(M)$ for $N$ fixed is a simpler limit since now the number of degrees of freedom is fixed. It will be found that $H_N(M)$ has a limit as an operator for $M \to \infty$ (the limit will be denoted $H_{RN}$), which ensures that the eigenvalues of $H_N(M)$ have limits. There are other parts to the renormalization problem, namely, computing matrix elements of the operators $\tau^+, \tau^-$, $a_m$, etc. between eigenstates of the renormalized Hamiltonian. These other problems will not be discussed.

The effective Hamiltonians $H_N(M)$ [with $H_M(M)$ defined to be $H_M$] are all in the space $S$. Denote the decomposition of $H_N(M)$ by $(J_N(M), \mathcal{E}_N(M), N, P_N(M))$, where, in turn, $P_N(M)$ is a point in $S_A$ with the decomposition $(\theta_N(M), \mathbf{A}_{kN}(M), C_{kN}(M))$. Denote the decomposition of the original cutoff Hamiltonians $H_M$ by $(J_{0M}, \mathcal{E}_{0M}, M, P_{0M})$; the decomposition of $P_{0M}$ is

$(\theta_{0M},0,0)$ and $J_{0M}$ and $\theta_{0M}$ are

$$J_{0M}=\Lambda^M(1+2g_{0M}^2)^{1/2}, \tag{5.86}$$

$$\theta_{0M}=\tan^{-1}(\sqrt{2}g_{0M}). \tag{5.87}$$

Since $H_N(M)$ is defined as the transform by $T$ of $H_{N+1}(M)$, one has

$$P_N(M)=T_A(P_{N+1}(M)). \tag{5.88}$$

Since $P_N(N)$ lies on the curve $Q_0$, this means $P_N(M)$ is on $Q_{M-N}$:

$$P_N(M)=Q_{M-N}(\theta_N(M)). \tag{5.89}$$

Also, one has

$$J_{N-1}(M)=\Lambda^{-1}J_N(M)T_B(P_N(M)) \tag{5.90}$$

and

$$\mathcal{E}_{N-1}(M)=\mathcal{E}_N(M)+J_N(M)T_C(P_N(M)) \tag{5.91}$$

from Eqs. (5.17) and (5.18). Finally, one has from Eqs. (5.89), (5.45), and (5.42)

$$\theta_N(M)=f_{M-N}(\theta_{N+1}(M)), \tag{5.92}$$

$$\theta_{N+1}(M)=F_{M-N}(\theta_N(M). \tag{5.93}$$

The condition $H_M(M)=H_M$ means $J_M(M)=J_M$ and $\theta_M(M)=\theta_{0M}$.

One wants to choose the sequences $\theta_{0M}$ and $\mathcal{E}_M$ so that the Hamiltonians $H_N(M)$ have a limit for $M\to\infty$. Customarily one would fix $\theta_{0M}$ and $\mathcal{E}_M$ by requiring that the renormalized coupling constant and the ground-state energy be fixed independent of $M$. We cannot calculate the renormalized coupling constant since this requires knowing the ground-state matrix element of $\tau^{\pm}$, and these matrix elements are not discussed in this paper. So a more *ad hoc* procedure will be used. Clearly if $H_N(M)$ is to approach a limit for $M\to\infty$, the sequences $\mathcal{E}_N(M)$ and $\theta_N(M)$ must approach limits as $M\to\infty$. The simplest way to ensure this is for $\mathcal{E}_N(M)$ and $\theta_N(M)$ to be independent of $M$. This cannot be true for all $N$, but it can be arranged for one value of $N$, say $N=0$. So let $\theta_0(M)$ be a constant $\theta_R$ (between 0 and $\frac{1}{2}\pi$) and let $\mathcal{E}_0(M)$ be 0.

Given $\theta_0(M)=\theta_R$ and $\mathcal{E}_0(M)=0$, for all $M$, one can reconstruct the complete double sequence $H_N(M)$. First, one computes $\theta_N(M)$, for all $M$ and $1\leq N\leq M$, using Eq. (5.93). Secondly, one computes $g_{0M}=(1/\sqrt{2})\tan\theta_M(M)$ and $J_{0M}$ from Eq. (5.86). Third, one computes all the $J_N(M)$ $(0\leq N<M)$ from $J_M(M)=J_{0M}$ and Eq. (5.90). Finally, one computes $\mathcal{E}_N(M)$ $(1\leq N\leq M)$ from Eq. (5.91). The points $P_N(M)$ are given by Eq. (5.89).

Now one can consider the limit for $M\to\infty$ of $H_N(M)$. The results are stated in Theorem 11.

*Theorem 11.* Assume $\theta_R\neq\frac{1}{2}\pi$. Then

(a) $\lim_{M\to\infty}H_N(M)=H_{RN}$ exists for all $N$.

Let $H_{RN}$ have the decomposition $(J_{RN},\mathcal{E}_{RN},N,P_{RN})$ and let the $\theta$ coordinate of $P_{RN}$ be $\theta_{RN}$. Then

(b) $P_{RN}=R(\theta_{RN})$; (5.94)

(c) $H_{RN}=T(H_{RN+1})$, i.e., (5.95)

$$\theta_{RN}=f(\theta_{RN+1}), \tag{5.96}$$

$$J_{RN}=\Lambda^{-1}J_{RN+1}T_B(P_{RN+1}), \tag{5.97}$$

$$\mathcal{E}_{RN}=\mathcal{E}_{RN+1}+J_{RN+1}T_C(P_{RN+1}), \tag{5.98}$$

$$P_{RN}=T_A(P_{RN+1}). \tag{5.99}$$

*Proof of Theorem 11.* The first step uses induction in $N$. For $N=0$, $\theta_0(M)$ has a limit $\theta_R$ for $M\to\infty$ by definition. Hence the sequence $P_0(M)$ satisfies the assumptions of Theorem 8. Hence $P_0(M)$ has a limit for $M\to\infty$; from Theorem 10, this limit is $R(\theta_R)$. Now suppose that $\theta_N(M)$ and $P_N(M)$ have limits $\theta_{RN}$ and $P_{RN}=R(\theta_{RN})$, respectively. Consider the sequence $\theta_{N+1}(M)$ as a function of $M$. It is given by Eq. (5.93). Since $\theta_N(M)$ has a limit $\theta_{RN}$, since the function $F_{M-N}(\theta)$ has a limit $F(\theta)$, and since $F_{M-N}(\theta)$ is continuous in $\theta$ uniformly in $M$ [see Theorem 5(d)], the sequence $\theta_{N+1}(M)$ must have a limit $\theta_{RN+1}$. Also $\theta_{RN+1}=F(\theta_{RN})$. Hence $\theta_{RN}$ satisfies (c). Since $\theta_{N+1}(M)$ has a limit, $P_{N+1}(M)$ has a limit (Theorem 8); the limit is $R(\theta_{RN+1})$ (Theorems 9 and 10). Because $\theta_{RN}$ is $f(\theta_{RN+1})$, one has

$$R(\theta_{RN})=T_A(R(\theta_{RN+1})) \tag{5.100}$$

[Theorem 10(c)]. By induction, one has established limits for $\theta_N(M)$ and $P_N(M)$ for all $N$, as $M\to\infty$. The limit $P_{RN}$ satisfies (b) and (c) and $\theta_{RN}$ satisfies (c). The next step is to look at the scale factors $J_N(M)$. We use Theorem 3. The points of the curve $Q_0$ satisfy the prerequisites of Theorem 3; hence all the curves $Q_L$ have the property that $A_{k1}$ vanishes for all $k$ at any point on the curve. In particular, $A_{kN1}(M)$ vanishes for all $k$. Look at $T_B(P_N(M))$. Let $P_N(M)$ be the point $P_A$ of Theorem 2. Using the notation of Theorem 2 and the result of Theorem 3,

$$T_B(P_N(M))=(m^2+2g''^2)^{1/2}, \tag{5.101}$$

$$\tan\theta'=\sqrt{2}g''/m. \tag{5.102}$$

Note that $\theta'$ is $\theta_{N-1}(M)$, $\theta$ (notation of Theorem 2) is $\theta_N(M)$, and $m=\cos\theta$. One can eliminate $g''$ to obtain

$$T_B(P_N(M))=\cos\theta_N(M)/\cos\theta_{N-1}(M). \tag{5.103}$$

Using Eqs. (5.86), (5.87), and (5.90), one obtains[14]

$$J_N(M)=\Lambda^N[\cos\theta_N(M)]^{-1}. \tag{5.104}$$

Since $\theta_N(M)$ has a limit $\theta_{NR}$ for $M\to\infty$, so does $J_N(M)$, provided $\theta_{NR}$ is not $\frac{1}{2}\pi$. But from (c) and Theorem 10(f), one sees that $\theta_{NR}<\frac{1}{2}\pi$ if $\theta_R<\frac{1}{2}\pi$. So $J_N(M)$ has a limit $J_{RN}$:

$$J_{RN}=\Lambda^N(\cos\theta_{NR})^{-1}. \tag{5.105}$$

Using Eq. (5.91), one can now show that $\mathcal{E}_N(M)$ has

[14] Equation (5.104) means that the coefficient of $(a_0^\dagger a_0+b_0^\dagger b_0-1)$ in $H_N(M)$ is simply $\Lambda^N$ independently of the value of $\theta_N(M)$. This is also an immediate consequence of Eq. (5.33) of Theorem 2.

a limit $\mathcal{E}_{RN}$ for $M\to\infty$. It is easily seen that $J_{RN}$ and $\mathcal{E}_{RN}$ satisfy (c). This completes the proof of Theorem 11.

The existence of the renormalized energies has now been proved. The renormalized theory is defined by the sequence of renormalized cutoff Hamiltonians $H_{RN}$. Because of Theorem 11(c), this sequence has a common set of eigenvalues: $H_{RN}$ describes the first $2^{2N+3}$ of these. The complete set of eigenvalues defines the complete renormalized Hamiltonian $H_R$. Unlike the renormalized Lee model, the present renormalized theory has no ghost states: The bare coupling constants $g_{0M}$ are real for all $M$ and all the Hamiltonians $H_M$ [and $H_N(M)$ and $H_{RN}$] are Hermitian. The limit of $g_{0M}$ for $M\to\infty$ is $\infty$; this is proven in Sec. VII.

To conclude this section, it will be shown that the set of renormalized Hamiltonians $H_{RN}$ is independent of the choice of the unrenormalized cutoff Hamiltonians $H_M$, in the following sense.

*Theorem 12.* Suppose that the cutoff Hamiltonians $H_M$ have the decomposition $(J_{0M},\mathcal{E}_{0M},M,P_{0M})$, where $P_{0M}$ lies on a one-parameter curve $Q_0'(t)$:

$$P_{0M}=Q_0'(\theta_{0M}). \tag{5.106}$$

Suppose that the curve $Q_0'(t)$ is any curve in the space $S_A$ defined for $0\leq t\leq\frac{1}{2}\pi$, such that $t$ is the $\theta$ coordinate of $Q_0'(t)$ and the bound (5.52) of Theorem 5(d) is satisfied by $Q_0'$. Construct the sequence of effective Hamiltonians $H_N(M)$ starting from $H_M$, and let $H_N(M)$ have the decomposition $(J_N(M),\mathcal{E}_N(M),N,P_N(M))$. The points $P_N(M)$ lie on curves $Q_{M-N}'(t)$ defined by analogy with $Q_L(t)$. Let $\theta_N(M)$ be the $\theta$ coordinate of $P_N(M)$. Let $J_N(M)$, $\mathcal{E}_N(M)$, and $\theta_N(M)$ be determined by the boundary conditions

$$\theta_0(M)=\theta_R, \tag{5.107}$$

$$\mathcal{E}_0(M)=0, \tag{5.108}$$

$$J_0(M)=(\cos\theta_R)^{-1}. \tag{5.109}$$

Then Theorem 11 holds for these $H_N(M)$ and the limiting Hamiltonians $H_{RN}$ are independent of the choice of the curve $Q_0'$.

To prove Theorem 12 one first rederives Theorems 5–11 with $Q_L'$ replacing $Q_L$; the arguments are unchanged except in Theorem 11 where the scale factors $J_N(M)$ are computed using a different boundary condition. To show that $J_N(M)$ has a limit as $M\to\infty$ one must show that $T_B(P)$ is a continuous function of $P$. This is true; the proof will be omitted.

To show that the limiting Hamiltonians $H_{RN}$ are independent of the starting curve $Q_0'$, we show that the limiting Hamiltonians $H_{RN}$ are uniquely determined by their properties, as specified in Theorem 11, plus the boundary conditions. Using Theorem 11(c), one finds

$$\theta_{RN+1}=F(\theta_{RN}). \tag{5.110}$$

Therefore one can compute $\theta_{RN}$ for all $N$ given $\theta_{R0}=\theta_R$. Then by 11(b), $P_{RN}$ is determined. Then one can use 11(c) to determine $J_{RN}$ and $\mathcal{E}_{RN}$ starting from the boundary conditions (5.108) and (5.109).

The scale factors $J_0(M)$ were specified in this discussion instead of $J_M(M)$ simply to ensure that $H_{RN}$ would be independent of the choice of curve $Q_0'$.

## VI. APPROXIMATE SCALE INVARIANCE IN RENORMALIZED THEORY

When a quantum theory is invariant to the orientation of the coordinate system, it must be rotationally symmetric—that is, there must exist unitary operators $R$ which generate rotations and which commute with the Hamiltonian. One can then diagonalize the generators of infinitesimal rotations simultaneously with the Hamiltonian; one can classify the resulting eigenstates by angular momentum eigenvalues, etc.

Likewise, when a quantum theory contains no parameters with the dimensions of energy, it must be invariant to a choice of energy scale. This immediately implies that the theory is invariant to a set of unitary operators $U(s)$ which change all energies by a scale factor $s$. The Hamiltonian $H$ is not invariant to $U(s)$, since $H$ is itself an energy; instead, one has

$$U^\dagger(s)HU(s)=sH. \tag{6.1}$$

There will be an infinitesimal generator $D$ which generates infinitesimal scale transformations (a transformation with $s=1+\epsilon$, where $\epsilon$ is infinitesimal). However, $D$ does not commute with $H$ and cannot be simultaneously diagonalized with $H$. Instead, scale invariance is used to generate a set of energy levels with any energy $sE$ given a level with energy $E$.

In field-theoretic problems there are usually mass parameters in the theory, but sometimes these parameters become negligible at high energies or short distances. For example, the propagator of a free scalar or spinor field at short distances is independent of the free-field mass and is equal to the propagator of the zero-mass theory. The free zero-mass scalar and spinor field theories are scale invariant.[15] The standard interacting field theories (quantum electrodynamics or pseudoscalar-meson theory) have only masses as dimensional parameters, but when solved in perturbation theory they do not become scale invariant at short distances (large momenta). The propagators of the interacting theories involve logarithms of $(q^2/m^2)$, where $m$ is a renormalized mass and $q$ the argument of the propagator. However, if one holds the renormalized coupling constant $e$ fixed, then at very large $q^2$ the logarithmic terms become so large that the terms of order $e^{2k}\ln^k(q^2/m^2)$ in the perturbation expansion are much larger than the Born approximation. To determine the propagator for this range of $q^2$, in particular, in the limit $q^2\to\infty$, one must sum the complete perturbation expansion. There are presently no methods for doing

[15] J. Wess, Nuovo Cimento **18**, 1086 (1960).

this (see especially the remarks of Bogoliubov and Shirkov[16]). There is then a question of whether the mass dependence will disappear at values of $q^2$ so large that the complete perturbation expansion has to be summed. The best analysis of this problem in relativistic theory is that of Gell-Mann and Low.[8]

In the model, what happens is this. The energy levels of order $\Lambda^n$ expanded in powers of the renormalized coupling constant $g_R$ have terms of order $n^k g_R{}^k$ which prevent any scaling laws from holding. But when $n$ is so large that $ng_R \gg 1$, the complete series in $g_R$ must be summed, and then the theory becomes scale invariant, in a manner to be explained below. If $g_R$ itself is of order 1 rather than small, then scale invariance sets in for much smaller $n$; the only requirement is $n \gg 1$.

There is a feature of scale transformations which distinguishes them in a very fundamental way from all other symmetries of the theory. The other symmetries (charge symmetries, etc.) are well defined in the presence of the cutoff $M$ of the model. The scale transformations are not. The scale transformations of the model are transformations $U_l$ which take the creation and destruction operators $a_m{}^\dagger$, $a_m$, $b_m{}^\dagger$, and $b_m$, for any $m$, into the operators $a_{m+l}{}^\dagger$, $a_{m+l}$, $b_{m+l}{}^\dagger$, and $b_{m+l}$. (Because the momentum continuum has been replaced by a discrete index $m$, the scale transformations are labeled by a discrete variable $l$ instead of a continuous variable $s$.) Since the creation and destruction operators satisfy the same commutation relations for any $m$, the transformation should exist and be unitary, except for end point effects. That is, in the cutoff theory there are no operators $a_m$, etc., with $m > M$, or $m < 0$. Thus the operators $a_M$, etc. cannot be transformed. To have scale transformations well defined, one must have operators $a_m$, etc. defined for $-\infty < m < \infty$. But this raises a new problem: If there are an infinite number of $a_m$, then they act in an infinite product Hilbert space, which is inseparable and therefore hard to work with.[10] This problem has not been mentioned up to now since it was evident once the unrenormalized Hamiltonian was defined that one could only solve it by introducing a cutoff $M$. Furthermore, when the limit $M \to \infty$ was defined in Sec. V, it was defined only for the effective Hamiltonians $H_N(M)$ for fixed $N$, which act on Hilbert spaces with a fixed and finite number of meson degrees of freedom.

The natural way to show that a theory has an approximate symmetry is to show that it departs only a small amount from a theory with the exact symmetry. In the present example of scale invariance, this would require constructing a version of the model which is exactly scale invariant. But this is very difficult precisely because of the problems of the infinite number of degrees of freedom. The problem is not the problem of keeping the pions with arbitrary large $m$. It was shown at the end of Sec. V that one could define a renormalized Hamiltonian $H_R$ which includes all the renormalized energy levels including those involving $m$-mesons with arbitrarily large $m$. The set of such energy levels can be ordered by their energy and therefore form a countable set of states, which one can think of as defining a separable subspace of the original inseparable space. The problem is that the exactly scale-invariant theory would have to include degrees of freedom $m$ with $m \to -\infty$. With such terms present there would be on every gross energy level an infinite sequence of fine structure, hyperfine structure, hyper-hyperfine structure, etc., with the net result that in a finite energy interval there would be an uncountable number of distinct energy levels. These would not form a continuum because each energy must be the sum of terms of order $\Lambda^{-1}$, $\Lambda^{-2}$, $\Lambda^{-3}$, etc., with coefficients of order 1.

Rather than try to develop a formalism for handling the difficulties of the inseparable space of energy levels of a scale-invariant theory, we will define approximate scale invariance to mean simply that for each energy level of $H_R$ of sufficiently large energy, there is another energy level which is approximately a factor $s_0$ larger in energy. The factor $s_0$ will be determined below; it will be of order $\Lambda$. The correspondence will not be one to one; for an energy level of energy $E$, there will be four energy levels of approximately energy $s_0 E$ due to the fact that the energy levels of energy $s_0 E$ involve one more meson degree of freedom.

One can try to predict the value of $s_0$ by considering the unrenormalized Hamiltonians $H_M$. If one applies the scaling operator $U_1$ to $H_M$ one gets

$$U_1{}^\dagger H_M(g_0) U_1 = \Lambda^{-1} H_{M+1}(g_0) - \Lambda^{-1} O_0, \qquad (6.2)$$

where $U_1$ is the operator that takes $a_m$ into $a_{m+1}$, and $O_0$ is the term of order 1 in $H_{M+1}$. Since the eigenvalues of $\Lambda^{-1}H_{M+1}(g_0) - \Lambda^{-1}O_0$ differ in order $\Lambda^{-1}$ from the eigenvalues of $\Lambda^{-1}H_{M+1}(g_0)$, it follows that $H_M(g_0)$ and $\Lambda^{-1}H_{M+1}(g_0)$ have the same eigenvalues except for fine structure of order $\Lambda^{-1}$.

Suppose that $H_M(g_0)$ had a well-defined limit as $M \to \infty$ for fixed $g_0$. Then, in particular, the energy levels of $H_M(g_0)$ and $H_{M+1}(g_0)$ would be the same for sufficiently large $M$ (excluding energies of order of the cutoff, that is, energies of order $\Lambda^M$). But then a given energy level of $H_M$ would be $\Lambda^{-1}$ times the energy of a level of $H_{M+1}$, from Eq. (6.2). For sufficiently large $M$, this level of $H_{M+1}$ is also a level of $H_M$. Hence for every energy level of $H_M$ of energy $E$, there would be another level with energy $\Lambda E$. Thus the factor $s_0$ would be $\Lambda$.

The prediction is wrong; $s_0$ is not $\Lambda$. The reason for the failure is that the renormalized energy levels are obtained by solving Hamiltonians $H_M(g_{0M})$ where $g_{0M}$ changes with $M$. It will be shown later that $g_{0M} \to \infty$ as $M \to \infty$; therefore even for large $M$, $g_{0M}$ is not constant.

The idea that operators do not scale as predicted from an unrenormalized theory was used in a recent

[16] N. N. Bogoliubov and V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959), pp. 528–529.

discussion of approximate scale invariance in strong interactions.[6] However, the analogy to the model of this paper is inexact since in the strong interaction problem, the scaling law for the Hamiltonian is fixed by general arguments; it is the other fields in the theory, such as the pion field, whose scaling laws (dimensions) were permitted to be arbitrary.

The remainder of this section is devoted to the technical problem of computing the nature of the energy levels with energies of order $\Lambda^n$ with $n$ large, and extracting the scale factor $s_0$. It will be shown not only that these energies scale by a factor $\Lambda\beta^{-1}$, where $\beta$ is approximately $\frac{1}{2}$, but also that the error to this scaling law itself scales like $\Lambda$, as if the Hamiltonian consisted of two terms, one scaling as $\Lambda\beta^{-1}$, the other as $\Lambda$ under a scale transformation [cf. Eqs. (6.24) and (6.25)].

In the following, it is assumed that the function $f(t)$ and the "curve" $R(t)$ defined in Sec. V are differentiable. I have not proved this.

The renormalized theory is defined by a sequence of Hamiltonians $H_{RN}$. These Hamiltonians are determined by the three parameters $J_{RN}$, $\mathcal{E}_{RN}$, and $\theta_{RN}$. We study $H_{RN}$ when $N$ is large. This requires knowledge of $J_{RN}$, $\mathcal{E}_{RN}$, and $\theta_{RN}$ for large $N$.

First look at the sequence $\{\theta_{RN}\}$. Since $\theta_{RN}$ is the $\theta$ coordinate of $P_{RN}$, and since $P_{RN}$ is the transform $T_A$ of $P_{RN+1}$, one can apply the inequality (5.31) to obtain

$$(1-0.51\sin^2\theta_{RN+1})\tan\theta_{RN+1}\leq\tan\theta_{RN}\leq(1-0.48\sin^2\theta_{RN+1})\tan\theta_{RN+1}. \quad (6.3)$$

First of all, this implies that

$$\theta_{RN}<\theta_{RN+1}<\tfrac{1}{2}\pi \quad (6.4)$$

(we assume $\theta_{R0}<\frac{1}{2}\pi$, which then forces $\theta_{RN}$ to be less than $\frac{1}{2}\pi$; see the proof of Theorem 11). Thus $\{\theta_{RN}\}$ is an increasing and bounded sequence. Therefore it has a limit for $N\rightarrow\infty$. The limit must be $\frac{1}{2}\pi$. The reason is that since $\theta_{RN}=f(\theta_{RN+1})$, the limit $\theta$ must satisfy $\theta=f(\theta)$. Also, $\theta_{R0}<\theta\leq\frac{1}{2}\pi$. But from Theorem 10(d), the only such $\theta$ is $\theta=\frac{1}{2}\pi$. Therefore, when $N$ is sufficiently large, $\theta_{RN}$ is approximately $\frac{1}{2}\pi$. Write

$$\theta_{RN}=\tfrac{1}{2}\pi-\phi_N. \quad (6.5)$$

When $\phi_N$ is small, the inequality (6.3) is approximately

$$0.49(\phi_{N+1})^{-1}\leq\phi_N^{-1}\leq0.52(\phi_{N+1})^{-1}, \quad (6.6)$$

e.g.,

$$\phi_{N+1}\approx\tfrac{1}{2}\phi_N. \quad (6.7)$$

To be more precise, consider the formula $\theta_{RN}=f(\theta_{RN+1})$ and expand in powers of $\phi_{N+1}$:

$$\tfrac{1}{2}\pi-\phi_N=f(\tfrac{1}{2}\pi-\phi_{N+1})=f(\tfrac{1}{2}\pi)-\phi_{N+1}f'(\tfrac{1}{2}\pi)+O(\phi_{N+1}^2). \quad (6.8)$$

Since $f(\frac{1}{2}\pi)$ is $\frac{1}{2}\pi$, one gets

$$\phi_N=f'(\tfrac{1}{2}\pi)\phi_{N+1}+O(\phi_{N+1}^2), \quad (6.9)$$

and Eq. (6.7) shows that $f'(\frac{1}{2}\pi)$ is approximately 2. Let

$$\beta=[f'(\tfrac{1}{2}\pi)]^{-1}\simeq0.5. \quad (6.10)$$

One can rewrite Eq. (6.9) to read

$$\phi_{N+1}=\beta\phi_N+O(\phi_{N+1}^2). \quad (6.11)$$

An analysis of this equation shows that

$$\phi_N=a\beta^N+O(\phi_N^2), \quad (6.12)$$

where $a$ is a constant ($a$ will depend on $\theta_{R0}$).

Now look at $J_{RN}$, $\mathcal{E}_{RN}$, and $P_{RN}$. From Eq. (5.105), assuming $N$ is large, one has

$$J_{RN}=\Lambda^N(\cos\theta_{RN})^{-1}\simeq\Lambda^N\phi_N^{-1}\simeq\Lambda^N a^{-1}\beta^{-N}. \quad (6.13)$$

To compute $P_{RN}$, one must study the curve $R(t)$. One has

$$R(\theta_{RN})\simeq R(\tfrac{1}{2}\pi)-\phi_N R'(\tfrac{1}{2}\pi). \quad (6.14)$$

Let

$$P_c=R(\tfrac{1}{2}\pi), \quad (6.15)$$

$$P_d=-R'(\tfrac{1}{2}\pi). \quad (6.16)$$

Then

$$P_{RN}\simeq P_c+a\beta^N P_d. \quad (6.17)$$

Finally, from Theorem 11(c) one has

$$\mathcal{E}_{RN}=-\sum_{n=1}^{N}J_{Rn}T_c(P_{Rn}) \quad (6.18)$$

(using the definition $\mathcal{E}_{R0}=0$). The dominant terms in this sum are for large $n$ since $J_{Rn}\sim\Lambda^n$ and $T_c\sim1$ [Eq. (5.30)]. For large $n$, $P_{Rn}\simeq P_c$. Let

$$\gamma=T_c(P_c). \quad (6.19)$$

Then for large $N$ [using Eq. (6.13)],

$$\mathcal{E}_{RN}\simeq-\Lambda^N\gamma a^{-1}\beta^{-N}\Lambda(\Lambda-\beta)^{-1}. \quad (6.20)$$

A more careful calculation gives the first correction to Eq. (6.20) as

$$\mathcal{E}_{RN}\simeq-\Lambda^N\beta^{-N}\gamma a^{-1}\Lambda(\Lambda-\beta)^{-1}+\Lambda^N\gamma_1\Lambda(\Lambda-1)^{-1}, \quad (6.21)$$

where $\gamma_1$ is a constant; also,

$$J_{RN}\simeq a^{-1}\beta^{-N}\Lambda^N+\gamma_2\Lambda^N, \quad (6.22)$$

where $\gamma_2$ is a constant.

With the above approximations for $P_{RN}$, $J_{RN}$, and $\mathcal{E}_{RN}$, one can write

$$H_{RN}=a^{-1}\Lambda^N\beta^{-N}\mathcal{H}_{cN}+\Lambda^N\mathcal{H}_{dN}, \quad (6.23)$$

where $\mathcal{H}_{cN}$ is a Hamiltonian with the decomposition $(J,\mathcal{E},N,P)=(1,-\gamma\Lambda(\Lambda-\beta)^{-1},N,P_c)$, and $\mathcal{H}_{dN}$ is a Hamiltonian with the decomposition $(1,\gamma_1\Lambda(\Lambda-1)^{-1},N,\gamma_2P_c+P_d)$. The only $N$ dependence of $\mathcal{H}_{cN}$ and $\mathcal{H}_{dN}$ is in terms of how many degrees of freedom are kept in Eq. (5.2), since neither the $J$, $\mathcal{E}$, nor $P$ components of $\mathcal{H}_{cN}$ or $\mathcal{H}_{dN}$ depend on $N$.

Now compare $H_{RN}$ with $H_{RN+1}$. The difference between $\mathcal{H}_{cN}$ and $\mathcal{H}_{cN+1}$ is only in terms containing

$a_{N+1}$, $b_{N+1}$, etc., and such terms are of order $\Lambda^{-N-1}$ or less. So the energy levels of $\mathcal{H}_{cN}$ are approximately the same as the energy levels of $\mathcal{H}_{cN+1}$, only each level of $\mathcal{H}_{cN}$ corresponds to four of $\mathcal{H}_{cN+1}$ because of the extra degrees of freedom in $\mathcal{H}_{cN+1}$. In $H_{RN}$, the $\mathcal{H}_{cN}$ term dominates the $\mathcal{H}_{dN}$ term by a factor $\beta^N$; neglecting the $\mathcal{H}_{dN}$ term, the energy levels of $\beta^{-1}\Lambda H_{RN}$ and $H_{RN+1}$ are approximately equal. This establishes the basic claim of this section. The energy levels of $H_{RN}$ and $H_{RN+1}$ are both subsets of the energy levels of $H_R$. So the energy levels of $\Lambda\beta^{-1}H_R$ are approximately equal to the energy levels of $H_R$. In scaling the Hamiltonian an extra factor $\beta$ has appeared.

Let us consider the errors in approximate scale invariance. One is comparing $\Lambda\beta^{-1}H_{RN}$ with $H_{RN+1}$. The basic energy scale for these Hamiltonians is $\Lambda^{N+1}\beta^{-N-1}$. Neglecting the $\mathcal{H}_{dN}$ and $\mathcal{H}_{dN+1}$ terms in $H_{RN}$ and $H_{RN+1}$ means that one has an error of order $\Lambda^N$. This is small by a factor $\beta^{N+1}$ from the basic energy scale but huge on an absolute scale (remember $N$ must be large for all our approximations to hold). There is also an error which is of order $\Lambda^{-N-1}$ in $\mathcal{H}_{cN+1}$ when one neglects the $a_{N+1}$ terms; this becomes an error of order $\beta^{-N-1}$ in $H_{RN+1}$ which is negligible compared to $\Lambda^N$ ($\beta\sim\frac{1}{2}$ while $\Lambda>4\times10^6$). Owing to the error of order $\Lambda^N$, the matching between $H_{RN+1}$ and $\beta^{-1}\Lambda_{HRN}$ is close only for energy levels with energies large compared to $\Lambda^N$, i.e., only highly excited states.

One can now get a scaling law for the leading correction to scale invariance. That is, one can take $\mathcal{H}_{dN}$ into account but still neglect the difference between $\mathcal{H}_{cN}$ and $\mathcal{H}_{cN+1}$ and the difference between $\mathcal{H}_{dN}$ and $\mathcal{H}_{dN+1}$. In this case one can write

$$H_{RN}=H_{cN}+H_{dN}, \tag{6.24}$$

with $H_{cN}=a^{-1}\Lambda^N\beta^{-N}\mathcal{H}_{cN}$, and $H_{dN}=\Lambda^N\mathcal{H}_{dN}$; then

$$H_{RN+1}\simeq\Lambda\beta^{-1}H_{cN}+\Lambda H_{dN}. \tag{6.25}$$

Since $H_{dN}$ is small compared to $H_{cN}$, the energies of $H_{cN}+H_{dN}$ consist, to a first approximation, of energies of $H_{cN}$ plus expectation values of $H_{dN}$. The correction therefore scales by a factor $\Lambda$ when $N\to N+1$ while the dominant term in the energy scales by $\Lambda\beta^{-1}$.

The unrenormalized Hamiltonian had two parts, the free meson energy term and the interaction term, but both parts scaled by $\Lambda$ when $N\to N+1$. The renormalized Hamiltonian also has two parts to a first approximation but the two parts scale differently, the dominant term scaling by $\Lambda\beta^{-1}$ while the leading correction scales by $\Lambda$.

It was crucial for the proof of scale invariance that the constants $\theta_{RN}$ approach a limit $\frac{1}{2}\pi$ for $N\to\infty$. As long as $\theta_{RN}$ changes with $N$, the energy-level structure of $H_R$ on the scale $\Lambda^N$ will differ by more than a scale factor from the structure on the scale $\Lambda^{N+1}$. This is due to the nontrivial dependence of the energy levels of $H_{RN}$ on $\theta_{RN}$. In particular, in perturbation theory, when $\theta_{RN}$ is small, the change from $\theta_{RN}$ to $\theta_{RN+1}$ is nonnegligible in order $\theta_{R0}^3$ (see Sec. VII for details). Hence in third order or higher in $\theta_{R0}$, $H_R$ does not show scale invariance. It is only when $N$ is so large that $\theta_{RN}\simeq\frac{1}{2}\pi$ that scale invariance becomes apparent; but for these values of $N$ an expansion in $\theta_{R0}$ is absurd even if $\theta_{R0}$ is small: The true expansion parameter turns out to be $(\sqrt{N})\theta_{R0}$, which is huge, instead of $\theta_{R0}$.

## VII. RENORMALIZATION AND ROLE OF TRANSFORMATION $T$

The renormalization program carried out in this paper followed the conventional pattern in that a renormalized coupling constant was defined and held fixed in the limit of infinite cutoff. The transformations $T$ and $T_A$ were introduced as part of the technique of solving the cutoff Hamiltonians; their properties were useful in proving the existence of the renormalized Hamiltonian. An analysis of the renormalization program of Sec. V shows that the transformations $T$ and $T_A$ play a more fundamental role in the renormalization than one might think. In Sec. VII A it is shown that the renormalized Hamiltonian is determined more by the properties of the transformation $T_A$ than by properties of the original unrenormalized Hamiltonian of Sec. II. In Sec. VII B, the problem of "why renormalization?" is considered; it is shown that three features of the model Hamiltonian cause the renormalization program to be nontrivial. These three features are (1) the model has an infinite number of degrees of freedom, (2) the $m$th degree of freedom with $m$ large dominates the degrees of freedom with $m$ small, and (3) scale invariance makes the behavior of the degrees of freedom for large $m$ similar for different $m$. In Sec. VII C, the renormalization theory of this paper is compared with the theory of Gell-Mann and Low for quantum electrodynamics.[8]

### A. Renormalization and Transformation $T_A$

The analysis of the renormalization program to be given here concerns very basic questions. To set the stage for these questions, it is worth reviewing the role of the Hamiltonian in ordinary quantum mechanics. In nonrelativistic quantum mechanics, a system is well defined once the Hamiltonian is specified. Any Hermitian (self-adjoint) Hamiltonian defines a unique and acceptable quantum mechanics. To specify the Hamiltonian, one must first define the basic observables of the system (e.g., position, momentum, or spin operators). Then one specifies the Hamiltonian as a function of these observables. In principle one could define the Hamiltonian in a different way, by giving a list of its eigenvalues and eigenvectors. This is rarely done in practice because the eigenvalues and eigenvectors are generally very complicated expressions, often not expressible in closed form. In contrast, the

Hamiltonian is often a simple function of the observables (for example, compare the Coulomb Hamiltonian of the helium atom with its eigenvalues and eigenvectors).

In Sec. II we defined a model quantum theory in an entirely conventional manner. The "observables" $a_n$, $a_n^\dagger$, $b_n$, $b_n^\dagger$, and $\tau^\pm$ were defined, and the Hamiltonian written as a simple function of these observables, with one free parameter $g_0$. Then in Secs. IV and V the techniques for solving the model were defined, and it was shown that after renormalization the theory had finite eigenvalues. The finite theory again depended on one free parameter, which, however, was the renormalized constant $\theta_{R0}$ instead of $g_0$.

The construction of the renormalized Hamiltonian in Sec. V was a complicated process. In summary, one chose a renormalized coupling constant $\theta_{R0}$. One constructed a sequence of Hamiltonians $H_{RN}$ by starting with the point $P_{R0}=R(\theta_{R0})$ and constructing the sequence $P_{RN}$ through the relation $P_{RN}=T_A(P_{RN+1})$. The full renormalized Hamiltonian consisted of a limit of $H_{RN}$ for $N\to\infty$ suitably defined. This construction leaves unclarified some fundamental questions. Does the renormalized theory solve the unrenormalized Hamiltonian of Sec. II? If not, what problem does it solve? Is the renormalized coupling constant a fundamental parameter in the theory? If not, can it be replaced by one that is? Is the unrenormalized Hamiltonian the simple expression which underlies and defines the rather complicated spectrum of renormalized energy levels; if not, where do we look for simplicity?

It is difficult to answer these questions conclusively because there are problems of interpretation. For example, one must decide what is a "fundamental" parameter, and what is "simple." However, in trying to answer the questions of the previous paragraph, two results become clear. The first is that the relation of the unrenormalized, uncutoff Hamiltonian to the renormalized energy levels is fundamentally different than the relation of a simple Coulomb Hamiltonian to its eigenvalues. How to characterize the new relationship can be debated, but certainly it is not the old and comfortable relationship of elementary quantum mechanics. The second result is this: There is a key fact which must figure in any discussion of the new relationship of Hamiltonian to energy levels, a key idea which must be used to obtain any fundamental understanding of why we must introduce an essentially phenomenological parameter (the renormalized coupling constant) in defining the renormalized theory. The crucial fact is the existence of a fixed point of the transformation $T_A$, namely, the point $P_c=R(\frac{1}{2}\pi)$. The point $P_c$ has already been encountered in Sec. VI: It is the limit of the points $P_{RN}$ (involved in the definition of $H_{RN}$) as $N\to\infty$. The role of the fixed point cannot be summarized in a few words; a detailed analysis of its function will be given later in this section.

The relation of the unrenormalized uncutoff Hamiltonian to the renormalized theory can be summarized in terms of the following two results which will be proven later in this section.

(1) If $\{g_{0M}\}$ is a sequence of coupling constants which approach a finite limit $g_0$ as $M\to\infty$, then the energy levels of the unrenormalized cutoff Hamiltonians $H_M(g_{0M})$ approach the energy levels of the uncutoff *free* Hamiltonian [Eq. (2.1) with $g_0=0$] as $M\to\infty$, except for an additive constant.

(2) If $\{g_{0M}\}$ is a sequence of coupling constants which approach $\infty$ as $M\to\infty$, the energy levels of $H_M(g_{0M})$ may or may not approach a limit as $M\to\infty$. For any $\theta_{R0}$ with $0<\theta_{R0}<\frac{1}{2}\pi$, there exists a sequence $\{g_{0M}\}$ with $g_{0M}\to\infty$ as $M\to\infty$, such that the energy levels of $H_M(g_{0M})$ approach the energy levels of the renormalized Hamiltonian $H_R(\theta_{R0})$ as $M\to\infty$ (apart from an additive constant).

The first result means that if the uncutoff unrenormalized Hamiltonian with finite $g_0$ is defined as a limit of cutoff Hamiltonians, then its solution is the same as the solution of the free uncutoff Hamiltonian and, in particular, is not related to any of the renormalized theories with interaction. The second result means that a single uncutoff unrenormalized Hamiltonian, the one with $g_0=\infty$, has an infinite number of possible solutions depending on what sequence $\{g_{0M}\}$ is used in the cutoff Hamiltonians. Therefore instead of each renormalized Hamiltonian corresponding to a separate unrenormalized Hamiltonian, one finds that all the renormalized Hamiltonians solve a single unrenormalized Hamiltonian. The nonuniqueness of the solution of the unrenormalized Hamiltonian with $g_0=\infty$ is discussed further below.

Now the results quoted above will be proven. It is helpful to prove the following. If $\theta<\theta_1$ and both lie between 0 and $\frac{1}{2}\pi$, then

$$f_L(\theta)<f_L(\theta_1)\quad(\text{for }\theta<\theta_1). \tag{7.1}$$

The proof is based on Theorem 5. From 5(b), $f_L(\theta_1)-f_L(0)$ is positive for $\theta_1>0$. From 5(d) [Eq. (5.54)]

$$|f_L(\theta_1)-f_L(\theta)|>0.025|\theta_1-\theta|. \tag{7.2}$$

From 5(b), $F_L(\theta)$ is continuous in $\theta$. Hence $f_L(\theta_1)-f_L(\theta)$ cannot change sign anywhere in the range $0\leq\theta<\theta_1$. Hence Eq. (7.1) holds. To prove the first result, consider a sequence $\{g_{0M}\}$ with a finite limit $g_0$ as $M\to\infty$. Consider the unrenormalized Hamiltonians $H_M(g_{0M})$. Using the transformation $T$ one can generate effective Hamiltonians $H_N(M)$ with coupling constants $\theta_N(M)$ having the same energy levels as $H_M(g_{0M})$. The constants $\theta_N(M)$ satisfy Eqs. (5.92) and (5.93) and

$$\tan\theta_M(M)=\sqrt{2}g_{0M}. \tag{7.3}$$

Let $\theta$ be an upper bound to $\theta_M(M)$; since $g_{0M}$ has a finite limit, one can choose $\theta$ to be less than $\frac{1}{2}\pi$. Define

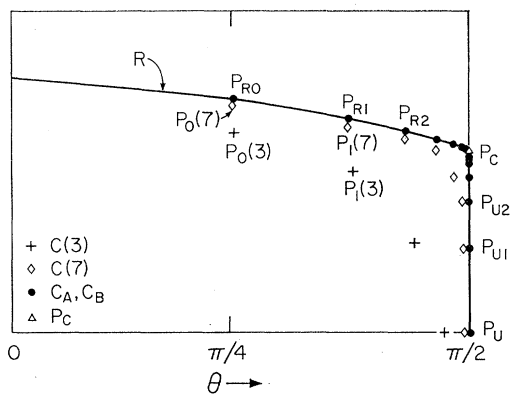


FIG. 1. Artist's conception of the trajectories $C(3)$, $C(7)$, $C_A$, and $C_B$ projected on a two-dimensional space. The renormalized coupling constant is $\frac{1}{4}\pi$. The curve $R$ is also shown. The first few points on $C_A$, $C_B$, $C(3)$, and $C(7)$ are labeled explicitly: $P_U$ is the first point on $C_A$; $P_{R0}$ is the first point on $C_B$.

a sequence $\{\theta_L\}$ to be $\theta_0=\theta$, $\theta_L=f_L(\theta_{L-1})$. Because of Eq. (7.1), $\theta_{M-N}$ is an upper bound for $\theta_N(M)$. The sequence $\{\theta_L\}$ is a decreasing sequence with limit 0 as $L\to\infty$; this follows from the inequality (5.31). Hence $\theta_N(M)\to 0$ as $M\to\infty$ for fixed $N$. Hence in the limit $M\to\infty$, $H_N(M)$ becomes a free Hamiltonian, which is result (1). To prove the second result, consider the sequence $\{g_{0M}\}$, defined in Sec. V following Eq. (5.93), corresponding to a given nonzero renormalized constant $\theta_{R0}$. Again one has constants $\theta_N(M)$ satisfying Eqs. (5.92), (5.93), and (7.3), but now $\theta_0(M)$ is fixed to be $\theta_{R0}$. From Eq. (5.31), $\theta_1(M+1)>\theta_0(M+1)=\theta_0(M)$; using Eq. (7.1) repeatedly, one gets $\theta_{N+1}(M+1)>\theta_N(M)$ for all $N$, and hence $g_{0M+1}>g_{0M}$. Thus $\{g_{0M}\}$ is an increasing sequence. It cannot have a finite upper bound, for if it did, $\theta_{R0}$ would have to be zero. Hence $g_{0M}\to\infty$ as $M\to\infty$. By the analysis of Sec. V the Hamiltonians $H_N(M)$ have well-defined limits as $M\to\infty$. [In Sec. V the unrenormalized Hamiltonians $H_M$ have a ground-state energy subtraction; if this subtraction is not made, then only the energy differences of levels of $H_N(M)$ have a limit as $M\to\infty$.] Such a sequence $\{g_{0M}\}$ exists for any $\theta_{R0}$, so result (2) is proved.

The fact that the uncutoff Hamiltonian with $g_0=\infty$ has an infinite number of solutions can be blamed on the fixed point $P_c$ of $T_A$. This result can be seen by studying the behavior of the double sequence $P_N(M)$ of points in $S_A$ defined in Sec. V as part of the renormalization analysis. The points $P_N(M)$ have the following properties.

(a) $P_M(M)$ has the decomposition $(\theta_M(M),0,0)$, i.e., the components $\mathbf{A}_k$ and $C_k$ are all zero. The point $P_M(M)$ corresponds to the unrenormalized Hamiltonian $H_M(g_{0M})$ with finite cutoff $M$ and $g_{0M}$ given by Eq. (7.3).

(b) $P_0(M)$ has $\theta$ coordinate $\theta_{R0}$, by definition.

(c) $P_{N-1}(M)=T_A[P_N(M)]$.

When $M\to\infty$, $\theta_M(M)\to\frac{1}{2}\pi$, so $P_M(M)$ has a limit $(\frac{1}{2}\pi,0,0)$ when $M\to\infty$. Denote this point by $P_U$. The point $P_U$ corresponds to the unrenormalized, uncutoff Hamiltonian with $g_0=\infty$.

The point $P_c=R(\frac{1}{2}\pi)$ (the fixed point of $T_A$) also has $\theta$ coordinate $\frac{1}{2}\pi$, but it is easily seen that the components $\mathbf{A}_k$ and $C_k$ of $P_c$ cannot vanish. Hence $P_c$ is distinct from $P_U$.

One can think of the points $P_N(M)$, for fixed $M$, as defining a trajectory $C(M)$. If one takes the limit of the trajectories $C(M)$ for $M\to\infty$, one gets a double trajectory $C_A\oplus C_B$. The trajectory $C_A$ goes from $P_U$ to $P_c$, i.e., it connects the point $P_U$ representing the unrenormalized Hamiltonian to the fixed point $P_c$. The trajectory $C_B$ connects the renormalized point $P_{R0}$ to the fixed point $P_c$. The first trajectory is an infinite sequence of points $(P_U,P_{U1},P_{U2},\ldots)$, all with $\theta=\frac{1}{2}\pi$, satisfying $P_{UN}=T_A(P_{UN-1})$, and with the limit $P_c$ as $N\to\infty$. The trajectory $C_B$ consists of the renormalized points $P_{RN}$ lying on the curve $R$, again with limit $P_c$ as $N\to\infty$. The trajectories $C(M)$ with $M$ large lie close to the limiting trajectories: The first few points on $C(M)$ [e.g., $P_M(M)$, $P_{M-1}(M)$, etc.] lie close to the first few points on $C_A$. The last few points on $C(M)$ [e.g., $P_1(M)$, $P_2(M)$, etc.] lie close to the first few points on $C_B$. The points near the middle of the trajectory $C(M)$ [e.g., $P_{M/2}(M)$] all lie close to $P_c$.

The trajectories $C(M)$, $C_A$, and $C_B$ are illustrated in Fig. 1. Figure 1 is an artist's conception of what these trajectories might look like if the space $S_A$ was a two-dimensional space instead of an infinite-dimensional space. The two dimensions are $\theta$ and a coordinate $x$ replacing the infinite-dimensional space defined by the sequences $\{\mathbf{A}_k\}$ and $\{C_k\}$. One can see explicitly in Fig. 1 that the points $P_N(M)\to P_{RN}$ as $M\to\infty$ and $P_{M-N}(M)\to P_{UN}$ as $M\to\infty$. One can also see the clustering of points about $P_c$.

Now return to the problem of the infinite number of solutions of the unrenormalized Hamiltonian. The nonuniqueness is connected with the fixed point $P_c$, because the limiting trajectory $C_A\oplus C_B$ is nonunique only on the section $C_B$. The trajectory $C_A$ connecting $P_U$ to $P_c$ is uniquely determined by $P_U$ and the recursion formula $P_{UN}=T_A(P_{UN-1})$. The trajectory $C_B$ connecting $P_c$ to $P_{R0}$ is nonunique; it is a different trajectory for each different value of $\theta_{R0}$. Thus the nonuniqueness arises at the point $P_c$.

The next question is: How is the nonuniqueness related to the properties of the fixed point $P_c$. In order to discuss this question it is necessary to know the behavior of the transformation $T_A$ in the neighborhood of $P_c$; this behavior will now be investigated.

Assume that the transformation $T_A$ is differentiable in the vicinity of $P_c$, so that if $P$ is any point near $P_c$, one can write

$$T_A(P)=P_c+U_A(P-P_c)+\text{order }(P-P_c)^2, \quad (7.4)$$

where $U_A$ is a *linear* transformation. Now consider a trajectory of points $P_N$, namely, a sequence of points satisfying

$$P_{N+1}=T_A(P_N), \tag{7.5}$$

and suppose that the trajectory lies in the vicinity of $P_c$. Then approximately,

$$P_{N+1}-P_c=U_A(P_N-P_c). \tag{7.6}$$

Consider therefore the trajectories defined by $U_A$, that is, sequences of points $Q_N$ satisfying

$$Q_{N+1}=U_A(Q_N). \tag{7.7}$$

Since this is a linear equation, an arbitrary solution can be written as a linear combination of a set of linearly independent "basic" solutions $Q_{N\alpha}$ ($\alpha=1, 2, 3, \ldots$ labels different linearly independent trajectories). The simplest type of solution is of the form

$$Q_{N\alpha}=Q_{0\alpha}(r_\alpha)^N, \tag{7.8}$$

where $Q_{0\alpha}$ is a point (determined up to a scale factor) and $r_\alpha$ is a constant. $Q_{0\alpha}$ is an eigenvector of the transformation $U_A$,

$$r_\alpha Q_{0\alpha}=U_A(Q_{0\alpha}), \tag{7.9}$$

and $r_\alpha$ is an eigenvalue. Since $U_A$ does not have to be a self-adjoint transformation, the eigenvalues $r_\alpha$ need not be real; also, there may be trajectories $Q_{N\alpha}$ behaving as $N(r_\alpha)^N$, $N^2(r_\alpha)^N$, etc., under special circumstances. Since $U_A$ is a transformation on a space with an infinite number of dimensions, there will be an infinite set of basic solutions $Q_{N\alpha}$. These solutions divide into three possible categories. Those with $|r_\alpha|>1$ are called "unstable" trajectories; these trajectories move away from $P_c$ as one keeps applying the transformation $T_A$. Those with $|r_\alpha|<1$ are stable trajectories; the stable trajectories approach $P_c$ as one keeps applying $T_A$. For example, the trajectory $C_A$ connecting $P_U$ with $P_c$ is a stable trajectory; the trajectory $C_B$ is an unstable trajectory. There can also be "neutral" trajectories with $|r_\alpha|=1$, in special cases.

A crucial question is that of how many linearly independent unstable trajectories $U_A$ has. The answer is one; the proof is as follows. There must be at least one basic unstable trajectory, for if all the basic trajectories were stable then all linear combinations of the basic trajectories would also be stable, i.e., all solutions of Eq. (7.6) would be stable. But we know there are unstable solutions, namely, the trajectories $C_B$ for any $\theta_{R0}$ (to be precise, the parts of these trajectories lying near $P_c$). On the other hand, there cannot be more than one basic unstable trajectory. For if there were two linearly independent unstable trajectories, say $Q_{N1}$ and $Q_{N2}$, then one could form a linear combination of these, say $\beta_1Q_{N1}+\beta_2Q_{N2}$, such that the $\theta$ coordinate of $\beta_1Q_{11}+\beta_2Q_{12}$ is 0. This means the $\theta$ coordinate of $(P_c+\beta_1Q_{11}+\beta_2Q_{12})$ is $\frac{1}{2}\pi$. But now the $\theta$ coordinate of $P_c+\beta_1Q_{N1}+\beta_2Q_{N2}$ will be $\frac{1}{2}\pi$ for all $N$ because $T_A$ does not change $\theta$ if $\theta=\frac{1}{2}\pi$. But then the sequence of points $P_c+\beta_1Q_{N1}+\beta_2Q_{N2}$ must approach $P_c$ as $N\to\infty$, using Theorems 8–10 of Sec. V. This means $\beta_1Q_{N1}+\beta_2Q_{N2}$ is a stable trajectory. Then we could use $\beta_1Q_{N1}+\beta_2Q_{N2}$ as a basic trajectory instead of $Q_{N2}$, for example, which leaves only one unstable trajectory. The trajectories $C_B$ for different $\theta_{R0}$ must all be multiples of the single unstable trajectory. This result has already been demonstrated in Sec. VI [see Eqs. (6.12) and (6.17)].

It will now be shown that the number of linearly independent unstable trajectories of $U_A$ determines the number of free parameters in the renormalized Hamiltonian. In other words, the degree of nonuniqueness of the solution of the unrenormalized Hamiltonian is determined by the number of unstable solutions of the linearized transformation $U_A$.

To show this, we must discuss what would have happened if $U_A$ had two or more linearly independent unstable trajectories. It will be shown that in this case the nonuniqueness of the solution of the unrenormalized Hamiltonian involves two or more free parameters. To be precise, we show that one can construct sequences $P_N(M)$ such that

$$(1)\quad \lim_{M\to\infty} P_M(M)=P_U,$$

$$(2)\quad \lim_{M\to\infty} P_N(M)=P_{RN}(a_1\cdots a_k),$$

$$(3)\qquad P_{N-1}(M)=T_A(P_n(M)),$$

where the point $P_{RN}$ depends on $k$ phenomenological parameters $a_1\cdots a_k$, $k$ being the number of linearly independent unstable solutions of $U_A$. Having shown that such sequences exist for any choice of the parameters $a_1\cdots a_k$, it is clear that there is a $k$-parameter family of renormalized Hamiltonians, defined by the points $P_{RN}(a_1\cdots a_k)$ for all $N$, all of which can be considered solutions of the single unrenormalized Hamiltonian $P_U$.

To prove the existence of the sequences $P_N(M)$, it is sufficient to consider the part of the sequence lying near $P_c$, say, the points $P_N(M)$ with

$$L<N<M-L,$$

where $L$ is large but held fixed as $M\to\infty$. So long as

$$(1')\quad \lim_{M\to\infty} P_{M-L}(M)=P_{UL}$$

($P_{UL}$ is the $L$th point on the trajectory $C_A$),

$$(2')\qquad \lim_{M\to\infty} P_L(M)=P_{RL}(a_1\cdots a_k),$$

one can reconstruct the remainders of the sequences using $T_A$ or $T_A^{-1}$ and satisfy the original requirements. If $L$ is large enough, $P_{UL}$ and $P_{RL}$ will be near $P_c$ and we can assume that

$$(3')\qquad P_{N-1}(M)=P_c+U_A(P_N(M)-P_c).$$

Since $P_N(M)-P_c$ satisfies the linearized equation, it must be a linear combination of the basic solutions for each $M$:

$$P_N(M)-P_c=\sum_\alpha \beta_\alpha(M)Q_{M-N\alpha}. \tag{7.10}$$

($Q$ depends on $M-N$ rather than $N$ so that the index of $Q$ increases as one applies $U_A$.) The sequence $P_{UN}$ must also be a linear combination of the basic solutions:

$$P_{UN}=\sum_\alpha \gamma_\alpha Q_{N\alpha}+P_c. \tag{7.11}$$

Furthermore, since $P_{UN}\to P_c$ as $N\to\infty$, the coefficients $\gamma_\alpha$ must be zero for all unstable trajectories. Suppose, to be specific, that the unstable trajectories correspond to $1\leq\alpha\leq k$ and that the trajectories for $\alpha>k$ are stable. Then $\gamma_\alpha=0$ for $\alpha\leq k$. The requirement that $P_{M-L}(M)\to P_{UL}$ as $M\to\infty$ means that $\beta_\alpha(M)$ must satisfy

$$\lim_{M\to\infty}\beta_\alpha(M)=\gamma_\alpha. \tag{7.12}$$

The requirement that $P_L(M)$ have a limit as $M\to\infty$ means that $\sum_\alpha \beta_\alpha(M)Q_{M-L\alpha}$ must have a limit for $M\to\infty$. For the stable trajectories, $Q_{M-L\alpha}\to 0$ as $M\to\infty$ and since $\beta_\alpha(M)\to\gamma_\alpha$, which is finite, the stable trajectories drop out in this limit. Assume that the unstable trajectories have pure exponential form [Eq. (7.8); the author has not examined alternative forms in detail]. Then the limit is $\sum_{\alpha=1}^{k}\beta_\alpha(M)(r_\alpha)^{M-L}Q_{0\alpha}$. For this to have a limit, it is sufficient to have

$$\beta_\alpha(M)=a_\alpha(r_\alpha)^{-M}\quad(1\leq\alpha\leq k), \tag{7.13}$$

where the constants $a_\alpha$ are arbitrary. Since $|r_\alpha|>1$ for $\alpha\leq k$, the constants $\beta_\alpha(M)$ for $\alpha\leq k$ have the limit 0 as $M\to\infty$, as required by Eq. (7.12). To complete the specification of $\beta_\alpha(M)$, put

$$\beta_\alpha(M)=\gamma_\alpha\quad(\alpha>k). \tag{7.14}$$

With this specification of $\beta_\alpha(M)$, the points $P_N(M)$ satisfy the requirements 1′–3′. The limit $P_{RL}$ has the form

$$P_{RL}=P_c+\sum_{\alpha=1}^{k}a_\alpha Q_{0\alpha}(r_\alpha)^{-L}, \tag{7.15}$$

which has $k$ arbitrary constants, as was stated at the beginning. In fact the renormalized points $P_{RL}$ (for sufficiently large $L$) are just a linear combination of the $k$ unstable trajectories of $U_A$, with the coefficients representing free parameters in the renormalized Hamiltonian.

In fact, the transformation $U_A$ has only one unstable trajectory, the renormalized Hamiltonian has only one free parameter, and Eq. (7.15) reduces to Eq. (6.17), where the free parameter is $a$ (which depends on $\theta_{R0}$). It was also shown in Sec. VI that the eigenvalue of $U_A$ [$r_1$ in Eq. (7.15) or $\beta^{-1}$ in Eq. (6.17)] determines the scaling properties of the renormalized Hamiltonian at small distances.

As a final comment, one notes that the unrenormalized Hamiltonian could be chosen to be any point $P$ with $\theta=\frac{1}{2}\pi$; the renormalized Hamiltonians are independent of the choice of the unrenormalized Hamiltonian since the sequences $P_N(M)$ will in the limit of large $M$ go from the unrenormalized point to $P_c$ and then along the unstable trajectory to a renormalized point $P_{R0}$.

In summary, the renormalized Hamiltonian is determined by properties of the fixed point $P_c$ rather than those of a particular unrenormalized Hamiltonian. The sequence of renormalized Hamiltonians $P_{RN}$ approaches $P_c$ as $N\to\infty$; for large $N$, $P_{RN}-P_c$ must be a linear combination of the unstable trajectories leaving $P_c$, and the different renormalized theories can be labeled by the coefficients $a_\alpha$ relating $P_{RN}-P_c$ to unstable trajectories. I think it is this relation of the renormalized theory to unstable trajectories leaving a fixed point, which is simple, to answer the question raised earlier. The coefficients $a_\alpha$ are, I think, as close as one can get to being fundamental parameters in the theory.

### B. Why Renormalization?

In this part we shall try to understand what features of the model Hamiltonian make renormalization necessary. The first step in the analysis will be to show that the transformation $T$ is divergence free. Then the reason for the appearance of divergences in perturbation theory will be examined.

The statement that the transformation $T$ is divergence free means the following. Let $H$ be a Hamiltonian in $S$. Let $H'$ be $T(H)$. Let $H$ have a decomposition $(J,\mathcal{E},N,\theta,\mathbf{A}_k,C_\lambda)$ and $H'$ have a decomposition $(J',\mathcal{E}',N-1,\theta',\mathbf{A}_k',C_k')$. Then, as discussed in Sec. V, if $J$, $\mathcal{E}$, $\theta$, $\mathbf{A}_k$, and $C_k$ are held fixed while $N$ varies, the quantities $J'$, $\mathcal{E}'$, $\theta'$, $\mathbf{A}_k'$, and $C_k'$ are independent of $N$ and cannot diverge for $N\to\infty$. Furthermore, the transformation is continuous, that is, if $H$ and $H''$ are two Hamiltonians with transforms $H'$ and $H'''$, then $H'\to H'''$ when $H\to H''$. This continuity is uniform in $N$.

To understand the significance of $T$ being divergence free, one can study the divergences that appear in ordinary perturbation theory and see that they arise despite the finiteness of $T$. Consider the unrenormalized cutoff Hamiltonian $H_M$ with a small, bare coupling constant $g_0$ and large cutoff $M$. Consider also the effective Hamiltonian $H_0(M)$ which describes the ground state and first few excited states of $H_M$. That $g_0$ is small means the angle $\theta_M(M)$ [also called $\theta_{0M}$, as in Eq. (5.87)] is small, and an expansion in $g_0$ can easily be converted into an expansion in $\theta_M(M)$. The effective Hamiltonian $H_0(M)$ is known if one knows the three parameters $J_0(M)$, $\mathcal{E}_0(M)$, and $\theta_0(M)$ and the curve

$Q_M(t)$ in $S_A$. The curve $Q_M(t)$ is well behaved for large $M$: As $M\to\infty$ it approaches the limit curve $R(t)$. From Eq. (5.104), $J_0(M)$ is a simple function of $\theta_0(M)$. Thus any divergences in the low-lying energy levels of $H_M$ as $M\to\infty$ must be due to divergences in $\mathcal{E}_0(M)$ or $\theta_0(M)$ as $M\to\infty$. A divergence in $\mathcal{E}_0(M)$ affects only the ground-state energy but not energy differences between the ground state and excited states. A divergence in $\theta_0(M)$ means a divergence in differences of energy levels at least through the scale factor $J_0(M)$. The divergence in $\theta_0(M)$ can be identified as a coupling-constant divergence while a divergence in $\mathcal{E}_0(M)$ is a ground-state energy divergence.

To study the divergences in $\mathcal{E}_0(M)$ and $\theta_0(M)$ one uses Eqs. (5.91) and (5.92) of Sec. V. Let $\theta_M(M)$ be denoted $\theta_M$; $\mathcal{E}_M(M)$ is zero (we do not make an energy subtraction in $H_M$). From the inequality (5.31), one finds that for $\theta$ small,

$$f_L(\theta)=\theta-\eta_L\theta^3, \tag{7.16}$$

with $\eta_L\simeq\frac{1}{2}$. For $L\to\infty$, $\eta_L$ approaches a limit $\eta$ since $f_L(\theta)$ has a limit. To a first approximation one neglects the $\theta^3$ term in Eq. (7.16); then one gets $\theta_0(M)\simeq\theta_M$. To a second approximation one replaces $\theta^3$ by ${\theta_M}^3$; then Eq. (5.92) becomes

$$\theta_N(M)=\theta_{N+1}(M)-\eta_{M-N}{\theta_M}^3, \tag{7.17}$$

which gives

$$\theta_0(M)=\theta_M-(\sum_{n=1}^{M}\eta_n){\theta_M}^3. \tag{7.18}$$

For large $M$ this becomes

$$\theta_0(M)\simeq\theta_M-M\eta{\theta_M}^3, \tag{7.19}$$

and one has a divergence linear in $M$. This corresponds to a logarithmic divergence in the cutoff momentum (since the cutoff momentum is $\Lambda^M$). The energy $\mathcal{E}_0(M)$ is dominated by a contribution from $J_M(M)$:

$$\mathcal{E}_0(M)\simeq\Lambda^M(\cos\theta_M)^{-1}T_c(P_M(M). \tag{7.20}$$

Since $T_c\simeq-1$ for any argument, $\mathcal{E}_0(M)$ is linearly divergent in the cutoff momentum. These are the divergences one expects.

The divergence in $\mathcal{E}_0(M)$ is easy to understand. The ground-state energy of $H_M$ gets contributions from each meson degree of freedom represented in $H_M$. The degree of freedom $m$ contributes an energy of order $\Lambda^m$, for that is the energy scale for mesons in state $\psi_m$. The dominant energy is $\Lambda^M$ associated with mesons having the cutoff momentum. Therefore $\mathcal{E}_0(M)$ is of order $\Lambda^M$. In any case the divergence in $\mathcal{E}_0(M)$ as $M\to\infty$ arises because the scale factor $J_M(M)\to\infty$ as $M\to\infty$. This type of divergence occurs also in relativistic theories as mass renormalization. In some field theories the mass is linearly divergent. The cause of this is that when the cutoff is large the natural energy scale for self-mass effects is the cutoff. Then one must let the bare mass in the Lagrangian be of the order of the cutoff and chosen very carefully so that all cutoff-dependent self-masses cancel and the physical mass is much smaller than the cutoff.

The coupling-constant divergence in $\theta_0(M)$ is more subtle. There is no question of a cutoff-dependent scale here; $\theta$ is a dimensionless variable. The divergence is proportional to the number of degrees of freedom. It arises because the transformation $T$ must be iterated $M$ times to give $H_0(M)$ starting from $H_M$. These iterations define a sequence of constants $\theta_N(M)$. The difference between $\theta_N(M)$ and $\theta_{N+1}(M)$ is finite for all $N$ and small in perturbation theory. However, these differences add in going from $\theta_M$ to $\theta_0(M)$; hence the divergence.

One sees from the above discussion that the divergences of perturbation theory derive from two causes. The linear divergence is due to the energy scale of the cutoff Hamiltonian $H_M$ being $\Lambda^M$ instead of the pion mass. The logarithmic divergence arises because the transformation $T$ is iterated $M$ times in going from $\theta_M$ to $\theta_0(M)$. The cause of the logarithmic divergence must be pursued further. Why was it necessary to compute $\theta_0(M)$ by an iterative process? Will an iterative method in which $\theta_0(M)$ is calculated in $M$ steps always make $\theta_0(M)$ divergent when $M\to\infty$?

To set up the discussion, pretend that the details of the analysis of the model had been different from what was reported in Sec. V. Suppose that the cutoff energy $\Lambda^M$ had not been crucial for the discussion of the model, but that still one defined a sequence of constants $\theta_N(M)$ in going from $\theta_M$ to $\theta_0(M)$. What might one expect in this case? Then, when $M$ and $N$ are large, one would expect that there could be no appreciable difference between $\theta_N(M)$ and $\theta_{N+1}(M)$, for in both cases the effective cutoff ($\Lambda^N$ or $\Lambda^{N+1}$) is large compared to the only important length. Most of the difference between $\theta_0(M)$ and $\theta_M$ would be due to the difference $\theta_0(M)-\theta_1(M)$ or $\theta_1(M)-\theta_2(M)$; the differences $\theta_N(M)-\theta_{N+1}(M)$ for large $N$ would go to zero and could not accumulate to make $\theta_0(M)$ diverge for $M\to\infty$.

Thus the essential question is why the difference $\theta_N(M)-\theta_{N+1}(M)$ does not go to zero for large $N$, at least in perturbation theory. The answer lies in two features of the cutoff Hamiltonian $H_M$ and the effective Hamiltonians $H_N(M)$. The first is that meson degrees of freedom of order $N$ dominate the Hamiltonian $H_N(M)$ rather than meson degrees of freedom of order 1. As a result, the change from $H_N(M)$ to $H_{N-1}(M)$, which means eliminating the $N$th degree of freedom, is a nontrivial change. Thus one can hardly expect $\theta_{N-1}(M)$ to be the same as $\theta_N(M)$ no matter how large $N$ is. If by contrast the meson degrees of freedom of order 1 had been the dominant degrees of freedom in $H_N(M)$ for large $N$, then dropping the $N$th degree of freedom would have been a negligible change and $\theta_{N-1}(M)$ would probably have been equal to $\theta_N(M)$.

The second important feature is scale invariance. Scale invariance means that if the degrees of freedom of order 1 can be neglected (which is true for large $N$), then the process of going from $H_N(M)$ to $H_{N-1}(M)$ is indistinguishable from the process of going from $H_{N-1}(M)$ to $H_{N-2}(M)$. In particular, if $H_{N-1}(M)$ differs from $H_N(M)$ only by a scale factor and an additive constant, then $H_{N-2}(M)$ differs from $H_{N-1}(M)$ only by the same scale factor and another additive constant. Now if $\theta_N(M)$ is small, $N$ is large and $M \gg N$, $H_{N-1}(M)$ does differ from $H_N(M)$ by little more than a scale factor and an additive constant. This is because $H_N(M)$ is defined by the constants $J_N(M)$, $\mathcal{E}_N(M)$, $\theta_N(M)$, and the point $Q_{M-N}(\theta_N(M))$, while $H_{N-1}(M)$ is defined by $J_{N-1}(M)$, $\mathcal{E}_{N-1}(M)$, $\theta_{N-1}(M)$, and $Q_{M-N+1}(\theta_{N-1}(M))$. If $\theta_N(M)$ is small then $\theta_{N-1}(M) \simeq \theta_N(M)$; since $Q_L(t) \approx R(t)$ when $L$ is large,

$$Q_{M-N}(\theta_N(M)) \simeq Q_{M-N+1}(\theta_{N-1}(M)).$$

Thus only the scale factor $J_N(M)$ and constant $\mathcal{E}_N(M)$ can differ appreciably from $J_{N-1}(M)$ and $\mathcal{E}_{N-1}(M)$. But under these circumstances the effect of the transformation $T$ on $H_N(M)$ and $H_{N-1}(M)$ is essentially the same, except for the effect on the scale factors $J$ and the constants $\mathcal{E}$. This is scale invariance, and it means in particular that the difference $\theta_{N-2}(M) - \theta_{N-1}(M)$ is the same as the difference $\theta_{N-1}(M) - \theta_N(M)$ when $\theta_N(M)$ is small; hence the divergence in $\theta_0(M)$ in perturbation theory is proportional to $M$ rather than some other function of $M$.

In conclusion, the fact that meson degrees of freedom of the order of the cutoff dominate the cutoff Hamiltonians makes renormalization inevitable. The divergence problem is not just an artifact of perturbation theory. Since the dominance of the degrees of freedom of order of the cutoff is due to the energy of a meson increasing as its momentum increases, which is also true in relativistic theories, one expects that renormalization will also be inevitable for strongly coupled relativistic theories. We note also that not only does the transformation $T$ determine basic properties of the renormalized theory, as shown in Sec. VII A, it is also divergence free. Clearly one will want to define an analogous transformation for relativistic theories.

### C. Analogy to Renormalization Theory of Gell-Mann and Low

Gell-Mann and Low, in 1954, presented an analysis of the renormalization of quantum electrodynamics, and predicted that there would be an "eigenvalue condition" for the bare coupling constant.[8] That is, the bare coupling constant $e_0$ would have to have a fixed value independent of the value of the renormalized coupling constant. To be precise, they predicted that there would be a function $\psi(x)$ with the property that if $e_0$ is finite, then $e_0$ is a root of the equation $\psi(e_0^2)=0$. To show this, Gell-Mann and Low of necessity had to obtain ideas from perturbation theory and then extrapolate to the region of strong bare coupling constant. This involves several speculations, some of which will be criticized below. Nevertheless, the analysis of Gell-Mann and Low remains after 16 years the most sensible discussion in the literature of nonperturbative renormalization theory for relativistic field theory.

Here is a brief review of the Gell-Mann–Low theory. Let $e$ be the physical (renormalized) electron charge and let $m$ be the physical electron mass. Let $d_c(k^2/m^2,e^2)$ be the renormalized photon propagator apart from a factor $k^{-2}$. The customary normalization requirement for $d_c$ is assumed:

$$d_c(0,e^2)=1. \tag{7.21}$$

Gell-Mann and Low define a generalization of the usual renormalization procedure for electrodynamics, with a different definition of the renormalized charge. In the Gell-Mann–Low program, the renormalized charge is a quantity $e_\lambda$ depending on a subtraction point $\lambda$. The photon propagator is (apart from the factor $k^{-2}$) a function $d(k^2/\lambda^2,m^2/\lambda^2,e_\lambda^2)$ with the normalization condition

$$d(1,m^2/\lambda^2,e_\lambda^2)=1. \tag{7.22}$$

The propagator $d$ is related to the usual propagator $d_c$ through the relation

$$e^2 d_c(k^2/m^2,e^2) = e_\lambda^2 d(k^2/\lambda^2,m^2/\lambda^2,e_\lambda^2). \tag{7.23}$$

In particular, putting $k^2=\lambda^2$ gives

$$e_\lambda^2 = e^2 d_c(\lambda^2/m^2,e^2), \tag{7.24}$$

which gives the definition of $e_\lambda$ in terms of $e$. In the Gell-Mann–Low program, all other amplitudes (electron propagator, vertex function, etc.) are functions of $e_\lambda$, and all depend on the reference momentum $\lambda$ as well as $m$ and various momenta. The subtraction procedure of Gell-Mann and Low is defined so that the bare coupling constant $e_0$ is the limit of $e_\lambda$ as $\lambda \to \infty$.

Gell-Mann and Low then argue that the function $d(k^2/\lambda^2,m^2/\lambda^2,e_\lambda^2)$ has a finite limit when $m \to 0$ holding $k^2$, $\lambda^2$, and $e_\lambda^2$ fixed. This should also be true of other amplitudes. They give an example of this from fourth-order perturbation theory, and then argue that it is true in general because the momentum $k$ provides an infrared cutoff. Whether the finiteness assumption is true is still an open question; the author knows of no reason to doubt it, and it will be assumed to be correct in the following.

If $d_c(k^2/m^2,e^2)$ is expanded in powers of $e^2$ for $k^2$ large, the coefficients involve logarithms of $k^2/m^2$, so that the effective expansion parameter is $e^2\ln(k^2/m^2)$ and not $e^2$; this means that radiative corrections become important when $\ln(k^2/m^2)$ is sufficiently large, no matter how small $e$ is. In contrast, as Gell-Mann and Low make clear, the fact that $d$ is independent of $m^2/\lambda^2$ when $m^2/\lambda^2$ is small means that the expansion

of $d(k^2/\lambda^2,m^2/\lambda^2,e_\lambda^2)$ in powers of $e_\lambda^2$ involves no large logarithms if $k$ and $\lambda$ are simultaneously large so that $k^2/\lambda^2$ is of order 1. In fact, in this case the coefficients of $e_\lambda^2$, $e_\lambda^4$, etc. are of order 1 no matter how large $k$ and $\lambda$ are.

To compute $e_0$ from Eq. (7.24) directly would be difficult since for any $e$ the radiative corrections to $d_e(\lambda^2/m^2,e^2)$ are infinite in the limit $\lambda\to\infty$. Therefore Gell-Mann and Low develop an indirect procedure which requires knowing only $d(k^2/\lambda^2,0,e_\lambda^2)$ for $k^2$ near $\lambda^2$. The radiative corrections to $d$ will be important because, as will be seen, one will have to consider coupling constants $e_\lambda$ of order 1. But unless one must consider the limit $e_\lambda\to\infty$, the radiative corrections will be finite. The trick of Gell-Mann and Low is to observe that one can use the function $d$ to set up an equation for $de_\lambda/d\lambda$. From Eq. (7.23) one finds that, for any $\lambda$ and $\lambda'$,

$$e_\lambda^2 d(k^2/\lambda^2,m^2/\lambda^2,e_\lambda^2)=e_{\lambda'}^2 d(k^2/\lambda'^2,m^2/\lambda'^2,e_{\lambda'}^2). \quad (7.25)$$

Putting $k=\lambda'$ gives

$$e_{\lambda'}^2=e_\lambda^2 d(\lambda'^2/\lambda^2,m^2/\lambda^2,e_\lambda^2). \quad (7.26)$$

If $\lambda$ and $\lambda'$ are both much larger than $m$ one can neglect the $m$ dependence. Differentiating with respect to $\lambda'$ and then putting $\lambda'=\lambda$ and approximating $m/\lambda$ by 0 gives

$$2e_\lambda(de_\lambda/d\lambda)=2\psi(e_\lambda^2)/\lambda, \quad (7.27)$$

where

$$\psi(x)=yx\frac{\partial d(y,0,x)}{\partial y}\bigg|_{y=1}. \quad (7.28)$$

The function $\psi(x)$ has a power-series expansion in $x$ for small $x$ with finite coefficients; Gell-Mann and Low assume it has a well-defined extrapolation to values of $x$ of order 1. To compute the limit of $e_\lambda$ for $\lambda\to\infty$, one must solve the differential equation (7.27). If $de_\lambda/d\lambda$ does not go to zero for finite $e_\lambda$, then necessarily an infinite increase in $\lambda$ will give an infinite increase in $e_\lambda$. Thus the only way $e_\lambda$ can stay finite as $\lambda\to\infty$ is for $\psi(e_\lambda^2)$ to have a zero. If $\psi(x)$ has a zero at $x=x_0$ and is positive for $x<x_0$ ($\psi$ is positive for small $x$ from perturbation theory), then the solution $e_\lambda^2$ of Eq. (7.27) will be an increasing function of $\lambda$ with the limit $x_0$ as $\lambda\to\infty$ (assuming $e_\lambda^2<x_0$ when $\lambda$ is of order $m$, as it will be if $e$ is small).

If $\psi(x)$ has a zero at $x_0$ then the function $e_\lambda^2$ will have the limit $x_0$ as $\lambda\to\infty$ for any value of $e$ sufficiently small. This demonstrates the main result of Gell-Mann and Low: The bare coupling constant $e_0$ is *independent* of the physical coupling constant $e$, at least over some finite range for $e$. Even if $\psi(x)$ does not have a zero, the solution $e_\lambda$ will have the limit $\infty$ for $\lambda\to\infty$ independently of the value of $e$; the bare coupling constant is again independent of the physical coupling constant. [This is true only for certain forms of the function $\psi(x)$. If the integral $\int_1^\infty dx/\psi(x)$ is finite, then $e_\lambda\to\infty$ for some finite value of $\lambda$ and the theory becomes nonsense for larger values of $\lambda$. This leads to contradictions discussed below.]

Thus Gell-Mann and Low predicted for electrodynamics the result that *one* unrenormalized Lagrangian would have an infinite number of solutions. This is exactly the result that was proved for the model in Sec. VII A.

The differential equation (7.27) can be regarded as analogous to the transformation equation

$$P_{RN}=T_A(P_{RN+1}) \quad (7.29)$$

that is involved in the definition of the renormalized Hamiltonian of the model. Equation (7.27) tells how a coupling constant $e_\lambda$ changes as $\lambda$ changes, while Eq. (7.29) tells how an infinite set of coupling constants change as $N$ changes. One can think of the function $\psi$ as defining an infinitesimal transformation on a one-dimensional coupling-constant space. In the limit $\lambda\to\infty$, $e_\lambda$ goes to a fixed point of the transformation defined by $\psi$ [if $\psi(e_0^2)=0$, then for $e_\lambda=e_0$, $de_\lambda/d\lambda=0$: thus $e_0$ is a fixed point]. This is analogous to the result that the limit of $P_{RN}$ as $N\to\infty$ is a fixed point of $T_A$. Thus Gell-Mann and Low discovered the idea that a fixed point of a transformation is important in renormalization. There are differences between Gell-Mann and Low's fixed point $e_0$ and the fixed point $P_e$; these differences will be emphasized below. These differences do not alter the fact that Gell-Mann and Low discovered the essential idea of a fixed point. Since they discovered the idea in the context of relativistic field theory, there is encouragement to believe that the analysis of the fixed point in the model is relevant to relativistic field theory and not just a consequence of the many simplifications which were made in defining the model.

There are two basic differences between the transformation $T_A$ defined for the model and the transformation $\psi$ of electrodynamics. First, the function $\psi$ can only be computed after electrodynamics has been solved, whether by a perturbation expansion or whatever. This is because $\psi$ is defined in terms of the renormalized propagator which is itself part of the solution of electrodynamics. In particular, if electrodynamics does not have a solution except as a perturbation expansion, then the $\psi$ function will not exist for strong coupling. In contrast, the transformation $T_A$ is defined before one knows whether the model has a solution. In the model of this paper the renormalized theory exists, but there are other models for which there is no renormalized theory (except one with no coupling). A particular example is a derivative of the Lee model constructed by analogy with the model of this paper. An earlier version of such a model was described in a previous paper[7] and from the analysis given there it is easy to see what happens in the truncated Lee model. One defines a transformation analo-

gous to $T_A$ and uses it to construct curves analogous to $Q_L(t)$. However these curves do not exist over the full range $0\leq t\leq\frac{1}{2}\pi$, but rather over a range $0\leq t\leq t_L$, where the constants $t_L$ form a decreasing sequence with the limit 0 as $L\to\infty$. The reason for this is that if a Hamiltonian has component $\theta$, the Lee model transformation takes $\theta$ into $\theta'$, where $\theta'<\theta$ for any $\theta>0$ including $\theta=\frac{1}{2}\pi$. This means also that the Lee model $T_A$ has no fixed point analogous to $P_c$. This analysis assumes that one does not permit complex coupling constants, as would be necessary if one wants to obtain nontrivial renormalized solutions. Since every time one considers a new theory the existence of a fixed point of the corresponding transformation $T_A$ is in doubt, and since renormalizability depends on there being such a fixed point (at least for the two examples considered; a general analysis of renormalization theory indicates renormalization could be possible for some types of transformations without fixed points), it is important that $T_A$ be defined without reference to the renormalized theory.

The second difference between $\psi$ and $T_A$ is that $\psi$ acts on a one-dimensional space, while $T_A$ acts on an infinite-dimensional space. In order to formulate the transformation $\psi$ as a transformation on one variable, one has to know that the renormalized theory depends on only one phenomenological parameter. For example, in pseudoscalar-meson theory where there are two phenomenological parameters, one must replace $\psi$ by a transformation on a two-dimensional space. But the lesson of the model of this paper is that the number of phenomenological parameters is not known until one has found the fixed point of $T_A$ and determined the number of unstable solutions of $T_A$ near the fixed point. The fact that $T_A$ is a transformation on an infinite set of coupling constants means one is not committed in advance to a particular number of phenomenological constants. Furthermore, one is not restricted to theories with interactions which are renormalizable. As long as $T_A$ is a transformation on the space of all possible couplings, renormalizable or not, the customary reason for considering only renormalizable interactions becomes irrelevant. The customary reason is that renormalizable interactions require an infinite set of noncounter terms to be renormalized; but now these counter terms are all present in the phenomenological Hamiltonians (or Lagrangians, perhaps). Thus if the renormalization theory of the model can be generalized to relativistic field theory, there is hope that pure quark models or the Fermi interaction can be studied, although there is no guarantee that the corresponding transformations will have fixed points.

## ACKNOWLEDGMENTS

I wish to thank Professor S. Drell and members of the SLAC theory group for their hospitality during the author's stay at SLAC.

## APPENDIX A

It is proven here that an iterative solution to Eq. (4.9) exists and that $1+R^\dagger-R$ has an inverse provided that

$$\|H_I\|<0.2\Delta E, \tag{A1}$$

where $\Delta E$ is the energy difference between the ground states and first excited state of $H_0$. Define a sequence of operators $\{R_n\}$ by

$$R_0=0, \tag{A2}$$

$$R_n=(1-P)(E_0-H_0)^{-1}(1-P-R_{n-1})\times H_I(P+R_{n-1}) \quad (n>0). \tag{A3}$$

Then

$$R=\lim_{n\to\infty}R_n. \tag{A4}$$

To prove the existence of the limit, the following equation is useful:

$$R_{n+1}-R_n=(1-P)(E_0-H_0)^{-1}\times[(1-P-R_n)H_I(R_n-R_{n-1})-(R_n-R_{n-1})H_I(P+R_{n-1})]. \tag{A5}$$

Now it is shown that

$$\|R_n\|<0.4. \tag{A6}$$

*Proof.* This is true for $n=0$. Suppose it is true for $n-1$. Now

$$\|(1-P)(E_0-H_0)^{-1}(1-P)\|=\Delta E^{-1}, \tag{A7}$$

$$\|1-P\|=1, \tag{A8}$$

$$\|P\|=1, \tag{A9}$$

and $R_{n-1}=(1-P)R_{n-1}$ from Eq. (A3). Thus

$$\|R_n\|\leq\Delta E^{-1}(1.4)(0.2\Delta E)(1.4)<0.4. \quad \text{Q.E.D.} \tag{A10}$$

Likewise one can show that

$$\|R_{n+1}-R_n\|\leq 0.4\times(0.56)^n. \tag{A11}$$

Hence from the Cauchy criterion, $R$ exists. It is easily shown that $R$ satisfies Eq. (4.9). The bound (A6) implies that

$$\|R^\dagger-R\|<0.8, \tag{A12}$$

which means that the inverse of $1+R^\dagger-R$ exists as a power series in $R^\dagger-R$.

## APPENDIX B

In this appendix the transformation $T$ will be defined in detail. It will be shown that $T$ has the form of Eqs. (5.16)–(5.19). Then Theorems 1–4 of Sec. V will be proven. The only assumption made in this appendix is that $\Lambda>4\times10^6$.

The first problem is to define $T$. Let $H$ be a Hamiltonian in $S$. Let $H$ have the decomposition $(J,\mathcal{E},N,P_A)$ where $P_A\in S_A$. Let $P_A$ have the decomposition $(\theta,\mathbf{A}_k,C_k)$. Let $H=H_0+H_I$ with $H_0$ given by Eq. (5.14). Define

$H_{\text{eff}}$ using Eq. (4.29). To define $T$, we must specify the decomposition of $H_{\text{eff}}$. The decomposition of $H_{\text{eff}}$ must be defined because it is not unique, as was pointed out in Sec. V. This nonuniqueness means that one must often prove properties for the decomposition of an operator which are obvious or already established for the operator itself. To define this decomposition, we will write out in detail the steps leading to $H_{\text{eff}}$, and specify the decomposition of each of the operators arising in the calculation. The operator $H_I$ has the form $H_I=J\mathcal{K}_I$, with

$$\mathcal{K}_I=\sum_{k=1}^{N}\mathbf{V}_k\cdot\mathbf{B}_{k-1}+\sum_{k=0}^{N}C_k, \tag{B1}$$

where

$$\mathbf{B}_0=(m,\sqrt{2}g\tau^+,\sqrt{2}g\tau^-)+\mathbf{A}_0, \tag{B2}$$

$$\mathbf{B}_k=\mathbf{A}_k \quad (k>0), \tag{B3}$$

and $m=\cos\theta$ and $g=(1/\sqrt{2})\sin\theta$. The equations which define $H_{\text{eff}}$ are as follows [including the iterative definitions of $R$ and $(1+R^\dagger R)^{\pm 1/2}$]:

$$R_0=0, \tag{B4}$$

$$R_n=(E_0-H_0)^{-1}(1-P-R_{n-1})H_I(P+R_{n-1}), \tag{B5}$$

$$R=\lim_{n\to\infty}R_n, \tag{B6}$$

$$Q_0=0, \tag{B7}$$

$$Q_n=\tfrac{1}{2}(R^\dagger R-Q_{n-1}{}^2), \tag{B8}$$

$$Q=\lim_{n\to\infty}Q_n, \tag{B9}$$

$$\bar{Q}_0=0, \tag{B10}$$

$$\bar{Q}_n=-Q-Q\bar{Q}_{n-1}, \tag{B11}$$

$$\bar{Q}=\lim_{n\to\infty}\bar{Q}_n, \tag{B12}$$

$$H_{\text{eff}}=(P+Q)H_I(P+R)(P+\bar{Q})+PE_0. \tag{B13}$$

In these formulas $P$ is the projection operator onto the two ground states of $H_0$, $(P+Q)$ is $(1+R^\dagger R)^{1/2}P$ and[17] $(P+\bar{Q})$ is $P(1+R^\dagger R)^{-1/2}$, and $E_0$ is the ground-state energy of $H_0$.

A particular form for $\mathcal{K}_I$ has been given in Eq. (B1). The operators $\mathbf{B}_k$ and $C_k$ will be called the decomposition of $\mathcal{K}_I$. Analogous decompositions will now be defined for $R_n$, etc. The equations (B4)–(B13) involve three basic operations: multiplication of $H_I$, $R_n$, etc. with themselves, multiplication with $P$, or multiplication with $(E_0-H_0)^{-1}$. Thus it is sufficient to define the decomposition of any of these products. Let $X$ be an operator with decomposition $(\mathbf{D}_k,F_k)$, for example. Then $PX$ has the obvious decomposition $(P\mathbf{D}_k,PF_k)$, and analogously for $(E_0-H_0)^{-1}X$. This is a legitimate decomposition since the only requirement on a decomposition $(\mathbf{D}_k,F_k)$ is that $\mathbf{D}_k$ and $F_k$ do not involve meson operators numbered above $k$ (no upper bounds on $\mathbf{D}_k$ and $F_k$ will be imposed now). Since $P$ and $(E_0-H_0)^{-1}$ act in the space of nucleons and 0-mesons (meson operators numbered 0), this restriction is satisfied by $P\mathbf{D}_k$ and $PF_k$, or $(E_0-H_0)^{-1}\mathbf{D}_k$ and $(E_0-H_0)^{-1}F_k$. Now let $Y$ be another operator with decomposition $(\mathbf{A}_k,C_k)$. One must define a decomposition $(\mathbf{G}_k,L_k)$ for the product $XY$. The decomposition is as follows:

$$\mathbf{G}_0=F_0\mathbf{A}_0+\mathbf{D}_0C_0, \tag{B14}$$

$$L_0=F_0C_0. \tag{B15}$$

For $k\geq 1$,

$$\mathbf{G}_k=\sum_{m=0}^{k}(F_m\mathbf{A}_k+\mathbf{D}_kC_m)+\sum_{m=0}^{k-1}(F_k\mathbf{A}_m+\mathbf{D}_mC_k)$$
$$+\sum_{n=1}^{k}\sum_{m=0}^{n-1}[\mathbf{D}_k(\mathbf{T}_n\cdot\mathbf{A}_m)+(\mathbf{T}_n\cdot\mathbf{D}_m)\mathbf{A}_k]$$
$$+\sum_{n=0}^{k-1}\sum_{m=0}^{k-1}[\mathbf{D}_n(\mathbf{T}_k\cdot\mathbf{A}_m)+(\mathbf{T}_k\cdot\mathbf{D}_n)\mathbf{A}_m], \tag{B16}$$

$$L_k=\sum_{n=0}^{k-1}\sum_{m=0}^{k-1}(\mathbf{T}_k\cdot\mathbf{D}_n)(\mathbf{T}_k\cdot\mathbf{A}_m)$$
$$+\sum_{n=1}^{k}\sum_{m=0}^{n-1}[(\mathbf{T}_n\cdot\mathbf{D}_m)C_k+F_k(\mathbf{T}_n\cdot\mathbf{A}_m)]$$
$$+\sum_{m=0}^{k}F_kC_m+\sum_{m=0}^{k-1}F_mC_k. \tag{B17}$$

$\mathbf{T}_n$ is defined by Eqs. (5.4)–(5.6). It is clear from these formulas that $\mathbf{G}_k$ and $L_k$ do not involve meson operators numbered above $k$. With some straightforward algebra, one can verify that the operator product $XY$ is given by

$$XY=\sum_{k=1}^{N}\mathbf{V}_k\cdot\mathbf{G}_{k-1}+\sum_{k=0}^{N}L_k. \tag{B18}$$

It can be shown that the decomposition is associative, i.e., a triple product $(XY)Z$ has the same decomposition as $X(YZ)$.

With the rules specified above and Eqs. (B1)–(B13), the decomposition of $H_{\text{eff}}-E_0P$ is uniquely defined. Note that the number of degrees of freedom $N$ nowhere enters into the calculation of $\mathbf{G}_k$ and $L_k$. Therefore if the operators $\mathbf{B}_k$ and $C_k$ in the decomposition of $\mathcal{K}_I$ are defined for *all* $k$ and are independent of $N$, then the decompositions of $R_n$, etc. (including $H_{\text{eff}}$) will also be defined for all $k$ and independent of $N$. It will be presumed from now on that decompositions are defined and computed for all $k$. Note also that $H_I$, $H_0$, and $E_0$ are all proportional to $J$. This makes $R_n$, $Q_n$, etc. independent of $J$ and $H_{\text{eff}}$ proportional to $J$. To be

[17] $\bar{Q}$ is a symbol completely independent of $Q$.

specific, $H_{\text{eff}}$ has the form

$$H_{\text{eff}}=PE_0+J(\sum_{k=1}^{N}\mathbf{V}_k\cdot\mathbf{G}_{k-1}+\sum_{k=0}^{N}L_k), \quad \text{(B19)}$$

where $\mathbf{G}_k$ and $L_k$ depend on $\tau^{\pm}$ and meson operators numbered 0 to $k$.

Since $H_{\text{eff}}$ acts within the subspace projected by $P$, the dependence of $H_{\text{eff}}$ on the 0-meson operators ($a_0$, etc.) and $\tau^{\pm}$ can be reduced to a dependence on $\tau_R{}^{\pm}$, the raising and lowering operators for the ground states of $H_0$. When this is done, $\mathbf{G}_k$ and $L_k$ depend only on $\tau_R{}^{\pm}$ and meson operators numbered 1 to $k$. To put $H_{\text{eff}}$ in a form in which it can be contained in the space $S$, one must renumber the meson operators 1–$N$ to run from 0 to $N-1$, e.g., $a_1\rightarrow a_0$, $a_2\rightarrow a_1$, etc. Also one replaces $\tau_R{}^{\pm}$ by $\tau^{\pm}$. Under this renumbering, $\mathbf{V}_k$ becomes $\Lambda^{-1}\mathbf{V}_{k-1}$; $H_{\text{eff}}$ is

$$H_{\text{eff}}=E_0+J\Lambda^{-1}(\sum_{k=0}^{N-1}\mathbf{V}_k\cdot\mathbf{G}_k+\Lambda\sum_{k=0}^{N}L_k), \quad \text{(B20)}$$

where $\mathbf{G}_k$ and $L_k$ depend on $\tau^{\pm}$ and meson operators numbered 0 to $k-1$. $PE_0$ is replced by $E_0$ because there is no longer any possible reference to states outside the subspace projected by $P$.

Now consider $\mathbf{G}_0$ and $L_0$. They involve no meson operators; they can be expressed purely in terms of $\tau^{\pm}$. Furthermore $\mathbf{G}_0$ and $L_0$ satisfy the appropriate Hermiticity, charge conservation, charge conjugation, and time-reversal requirements, because these requirements are preserved by the equations defining the decomposition of $H_{\text{eff}}$. These requirements force $L_0$ to be a real constant and $\mathbf{G}_0$ to have the form

$$\mathbf{G}_0=(m'',\sqrt{2}g''\tau^{+},\sqrt{2}g''\tau^{-}), \quad \text{(B21)}$$

where $m''$ and $g''$ are real constants.

It is now easy to define a decomposition of $H_{\text{eff}}$ in the space $S$. Denote the decomposition $(J',\mathcal{E}',N',P_A')$ with $P_A'$ having the decomposition $(\theta',\mathbf{A}_k',C_k')$. Comparison of Eq. (B20) with Eqs. (5.1), (5.2), (5.8), and (5.9) leads to the following formulas:

$$J'=\Lambda^{-1}J(m''^2+2g''^2)^{1/2}, \quad \text{(B22)}$$

$$\mathcal{E}'=E_0+JL_0, \quad \text{(B23)}$$

$$N'=N-1, \quad \text{(B24)}$$

$$\theta'=\tan^{-1}(\sqrt{2}g''/m''), \quad \text{(B25)}$$

$$\mathbf{A}_k'=(m''^2+2g''^2)^{-1/2}\mathbf{G}_{k+1}, \quad \text{(B26)}$$

$$C_k'=\Lambda(m''^2+2g''^2)^{-1/2}L_{k+1}. \quad \text{(B27)}$$

The quantities $m''$, $g''$, $L_0$, $\mathbf{G}_k$, and $L_k$ depend only on $P_A$, not on $J$, $\mathcal{E}$, or $N$. Hence, Eq. (B22) has the form of Eq. (5.17) with $T_B(P_A)=(m''^2+2g''^2)^{1/2}$. Also, Eqs. (B25)–(B27) define the transformation $T_A(P_A)$ of Eq. (5.19). Finally, the ground-state energy of $H_0$ [defined by Eq. (5.14)] is

$$E_0=\mathcal{E}-J \quad \text{(B28)}$$

(of Table I). Hence $\mathcal{E}'$ has the form of Eq. (5.18) with

$$T_c(P_A)=-1+L_0. \quad \text{(B29)}$$

The next problem is to prove Theorems 1–4 of Sec. V. The proofs involve a very large number of upper bounds and are quite complex. To guard against subtle errors, all bounds have been obtained as explicit numbers multiplying powers of $\Lambda$. In principle, it would have been sufficient to know that bounds existed in the form of unknown sufficiently large numbers multiplying known powers of $\Lambda$. In addition, the use of numbers saves symbols. In the following, $\leq$ means only $\not>$ (the equality need not be realized). These proofs are crucial to the renormalization of the model of this paper; they are condemned to an appendix because they are special to the model, whereas the analysis of Sec. V is of more general interest.

To start with, one needs an upper bound for the decomposition of the product $XY$ given bounds on $X$ and $Y$. Let $X$ and $Y$ have decompositions $(\mathbf{D}_k,F_k)$ and $(\mathbf{A}_k,C_k)$ as before. It is convenient to define an abstract bound for $X$. This bound will consist of three numbers $(d,e,f)$. By definition, $X$ has a bound $(d,e,f)$ if

$$\begin{aligned}&\|\mathbf{D}_0\|\leq\mathbf{m}d, \quad \|\mathbf{D}_k\|\leq\mathbf{m}e\Lambda^{-k} \quad (k\geq 1),\\ &\|F_0\|\leq f\Lambda^{-1}, \quad \|F_k\|\leq e\Lambda^{-2k} \quad (k\geq 1),\end{aligned} \quad \text{(B30)}$$

where $\|\mathbf{D}_0\|$ is a vector with components $\|D_{01}\|$, $\|D_{02}\|$, $\|D_{03}\|$, $\|D_{01}\|$ being the ordinary operator bound. Also, $\mathbf{m}$ is the vector $(m,\sqrt{2}g,\sqrt{2}g)$, and $m$ and $g$ are as defined before.

Suppose $X$ has a bound $(d,e,f)$ and $Y$ has a bound $(a,b,c)$. Then it can be shown that $XY$ has a bound $(g,h,l)$,[18] with

$$g=\Lambda^{-1}(af+dc), \quad \text{(B31)}$$

$$h=5ad+\Lambda^{-1}[bf+ec+(\sqrt{70})(ae+bd)]+\Lambda^{-2}(14be), \quad \text{(B32)}$$

$$l=cf\Lambda^{-1}. \quad \text{(B33)}$$

(These bounds were computed assuming only that $\Lambda>21$.) A brief summary of the proof of these bounds is as follows. The operators $T_{ki}$ have bounds

$$\|T_{ki}\|=\Lambda^{-k}. \quad \text{(B34)}$$

(This is proved by a straightforward calculation.) Next one puts bounds on the sums $\sum_{n=1}^{\infty}\|\mathbf{T}_n\|$, $\sum_{k=0}^{\infty}\|\mathbf{A}_n\|$, etc. (which are also bounds for finite sums such as $\sum_{n=0}^{k}\|\mathbf{A}_n\|$). One gets

$$\sum_{n=1}^{\infty}\|\mathbf{T}_n\|=\mathbf{1}\Lambda^{-1}(1-\Lambda^{-1})^{-1}\leq 1.05\Lambda^{-1}\mathbf{1}, \quad \text{(B35)}$$

[18] The symbol $g$ here has a different meaning from elsewhere in the paper.

where **1** is the vector (1,1,1). Also,

$$\sum_{n=0}^{\infty}\|\mathbf{A}_n\|\leq \mathbf{m}a+\mathbf{m}b\Lambda^{-1}(1-\Lambda^{-1})^{-1}$$
$$\leq \mathbf{m}(a+1.05\Lambda^{-1}b), \quad \text{(B36)}$$

using the definition of the bound $(a,b,c)$. Similar formulas can be obtained for sums of $\|C_n\|$, $\|\mathbf{D}_n\|$, and $\|F_n\|$. Now one constructs upper bounds for all the terms in Eqs. (B14)–(B17) for $\mathbf{G}_0$, $L_0$, $\mathbf{G}_k$ and $L_k$. For example, one term in $\mathbf{G}_k$ is

$$\|\sum_{n=1}^{k}\sum_{m=0}^{n-1}\mathbf{D}_k\mathbf{T}_n\cdot\mathbf{A}_m\|\leq\|\mathbf{D}_k\|\sum_{n=1}^{\infty}\|\mathbf{T}_n\|\cdot\sum_{m=0}^{\infty}\|\mathbf{A}_m\|$$
$$\leq \mathbf{m}e\Lambda^{-k}(1.05\Lambda^{-1})\mathbf{1}\cdot\mathbf{m}(a+1.05\Lambda^{-1}b). \quad \text{(B37)}$$

Now $\mathbf{1}\cdot\mathbf{m}=m+2\sqrt{2}g$. An upper bound on $m+2\sqrt{2}g$ results from

$$(m+2\sqrt{2}g)^2-5(m^2+2g^2)=-(2m-\sqrt{2}g)^2<0. \quad \text{(B38)}$$

Because $m^2+2g^2=1$ (normalization condition), one gets the bound

$$\mathbf{1}\cdot\mathbf{m}\leq\sqrt{5}. \quad \text{(B39)}$$

Hence

$$\|\sum_{n=1}^{k}\sum_{m=0}^{n-1}\mathbf{D}_k\mathbf{T}_n\cdot\mathbf{A}_m\|$$
$$\leq \mathbf{m}\Lambda^{-k-1}(1.05\sqrt{5})(ae+1.05\Lambda^{-1}be). \quad \text{(B40)}$$

Similarly one finds bounds for all terms in Eqs. (B14)–(B17); the result is that $(g,h,l)$ given by Eqs. (B31)–(B33) is an upper bound for the product $XY$. It is convenient to introduce a shorthand for Eqs. (B31)–(B33): We define the "product" $(a,b,c)(d,e,f)$ to be the quantities $(g,h,l)$ given by Eqs. (B31)–(B33). This product can be shown to be associative and commutative and thus algebraic expressions involving these products can be manipulated using ordinary algebra. This simplifies the calculations.

Using the bound quoted above for products, one can construct a set of upper bounds for the operators $R_n$, $Q_n$, etc. These bounds are listed in Table IV. They are not least upper bounds. The operator $\bar{H}$ in Table IV is defined in terms of $H_{\text{eff}}$ by

$$H_{\text{eff}}=PE_0+JP\mathbf{M}\cdot\mathbf{V}_1P+J\bar{H}, \quad \text{(B41)}$$

where $H_{\text{eff}}$ is the effective Hamiltonian, as of Eqs. (B13) and (B19), before renumbering the meson operators, and $\mathbf{M}$ is

$$\mathbf{M}=(m,\sqrt{2}g\tau^{+},\sqrt{2}g\tau^{-}). \quad \text{(B42)}$$

The operator $\mathfrak{IC}_J$ in Table IV is defined by

$$\mathfrak{IC}_J=\sum_{k=1}^{N}\mathbf{V}_k\cdot\mathbf{A}_{k-1}+\sum_{k=0}^{N}C_k, \quad \text{(B43)}$$

so that

$$\mathfrak{IC}_I=\mathbf{M}\cdot\mathbf{V}_1+\mathfrak{IC}_J. \quad \text{(B44)}$$

TABLE IV. Operator bounds obtained in the proofs of Theorems 1 and 2, assuming $\Lambda>4\times10^6$.

| Operator | Bound | Operator | Bound |
|---|---|---|---|
| $\mathfrak{IC}_I$ | (1.5,0.5,100) | $Q_n-Q_{n-1}$ | $13g^2\times10^{4-n}(\Lambda^{-1},1,\Lambda^{-1})$ |
| $\mathfrak{IC}_J$ | $200g^2(\Lambda^{-1},\Lambda^{-1},1)$ | $\bar{Q}_n$, $\bar{Q}$ | $14g^2\times10^3(\Lambda^{-1},1,\Lambda^{-1})$ |
| $R_n$, $R$ | $g(2,65,160)$ | $\bar{Q}_n-\bar{Q}_{n-1}$ | $14g^2\times10^{4-n}(\Lambda^{-1},1,\Lambda^{-1})$ |
| $R_n-R_{n-1}$ | $16g\times10^{2-n}(1,\Lambda,1)$ | $\bar{H}$ | $g^2(25\times10^3\Lambda^{-1},40,210)$ |
| $Q_n$, $Q$ | $13g^2\times10^3(\Lambda^{-1},1,\Lambda^{-1})$ | | |

The proofs of the bounds of Table IV are mostly straightforward and only examples of the proofs will be given here. In some cases the bounds of Table IV are gross overestimates of the bounds one calculates in the proofs quoted below.

The bounds on $\mathfrak{IC}_J$ and $\mathfrak{IC}_I$ are simple consequences of the definition of the space $S$, in particular, the bounds (5.10)–(5.13). In computing the bound on $\mathfrak{IC}_I$, one also uses the inequality $g^2<\frac{1}{2}$ which follows from the definition (5.9) of $g$ (also one uses $\Lambda>200$).

In proving the bound on $R_n$, it is convenient to eliminate the factor $J$ by defining

$$H_0=E_0+J\mathfrak{IC}_0, \quad \text{(B45)}$$

Now write the equation for $R_n$ as

$$R_n=(-\mathfrak{IC}_0)^{-1}(1-P)\mathbf{M}\cdot\mathbf{V}_1P$$
$$+(-\mathfrak{IC}_0)^{-1}(1-P)(\mathfrak{IC}_JP-R_{n-1}\mathfrak{IC}_IP+\mathfrak{IC}_IR_{n-1}$$
$$-R_{n-1}\mathfrak{IC}_IR_{n-1}). \quad \text{(B46)}$$

The proof of the bound on $R_n$ is by induction. It is true of $R_0$. Assume that it is true of $R_{n-1}$. To bound the first term of Eq. (B46), one needs the following bounds:

$$\|(-\mathfrak{IC}_0)^{-1}(1-P)\|=1, \quad \text{(B47)}$$
$$\|(1-P)\tau^{+}P\|=\|(1-P)\tau^{-}P\|=g. \quad \text{(B48)}$$

These bounds can be obtained by explicit calculation using Table I. With this information, one finds that $(-\mathfrak{IC}_0)^{-1}(1-P)\mathbf{M}\cdot\mathbf{V}_1P$ has a bound $(g,0,0)$. The second term in Eq. (B46) can be bounded using the bound $\|P\|=1$, Eq. (B47), and the bounds of Table IV for $\mathfrak{IC}_J$, $R_{n-1}$, and $\mathfrak{IC}_I$. Schematically one has

$$|R_n|\leq(g,0,0)+|\mathfrak{IC}_J|+2|R_{n-1}|\mathfrak{IC}_I|$$
$$+|\mathfrak{IC}_I|\,|R_{n-1}|^2, \quad \text{(B49)}$$

where $|\mathfrak{IC}_J|$ means $200g^2(\Lambda^{-1},\Lambda^{-1},1)$, etc. After calculating the products explicitly using Eqs. (B31)–(B33), one finds that this expression is less than the bound of Table IV for $|R_n|$. Hence the bound of Table IV for $R_n$ holds for all $n$, by induction. The same bound holds for $R$ because $R$ is the limit of $R_n$ for $n\to\infty$.

The bound on $R_n-R_{n-1}$ is also proven by induction. The bound on $R_1-R_0$ is true because it is larger than the bound of Table IV for $R_1$. Then one computes a bound on $R_{n+1}-R_n$, given the bound for $R_n-R_{n-1}$ and using Eq. (A5) of Appendix A. Since the bound for $|R_n-R_{n-1}|$ goes to zero as $n\to\infty$, the decomposi-

tion of $R_n$ approaches a limit for $n\to\infty$; the limit defines a decomposition for $R$.

To get a bound for $\bar{H}$, one writes

$$\bar{H}=P\mathbf{M}\cdot\mathbf{V}_1(1-P)R+P\mathfrak{IC}_J(P+R)+Q\mathfrak{IC}_I(P+R)(P+\bar{Q})+P\mathfrak{IC}_I(P+R)\bar{Q} \quad \text{(B50)}$$

and sums the bounds of each term.

Now one can get bounds on $m''$ and $g''$. Let $\bar{H}$ have a decomposition $(\mathbf{D}_k,F_k)$. Comparing Eq. (B41) with Eq. (B19), one gets

$$\mathbf{G}_0=P\mathbf{M}P+\mathbf{D}_0, \quad \text{(B51)}$$

$$\mathbf{G}_k=\mathbf{D}_k \quad (k>0), \quad \text{(B52)}$$

$$L_k=F_k \quad (\text{all } k). \quad \text{(B53)}$$

Explicit calculation using Table I gives

$$P\mathbf{M}P=(m,\sqrt{2}g(1-g^2)\tau_R^+,\sqrt{2}g(1-g^2)\tau_R^-). \quad \text{(B54)}$$

From Table IV, $\mathbf{D}_0$ has the bound (remember $\Lambda>4\times10^6$)

$$\|\mathbf{D}_0\|\leq\mathbf{m}g^2\times25\,000\Lambda^{-1}\leq0.01\mathbf{m}g^2. \quad \text{(B55)}$$

The bound on $\mathbf{D}_0$ is a bound on the difference $\mathbf{G}_0-P\mathbf{M}P$. $\mathbf{G}_0$ can be expressed in terms of $m''$ and $g''$ [Eq. (B21); for $\tau^\pm$ in Eq. (B21), read $\tau_R^\pm$, since in the present analysis we have not yet substituted $\tau^\pm$ for $\tau_R^\pm$]. Using Eqs. (B21), (B54), (B51), and (B55), one gets the bounds

$$|m''-m|\leq0.01mg^2, \quad \text{(B56)}$$

$$|g''-g(1-g^2)|\leq0.01g^3. \quad \text{(B57)}$$

From these bounds, one gets bounds on $\tan\theta'$ [Eq. (B25)]:

$$(\sqrt{2}g/m)(1-1.01g^2)(1+0.01g^2)^{-1}\leq\tan\theta'\leq(\sqrt{2}g/m)(1-0.99g^2)(1-0.01g^2)^{-1}. \quad \text{(B58)}$$

Using the bound $g^2\leq\frac{1}{2}$, one can simplify these bounds; inserting $\sqrt{2}g=\sin\theta$ and $m=\cos\theta$, one gets

$$\tan\theta(1-0.51\sin^2\theta)\leq\tan\theta'\leq\tan\theta(1-0.48\sin^2\theta). \quad \text{(B59)}$$

To complete the proof of Theorem 2, one notes that [see Eq. (B29)]

$$T_c(P_A)=-1+L_0=-1+F_0. \quad \text{(B60)}$$

From the bound on $\bar{H}$, $F_0$ is less than $210g^2\Lambda^{-1}$ which is less than 0.01. Hence one obtains Eq. (5.30).

To prove Theorem 1, one must have bounds for $\|\mathbf{A}_k'\|$ and $C_k'$ in terms of $m'$ and $g'$. One has bounds for $\|\mathbf{G}_k\|$ and $\|L_k\|$ in terms of $m$ and $g$ [from Table IV and Eqs. (B52) and (B53)]:

$$\|\mathbf{G}_k\|\leq\mathbf{m}g^2\times40\Lambda^{-k} \quad (k\geq1), \quad \text{(B61)}$$

$$\|L_k\|\leq g^2\times40\Lambda^{-2k} \quad (k\geq1). \quad \text{(B62)}$$

From Eq. (B59) one has $\theta'<\theta$, and therefore $m<m'$. To get a bound on $g$ in terms of $g'$, one uses Eq. (B59). Let

$$(1-0.51\sin^2\theta)^2=1-\beta. \quad \text{(B63)}$$

Then

$$g^2/g'^2=\sin^2\theta/\sin^2\theta'=\sin^2\theta[1+(\tan^2\theta')^{-1}]\leq\sin^2\theta[1+\cos^2\theta(1-\beta)^{-1}(\sin^2\theta)^{-1}]=1+(1-\sin^2\theta)\beta(1-\beta)^{-1}. \quad \text{(B64)}$$

The maximum value of $\beta$ occurs for $\sin\theta=1$ and is less than 0.8. Except for very small $\theta$, $1-\beta$ is larger than $1-\sin^2\theta$, making $g^2/g'^2$ less than $1+\beta$. Hence

$$g^2\leq1.8g'^2. \quad \text{(B65)}$$

Hence

$$\mathbf{m}\leq(1.8)^{1/2}\mathbf{m}' \quad \text{(B66)}$$

(the inequality is true for each component of the two vectors). Also, from Eqs. (B56) and (B57) and $m^2+2g^2=1$, one gets a minimum value for $m'^2+2g'^2$, which in turn gives a bound

$$(m'^2+2g'^2)^{-1/2}<2.03. \quad \text{(B67)}$$

The bounds (B61), (B62), (B66), and (B67), substituted in Eqs. (B26) and (B27), lead to the bounds

$$|\mathbf{A}_k'|\leq200\mathbf{m}'g'^2\Lambda^{-k-1}, \quad \text{(B68)}$$

$$|C_k'|\leq200g'^2\Lambda^{-2k-1}. \quad \text{(B69)}$$

To complete the proof of Theorem 1, one must show that $\mathbf{A}_k'$ and $C_k'$ satisfy Hermiticity requirements and symmetry requirements with respect to charge conservation, charge conjugation, and time reversal. The symmetry requirements are easily established since all the intermediate operators $R_n$ etc., have the same symmetries as $H$. One easily verifies that if $X$ and $Y$ are operators whose decompositions obey the symmetry requirements, then the product $XY$ has a decomposition obeying the symmetry requirements. The rest of the proof of symmetry is omitted. Hermiticity is more complicated because $R_n$ and $R$ are not Hermitian, and one must use Eq. (4.3) instead of Eq. (4.19) to show that $H_{\text{eff}}$ is Hermitian. However, a proof showing that $A_k'$ and $C_k'$ satisfy the Hermiticity requirements of $S_A$ can still be constructed. The basic result needed for the proof is that if $XY$ has a decomposition $(\mathbf{G}_k,L_k)$, then $Y^\dagger X^\dagger$ has the Hermitian conjugate decomposition $(G_{k1}\to G_{k1}^\dagger,\ G_{k2}\to G_{k3}^\dagger,\ G_{k3}\to G_{k2}^\dagger,\ L_k\to L_k^\dagger)$. The proof is omitted.

Now Theorem 3 will be proven. If an operator $X$ has a decomposition $(\mathbf{A}_k,C_k)$, we will call the $A_{k1}$ the "1-components" of $X$.

Note the following. Let operators $X$ and $Y$ have the decompositions $(\mathbf{D}_k,F_k)$ and $(\mathbf{A}_k,C_k)$, respectively. Let $A_{k1}$ vanish for all $k$ and $D_{k1}$ vanish for $k>0$, and let $D_{01}$ be a $c$ number. Let the product $XY$ have decomposition $(\mathbf{G}_k,L_k)$. Then from Eqs. (B14) and (B16),

$$G_{01}=D_{01}C_0, \quad \text{(B70)}$$

$$G_{k1}=D_{01}(C_k+\sum_{m=0}^{k-1}\mathbf{T}_k\cdot\mathbf{A}_m) \quad (k>0). \quad \text{(B71)}$$

Exactly the same formulas for $G_{01}$ and $G_{k1}$ result from decomposing the commuted product $YX$. This means that the commutator $[X,Y]$ has no 1-components in its decomposition.

If $D_{01}$ is zero also then $XY$ has no 1-components. Now consider Theorem 3. Let the 1-components of $H_J$ vanish. We prove by induction that $R_n$ has no 1-components. This is obviously true of $R_0$. Assume it is true of $R_{n-1}$. Consider Eq. (B5). The operator $(1-P)H_IP$ has no 1-components because

$$(1-P)H_IP=(1-P)\mathbf{M}\cdot\mathbf{V}_1P+(1-P)H_JP; \quad \text{(B72)}$$

$H_J$ has no 1-components by assumption and

$$(1-P)M_1P=(1-P)mP$$

vanishes. The product $R_{n-1}H_IR_{n-1}$ can be written $R_{n-1}PH_I(1-P)R_{n-1}$ and so is a product of terms none of which contain a 1-component. So this product has no 1-components. The remaining term in $R_n$ can be written $(E_0-H_0)^{-1}(1-P)[H_I,R_{n-1}]P$. The operators $X=H_I$ and $Y=R_{n-1}$ satisfy the conditions noted above, so the commutator $[H_I,R_{n-1}]$ has no 1-components. Hence $R_n$ has no 1-components. It follows that $R$ has no 1-components, nor do $Q$ and $\bar{Q}$. Now consider $H_{\text{eff}}$ [Eq. (B13)]. Using the fact that $Q=QP$, $\bar{Q}=P\bar{Q}$, $R=(1-P)R$, and that $PH_I(1-P)$ has no 1-components, one sees that the 1-components of $H_{\text{eff}}$ are contained in $(P+Q)PH_I(P+\bar{Q})$. This can be written

$$PH_I(P+\bar{Q})+PH_IQ(P+\bar{Q})+[Q,PH_I](P+\bar{Q}).$$

The commutator has no 1-components by the argument noted above. The other terms can be written

$$PH_I(P+Q)(P+\bar{Q})=PH_IP, \quad \text{(B73)}$$

since $(P+Q)(P+\bar{Q})$ is $P$. This means that the only 1-component in $H_{\text{eff}}$ comes from $P\mathbf{M}\cdot V_1P$. This means that in Eq. (B20), $G_{01}$ is $m$ and $G_{k1}$ vanishes for $k\geq 1$. This means that $m''=m$ and $A_{k1}'=0$, which is Theorem 3.

Finally Theorem 4 will be proven. This requires that two Hamiltonians, say, $H$ and $H_A$, be compared. Let $H$ and $H_A$ both be elements of $S$. Then for each of the operators $\mathfrak{C}_I$, $\mathfrak{C}_J$, $R_n$, etc., associated with $H$, there is a corresponding operator $\mathfrak{C}_{IA}$, $\mathfrak{C}_{JA}$, $R_{An}$, etc., associated with $H_A$. The decomposition of $\mathfrak{C}_J$ is $(\mathbf{A}_k,C_k)$; the decomposition of $\mathfrak{C}_{JA}$ is $(\mathbf{A}_{kA},C_{kA})$. One has

$$\mathfrak{C}_I=\mathbf{M}\cdot\mathbf{V}_1+\mathfrak{C}_J, \quad \text{(B74)}$$

$$\mathfrak{C}_{IA}=\mathbf{M}_A\cdot\mathbf{V}_1+\mathfrak{C}_{JA}, \quad \text{(B75)}$$

where

$$\mathbf{M}=(m,\sqrt{2}g\tau^+,\sqrt{2}g\tau^-), \quad \text{(B76)}$$

$$\mathbf{M}_A=(m_A,\sqrt{2}g_A\tau^+,\sqrt{2}g_A\tau^-), \quad \text{(B77)}$$

and

$$m^2+2g^2=m_A{}^2+2g_A{}^2=1. \quad \text{(B78)}$$

The assumptions of Theorem 4 are that

$$(m-m_A)^2+2(g-g_A)^2\leq d_1{}^2, \quad \text{(B79)}$$

$$\|\mathbf{A}_{kA}-\mathbf{A}_k\|<d_2\mathbf{u}\Lambda^{-k-1} \quad (\text{all } k), \quad \text{(B80)}$$

$$\|C_{kA}-C_k\|\leq d_2\Lambda^{-2k-1} \quad (\text{all } k), \quad \text{(B81)}$$

where $\mathbf{u}$ is the vector $(1/\sqrt{2})(1,1,1)$. The objective is to obtain bounds on $m_A'-m'$, $g_A'-g'$, $\mathbf{A}_{kA}'-\mathbf{A}_k'$, and $C_{kA}'-C_k'$, all in terms of $d_1$ and $d_2$. The bounds will be computed by the same techniques as in the proofs of Theorems 1 and 2. One change is that in defining the bound $(a,b,c)$ of an operator $X$, the vector $\mathbf{u}$ is substituted in Eq. (B30) for $\mathbf{m}$. From Eq. (B78) it follows that $\mathbf{m}\leq\sqrt{2}\mathbf{u}$, and $\mathbf{m}_A\leq\sqrt{2}\mathbf{u}$, which means that a bound $(a,b,c)$ from Table IV (which implies the use of the vector $\mathbf{m}$) can be changed into a bound using the vector $\mathbf{u}$ simply by the replacement $a\rightarrow\sqrt{2}a$, $b\rightarrow\sqrt{2}b$, $c\rightarrow c$. The bounds for $\mathfrak{C}_I$, etc., expressed in terms of $\mathbf{u}$, also bound $\mathfrak{C}_{IA}$, etc.

A problem arises in comparing $A_{kA}'$ with $A_k'$. As part of the definition of $\mathbf{A}_k'$, one took $\mathbf{G}_{k+1}$ and replaced $\tau_R{}^+$ by $\tau^+$. The operator $\tau_R{}^+$ is an operator in the full Hilbert space of mesons labeled 0–$N$ and the nucleons, although it is nonzero only in the subspace of the ground states $|P\rangle$ and $|N\rangle$ of $H_0$ plus mesons labeled 1–$N$. The operator $\tau^+$ acts in a separate space isomorphic to this subspace. Now when the operators $\mathbf{A}_{kA}$ are calculated, one starts from $\mathbf{G}_{k+1A}$ expressed in terms of operators $\tau_{RA}{}^\pm$ which are *different* from $\tau_R{}^\pm$. This is because $\tau_{RA}{}^\pm$ are raising and lowering operators for a different pair of states $|PA\rangle$ and $|NA\rangle$, namely, the ground states of $H_{0A}$. However, in $A_k'$ and $A_{kA}'$ the same operators $\tau^\pm$ appear. Thus it will simplify matters to make a unitary transformation on $\mathbf{G}_{k+1A}$ which takes $\tau_{RA}{}^\pm$ into $\tau_R{}^\pm$; after this has been done, one can make comparisons in the full space of 0–$N$ mesons plus nucleons instead of the separate space involving $\tau^\pm$ plus 1–$N$ mesons. Let the unitary transformation be $U_A$. One wants $U_A{}^\dagger$ to take eigenstates of $H_{0A}$ into eigenstates of $H_A$. In particular, if $P_A$ projects the ground states of $H_{0A}$, one wants

$$U_A{}^\dagger P_AU_A=P. \quad \text{(B82)}$$

Then one replaces $\mathbf{G}_{kA}$ by $U_A{}^\dagger\mathbf{G}_{kA}U_A$ before comparing with $\mathbf{G}_k$, and likewise for $L_{kA}$.

One can take Eqs. (B4)–(B13), replace $R_n$ by $R_{An}$, etc., and then transform them all by $U_A{}^\dagger\cdots U_A$. Note that $\mathfrak{C}_{0A}$ and $\mathfrak{C}_0$ [cf. Eq. (B45)] have the same eigenvalues (0, 1, and 2) (cf. Table I), so $U_A{}^\dagger\mathfrak{C}_{0A}U_A=\mathfrak{C}_0$. From now on, let $R_{An}$ stand for what was $U_A{}^\dagger R_{An}U_A$, and likewise for $R_A$, $Q_{An}$, $Q_A$, $\bar{Q}_{An}$, $\bar{Q}_A$, $\bar{H}_A$ [cf. Eq. (B41)], and $H_{A\,\text{eff}}$. However, $\mathfrak{C}_{IA}$ and $\mathfrak{C}_{JA}$ will still be the untransformed operators; denote $U_A{}^\dagger\mathfrak{C}_{IA}U_A$ by $\mathfrak{C}_{IA}'''$ and $U_A{}^\dagger\mathfrak{C}_{JA}U_A$ by $\mathfrak{C}_{JA}'''$. The equations for $R_{An}$, $Q_{An}$, etc. are now obtained from Eqs. (B4)–(B13) by replacing $H_I$ by $\mathfrak{C}_{IA}'''$ and by inserting an over-all scale factor $J_A$ in the formula for $H_{A\,\text{eff}}$.

TABLE V. Bounds on operators needed in the proof of Theorem 4. The bound $(d,e,f)$ is defined by Eq. (B30) except that the vector $\mathbf{u}=(1/\sqrt{2})$ (1,1,1) replaces the vector $\mathbf{m}$.

| Operator | Bound |
|---|---|
| $\mathbf{M}_a\cdot\mathbf{V}_1$ | $d_1(5.5,0,0)$ |
| $V_A$ | $\sqrt{2}d_1$ |
| $\mathfrak{IC}_{Ja}$ | $d_1(450\Lambda^{-1},450\Lambda^{-1},300)+d_2(\Lambda^{-1},\Lambda^{-1},1)$ |
| $R_{an},R_a$ | $d_1(8,800,1600)+d_2(30\Lambda^{-1},2200\Lambda^{-1},1.3)$ |
| $Q_{an},Q_a$ | $d_1(3.25\times10^5\Lambda^{-1},245,3.25\times10^5\Lambda^{-1})+d_2(7.9\Lambda^{-1},1100\Lambda^{-1},2200\Lambda^{-1})$ |
| $\bar{Q}_{an},\bar{Q}_a$ | $d_1(3.3\times10^5\Lambda^{-1},\ 250,\ 3.3\times10^5\Lambda^{-1})+d_2(8\Lambda^{-1},1200\Lambda^{-1},2400\Lambda^{-1})$ |
| $\bar{H}_a$ | $d_1(7200\Lambda^{-1},230,310)+d_2(16\Lambda^{-1},27\,000\Lambda^{-1},1.1)$ |
| $\bar{H},\bar{H}_A$ | $(20\,000\Lambda^{-1},40,120)$ |

Now define the following differences:

$$\begin{aligned}\mathfrak{IC}_{Ia}&=\mathfrak{IC}_{IA}'''-\mathfrak{IC}_I,\\ \mathfrak{IC}_{Ja}&=\mathfrak{IC}_{JA}'''-\mathfrak{IC}_J,\\ \mathbf{M}_a&=\mathbf{M}_A'''-\mathbf{M},\\ R_{an}&=R_{An}-R_n,\ \text{etc.},\end{aligned}\tag{B83}$$

where

$$\mathbf{M}_A'''=U_A^\dagger\mathbf{M}_AU_A.\tag{B84}$$

One can write equations for the differences $R_{an}$, etc., as follows:

$$R_{a0}=0,\tag{B85}$$

$$\begin{aligned}R_{an}=(-\mathfrak{IC}_0)^{-1}(1-P)[&(1-R_{An-1})\mathfrak{IC}_{Ia}(P+R_{An-1})\\&-R_{an-1}\mathfrak{IC}_I(P+R_{An-1})+(1-R_{n-1})\mathfrak{IC}_IR_{an-1}]\\&(n>0),\end{aligned}\tag{B86}$$

$$Q_{an}=\tfrac{1}{2}(R_A^\dagger R_a+R_a^\dagger R-Q_{An-1}Q_{an-1}-Q_{an-1}Q_{n-1})\qquad(n>0),\tag{B87}$$

$$\bar{Q}_{an}=-Q_a-Q_A\bar{Q}_{an-1}-Q_a\bar{Q}_{n-1},\tag{B88}$$

$$\begin{aligned}\bar{H}_a=&P\mathfrak{IC}_{Ja}P+Q_A\mathfrak{IC}_{Ia}(P+R_A)(P+\bar{Q}_A)\\&+Q_a\mathfrak{IC}_I(P+R_A)(P+\bar{Q}_A)+Q\mathfrak{IC}_IR_a(P+\bar{Q}_A)\\&+Q\mathfrak{IC}_I(P+R)\bar{Q}_a+P\mathfrak{IC}_{Ia}(R_A+\bar{Q}_A+R_A\bar{Q}_A)\\&+P\mathfrak{IC}_I(R_a+\bar{Q}_a+R_a\bar{Q}_A+R\bar{Q}_a).\end{aligned}\tag{B89}$$

Knowing upper bounds for $\bar{H}_a$, one easily obtains upper bounds for $U_A^\dagger\mathbf{G}_{kA}U_A-\mathbf{G}_k$ and $U_A^\dagger L_{kA}U_A-L_k$.

The first step in deriving upper bounds is to get upper bounds for $\mathfrak{IC}_{Ja}$ and $\mathbf{M}_a$. One has

$$\begin{aligned}\mathbf{M}_a=(&m_A-m,\sqrt{2}g_AU_A^\dagger\tau^+U_A-\sqrt{2}g\tau^+,\\&\sqrt{2}g_AU_A^\dagger\tau^-U_A-\sqrt{2}g\tau^-).\end{aligned}\tag{B90}$$

Write $U_A=1+V_A$; then

$$\begin{aligned}g_AU_A^\dagger\tau^+U_A-g\tau^+=&(g_A-g)U_A^\dagger\tau^+U_A\\&+gV_A^\dagger\tau^+U_A+g\tau^+V_A.\end{aligned}\tag{B91}$$

Thus one needs a bound for $V_A$. The operator $U_A$ is

$$U_A=\sum_{n=1}^{8}|n\rangle_A\langle n|,\tag{B92}$$

where $|n\rangle$ $(1\leq n\leq8)$ are the eigenstates of $H_0$ and $|n\rangle_A$ are the eigenstates of $H_{0A}$. These are known explicitly from Table I. An upper bound for $\|V_A\|^2$ is obtained by computing the trace of $V_A^\dagger V_A$. The trace is

$$\begin{aligned}\mathrm{Tr}V_A^\dagger V_A=&\mathrm{Tr}(2-U_A-U_A^\dagger)\\&=\sum_{n=1}^{8}(2-\langle n|n_A\rangle-\langle n_A|n\rangle).\end{aligned}\tag{B93}$$

In fact, one can compute the trace separately for states of a given charge; the maximum of these traces is still greater than $\|V_A\|^2$. The traces are as follows:

$$\text{charge}=2\text{ or }-1:\quad\mathrm{Tr}V_A^\dagger V_A=0,\tag{B94}$$

$$\begin{aligned}\text{charge}=0\text{ or }1:\quad&\mathrm{Tr}V_A^\dagger V_A\\&=2[(m-m_A)^2+2(g-g_A)^2].\end{aligned}\tag{B95}$$

The latter result was obtained using Table I and Eq. (B78). From this it follows that

$$\|V_A\|\leq\sqrt{2}d_1.\tag{B96}$$

Also, $|m-m_A|$ and $\sqrt{2}|g-g_A|$ are less than $d_1$ [from Eq. (B79)]; $\sqrt{2}g$ and $\sqrt{2}g_A$ are less than 1; and $\|U_A\|=1$. Using these results in Eq. (B90) gives

$$\|\mathbf{M}_a\|<5.5d_1\mathbf{u}.\tag{B97}$$

A similar calculation for $\mathfrak{IC}_{Ja}$, using the bound of Table IV for $\mathfrak{IC}_J$ and $\mathfrak{IC}_{JA}$, plus the bounds $\mathbf{m}<\sqrt{2}\mathbf{u}$ and $\mathbf{m}_A<\sqrt{2}\mathbf{u}$, gives the bound shown in Table V. One can now obtain bounds on $R_{an}$, $R_a$, etc., using Eqs. (B85)–(B89). One uses Eqs. (B31)–(B33) to obtain bounds on products [with $\mathbf{u}$ replacing $\mathbf{m}$ in the definition of the bound $(a,b,c)$]. The results are shown in Table V.

Write the decomposition of $\bar{H}_a$ as

$$\bar{H}_a=\sum_{k=1}^{N}\mathbf{V}_k\cdot D_{ak-1}+\sum_{k=0}^{N}F_{ak}.\tag{B98}$$

The bound of Table V for $\bar{H}_a$ gives the following bounds:

$$\begin{aligned}\|\mathbf{D}_{a0}\|&\leq(7200\Lambda^{-1}d_1+16\Lambda^{-1}d_2)\mathbf{u},\\ \|\mathbf{D}_{ak}\|&\leq(230d_1+27000\Lambda^{-1}d_2)\mathbf{u}\Lambda^{-k}\quad(k>0),\\ \|F_{ak}\|&\leq(230d_1+27000\Lambda^{-1}d_2)\Lambda^{-2k}\quad(k>0).\end{aligned}\tag{B99}$$

Consider the significance of $\mathbf{D}_{a0}$. It is a difference $\mathbf{D}_{A0}-\mathbf{D}_0$. From Eqs. (B51), (B54), and (B21) [one must substitute $\tau_R{}^{\pm}$ for $\tau^{\pm}$ in Eq. (B21)], $\mathbf{D}_0$ itself is

$$\mathbf{D}_0=(m''-m, \sqrt{2}[g''-g(1-g^2)]\tau_R{}^+, \sqrt{2}[g''-g(1-g^2)]\tau_R{}^-). \quad \text{(B100)}$$

Correspondingly,

$$\mathbf{D}_{A0}=(m_A''-m_A, \sqrt{2}[g_A''-g_A(1-g_A{}^2)]\tau_R{}^+, \sqrt{2}[g_A''-g_A(1-g_A{}^2)]\tau_R{}^-). \quad \text{(B101)}$$

Thus $\mathbf{D}_{a0}$ involves differences such as

$$(m_A''-m_A)-(m''-m).$$

We can use the bound on $\mathbf{D}_{a0}$ to prove the inequalities of Eq. (5.37) (the first inequality of Theorem 4). In the notation of this appendix the quantity $d_1'$ is defined as

$$(d_1')^2=(m_A'-m')^2+2(g_A'-g')^2, \quad \text{(B102)}$$

where

$$m'=\cos\theta'=m''/(m''^2+2g''^2)^{1/2}, \quad \text{(B103)}$$

$$g'=\sin\theta'=\sqrt{2}g''/(m''^2+2g''^2)^{1/2}, \quad \text{(B104)}$$

and analogous formulas hold for $m_A'$ and $g_A'$; $\theta'$ is the angle in the decomposition of $T(H)$ [cf. Eq. (B25)].

To get bounds on $d_1'$ requires some further manipulations which are most conveniently done with another set of vectors. Define the following two-dimensional vectors:

$$\mathbf{x}=(m, \sqrt{2}g(1-g^2)), \quad \text{(B105)}$$

$$\mathbf{x}''=(m'', \sqrt{2}g''), \quad \text{(B106)}$$

$$\mathbf{x}'=(m', \sqrt{2}g'), \quad \text{(B107)}$$

and analogously for $\mathbf{x}_A$, $\mathbf{x}_A''$, and $\mathbf{x}_A'$. Define

$$\hat{x}=|\mathbf{x}|^{-1}\mathbf{x}, \text{ etc.} \quad \text{(B108)}$$

Then

$$\mathbf{x}'=\hat{x}'', \quad \mathbf{x}_A'=\hat{x}_A''. \quad \text{(B109)}$$

Now one has

$$d_1'=|\mathbf{x}_A'-\mathbf{x}'|=|\hat{x}_A''-\hat{x}''|. \quad \text{(B110)}$$

The bound $d_1'$ will be computed in two parts, first relating $\hat{x}_A''-\hat{x}''$ to $\hat{x}_A-\hat{x}$ and then bounding $\hat{x}_A-\hat{x}$. Write

$$\phi=|\hat{x}_A''-\hat{x}''-\hat{x}_A+\hat{x}|, \quad \text{(B111)}$$

$$\psi=|\hat{x}_A-\hat{x}|. \quad \text{(B112)}$$

Then

$$\psi-\phi\leq d_1'\leq\psi+\phi. \quad \text{(B113)}$$

To compute $\psi$, it is convenient to let

$$\hat{x}=(\cos\omega, \sin\omega), \quad \hat{x}_A=(\cos\omega_A, \sin\omega_A), \quad \text{(B114)}$$

and $m=\cos\theta$, $\sqrt{2}g=\sin\theta$, $m_A=\cos\theta_A$, $\sqrt{2}g=\sin\theta_A$. Then[19]

$$\omega=\tan^{-1}[\tan\theta(1-\tfrac{1}{2}\sin^2\theta)]=f(\theta) \quad \text{(B115)}$$

[19] The function $f$ is not the function defined in Sec. V.

and $\omega_A$ is $f(\theta_A)$. The derivative $f'(\theta)$ has the form

$$f'(\theta)=N(y)/D(y), \quad \text{(B116)}$$

where

$$y=\sin^2\theta, \quad \text{(B117)}$$

$$N(y)=1-\tfrac{1}{2}y-y(1-y), \quad \text{(B118)}$$

$$D(y)=1-y^2+\tfrac{1}{4}y^3. \quad \text{(B119)}$$

Analyzing the form for $f'(\theta)$ one sees that the numerator $N$ decreases for $0<y<\frac{3}{4}$ and increases for $\frac{3}{4}<y$; the denominator $D$ decreases over the whole range $0<y<1$. Therefore one has the following bounds:

$$N(0.75)/D(0)<f'(\theta)<\max\{N(0)/D(0.75), N(1)/D(1)\}. \quad \text{(B120)}$$

Evaluated, this gives

$$0.4375<f'(\theta)<2. \quad \text{(B121)}$$

Hence by the mean-value theorem,

$$0.436|\theta_A-\theta|\leq|\omega_A-\omega|\leq 2|\theta_A-\theta|. \quad \text{(B122)}$$

Now the definitions of $d_1$ and $\psi$ are equivalent to

$$d_1=|2\sin\tfrac{1}{2}(\theta_A-\theta)|, \quad \text{(B123)}$$

$$\psi=|2\sin\tfrac{1}{2}(\omega_A-\omega)|. \quad \text{(B124)}$$

One can show that

$$(\sin az)>a\sin z \quad \text{(B125)}$$

when $0<z<\frac{1}{2}\pi$ and $0<a<1$. The result of Eqs. (B124), (B122), (B125), and then (B123) is

$$\psi\geq|2\sin 0.218(\theta_A-\theta)|\geq 0.436 d_1 \quad [z=0.5(\theta_A-\theta)], \quad \text{(B126)}$$

$$\psi\leq|2\sin(\theta_A-\theta)|\leq 2d_1 \quad [az=0.5(\theta_A-\theta)]. \quad \text{(B127)}$$

The next step is to bound $\phi$. It is convenient to define

$$\mathbf{g}(\mathbf{x})=\hat{x}. \quad \text{(B128)}$$

Then

$$\phi\leq\phi_1+\phi_2, \quad \text{(B129)}$$

where

$$\phi_1=|\mathbf{g}(\mathbf{x}_A'')-\mathbf{g}(\mathbf{x}_A+\delta\mathbf{x})|, \quad \text{(B130)}$$

$$\phi_2=|\mathbf{g}(\mathbf{x}_A+\delta\mathbf{x})-\mathbf{g}(\mathbf{x}+\delta\mathbf{x})-\mathbf{g}(\mathbf{x}_A)+\mathbf{g}(\mathbf{x})|, \quad \text{(B131)}$$

and

$$\delta\mathbf{x}=\mathbf{x}''-\mathbf{x}. \quad \text{(B132)}$$

Now by the mean-value theorem

$$\phi_1\leq\max(0\leq\lambda\leq 1)|\delta x_a\cdot\nabla\mathbf{g}(\mathbf{x}_A''-\lambda\delta\mathbf{x}_a)|, \quad \text{(B133)}$$

where

$$\delta\mathbf{x}_a=\mathbf{x}_A''-\mathbf{x}''-\mathbf{x}_A+\mathbf{x}, \quad \text{(B134)}$$

and by a second-order mean-value theorem

$$\phi_2\leq\max(0\leq\lambda\leq 1, 0\leq\mu\leq 1)\times|(\delta\mathbf{x}\cdot\nabla)(\mathbf{x}_a\cdot\nabla)\mathbf{g}(\mathbf{x}+\lambda\delta\mathbf{x}+\mu\mathbf{x}_a)|, \quad \text{(B135)}$$

where

$$\mathbf{x}_a=\mathbf{x}_A-\mathbf{x} \tag{B136}$$

(and $\delta\mathbf{x}\cdot\nabla$ acts on $\mathbf{g}$, not on $\mathbf{x}_a\cdot\nabla$). From Eqs. (B56), (B57), (B78), (B105), and (B106),

$$|\delta\mathbf{x}|\leq 0.005, \tag{B137}$$

and from Eq. (B99),

$$|\delta\mathbf{x}_a|\leq(7200\Lambda^{-1}d_1+16\Lambda^{-1}d_2). \tag{B138}$$

Furthermore,

$$|\mathbf{x}_a|^2=(m_A-m)^2+2(g_A-g)^2(1-g_A{}^2-g_Ag-g^2)^2, \tag{B139}$$

and since $g_A$ and $g$ are less than $1/\sqrt{2}$, one has

$$-\tfrac{1}{2}<(1-g_A{}^2-g_Ag-g^2)<1. \tag{B140}$$

Therefore

$$|\mathbf{x}_a|^2<d_1{}^2. \tag{B141}$$

Now let $\mathbf{y}$ be an arbitrary vector; one can most easily compute $\delta\mathbf{x}_a\cdot\nabla\mathbf{g}(\mathbf{y})$ and $(\delta\mathbf{x}\cdot\nabla)(\mathbf{x}_a\cdot\nabla)\mathbf{g}(\mathbf{y})$ using a coordinate system with the first parallel to $\mathbf{y}$. If $\delta x_{\parallel}$ and $\delta x_\perp$ are the components of $\delta\mathbf{x}$ parallel and perpendicular to $y$ (and likewise for $x_{a\parallel}$, etc.), one has

$$\delta\mathbf{x}_a\cdot\nabla\mathbf{g}(\mathbf{y})=(0,\delta x_{a\perp}|\mathbf{y}|^{-1}) \tag{B142}$$

and

$$(\delta\mathbf{x}\cdot\nabla)(\mathbf{x}_a\cdot\nabla)\mathbf{g}(\mathbf{y})=(-\delta x_\perp x_{a\perp}|\mathbf{y}|^{-2}, -(\delta x_\parallel x_{a\perp}+\delta x_\perp x_{a\parallel})|\mathbf{y}|^{-2}). \tag{B143}$$

In absolute value

$$|(\delta\mathbf{x}_a\cdot\nabla)\mathbf{g}(\mathbf{y})|\leq|\delta\mathbf{x}_a|\,|\mathbf{y}|^{-1}, \tag{B144}$$

$$|(\delta\mathbf{x}\cdot\nabla)(\mathbf{x}_a\cdot\nabla)\mathbf{g}(\mathbf{y})|\leq(2/\sqrt{3})|\delta\mathbf{x}|\,|\mathbf{x}_a|\,|\mathbf{y}|^{-2}. \tag{B145}$$

The second inequality is proved using the relation

$$2x_{a\parallel}x_{a\perp}\delta x_\parallel\delta x_\perp<\tfrac{4}{3}(\delta x_\parallel x_{a\parallel})^2 +\tfrac{1}{3}[(\delta x_\parallel x_{a\perp})^2+(\delta x_\perp x_{a\parallel})^2+(\delta x_\perp x_{a\perp})^2]. \tag{B146}$$

To use the bound (B144) to obtain a bound for $\phi_1$ [cf. Eq. (B133)], one puts $\mathbf{y}=\mathbf{x}_A{}''-\lambda\delta\mathbf{x}_a$; hence

$$|\mathbf{y}|^{-1}\leq(|\mathbf{x}_A{}''|-|\delta\mathbf{x}_a|)^{-1}. \tag{B147}$$

Now [from Eq. (B137) and the analogous bound for $|\mathbf{x}_A{}''-\mathbf{x}_A|$]

$$|\delta\mathbf{x}_a|\leq|\delta\mathbf{x}|+|\mathbf{x}_A{}''-\mathbf{x}_A|\leq 0.01 \tag{B148}$$

and $|\mathbf{x}_A{}''|>0.49$ from Eq. (B67) (which holds for $|\mathbf{x}_A{}''|$ as well as $|\mathbf{x}''|$). Thus (remember that $\Lambda>4\times10^6$)

$$\phi_1\leq 2.1(7200\Lambda^{-1}d_1+16\Lambda^{-1}d_2)\leq 0.004d_1+10^{-5}d_2. \tag{B149}$$

To get a bound for $\phi_2$, one uses Eq. (B145) with $\mathbf{y}$ being $(1-\mu)\mathbf{x}+\mu\mathbf{x}_A+\lambda\delta\mathbf{x}$. Since $\mathbf{x}\cdot\mathbf{x}_A\geq 0$ and since $\mu$ and $1-\mu$ are non-negative,

$$|(1-\mu)\mathbf{x}+\mu\mathbf{x}_A|^2\geq(1-\mu)^2|\mathbf{x}|^2+\mu^2|\mathbf{x}_A|^2. \tag{B150}$$

But $|\mathbf{x}|$ and $|\mathbf{x}_A|$, and $(1-\mu)^2+\mu^2$, are all larger than or equal to $\frac{1}{2}$. So

$$|(1-\mu)\mathbf{x}+\mu\mathbf{x}_A|\geq 1/\sqrt{8}. \tag{B151}$$

Hence

$$|\mathbf{y}|\geq(1/\sqrt{8})-0.005>\tfrac{1}{3}. \tag{B152}$$

Hence from Eqs. (B135), (B145), (B152), (B137), and (B141),

$$\phi_2\leq 6\sqrt{3}\,|\delta\mathbf{x}|\,|\mathbf{x}_a|\leq 0.052d_1. \tag{B153}$$

From Eqs. (B113), (B126), (B127), (B129), (B149), and (B153),

$$0.38d_1-10^{-5}d_2<d_1{}'<20d_1+10^{-5}d_2, \tag{B154}$$

which is the first inequality of Theorem 4.

To obtain the second inequality of Theorem 4, one starts from Eqs. (B26), (B27), (B52), (B53), (B106), and the corresponding equations for $A_{kA}{}'$, etc., from which one can obtain

$$A_{kA}{}'-A_k{}'=(|\mathbf{x}_A{}''|^{-1}-|\mathbf{x}''|^{-1})D_{Ak+1} +|\mathbf{x}''|^{-1}Da_{k+1}, \tag{B155}$$

$$C_{kA}{}'-C_k{}'=\Lambda(|\mathbf{x}_A{}''|^{-1}-|\mathbf{x}''|^{-1})F_{Ak+1} -\Lambda|\mathbf{x}''|^{-1}F_{ak+1}. \tag{B156}$$

Now, from Eqs. (B67), (B134), (B136), (B138), and (B141),

$$|\mathbf{x}_A{}''|^{-1}-|\mathbf{x}''|^{-1} \leq|\mathbf{x}_A{}''-\mathbf{x}''|\,|\mathbf{x}_A{}''|^{-1}|\mathbf{x}''|^{-1}\leq 4.07|\delta\mathbf{x}_a+\mathbf{x}_a| \leq 4.07(d_1+7200\Lambda^{-1}d_1+16\Lambda^{-1}d_2). \tag{B157}$$

From Table V, $\mathbf{D}_{Ak+1}$ and $F_{Ak+1}$ have bounds

$$\|\mathbf{D}_{Ak+1}\|\leq 40\mathbf{u}\Lambda^{-k-1}, \tag{B158}$$

$$\|F_{Ak+1}\|\leq 40\Lambda^{-2k-2}. \tag{B159}$$

From Eqs. (B155)–(B159), (B67), and (B99), one gets

$$\|A_{kA}{}'-A_k{}'\|\leq(1100d_1+0.06d_2)\mathbf{u}\Lambda^{-k-1}, \tag{B160}$$

$$\|C_{kA}{}'-C_k{}'\|-(1100d_1+0.06d_2)\Lambda^{-2k-1}, \tag{B161}$$

which proves the second inequality of Theorem 4.