

## DUALITY TRANSFORMATIONS FOR GENERAL ABELIAN SYSTEMS

Robert SAVIT

*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

Received 14 July 1981

We describe the general structure of duality transformations for a very broad set of abelian statistical and field theoretic systems. This includes theories with many different types of fields and a large variety of kinds of interactions including, but not limited to nearest neighbor, next nearest neighbor, multi-spin interactions, etc. We find that the dual form of a theory does not depend directly on the dimensionality of the theory, but rather on the number of fields and number of different kinds of interactions. The dual forms we find have a generalized gauge symmetry and possess the usual property of having a temperature (or coupling constant) which is inverted from that of the original theory. Our results reduce to the well-known results in those particular cases that have heretofore been studied. Our procedure also suggests variations capable of generating other forms of the dual theory which may be useful in various specific cases.

### 1. Introduction

Duality transformations have been very helpful in elucidating the properties of a large number of statistical and field theoretic systems [1]. These (exact) transformations map the high (low) temperature region of a theory into the low (high) temperature region of another, dual theory\*. The dual theory is expressed in terms of dual or disorder variables. Properly defined expectation values and correlation functions of these disorder variables measure the disorder of the original theory (or the order of the dual theory) in the same way that functions of the original variables constitute measures of the order of the theory. Intelligent and judicious use of the dual form of a theory can reveal much about the nature of its high and low temperature regions and its phase structure and can provide much information that is intrinsically non-perturbative.

Most of the theories that have been studied with the aid of duality transformations are of the usual type of abelian spin or gauge theories. For instance, for a scalar field the kinetic energy terms (in the field theoretic sense) involve the usual laplacian. If gauge fields are involved the kinetic energy has the usual field strength

\* In the body of the paper we will usually use the nomenclature appropriate to statistical mechanics. The field theorist should replace “hamiltonian” by “lagrangian”, “space dimensions” by “space-time dimensions”, and “high (low) temperature” by “strong (weak) coupling”.

form or some straightforward generalization thereof, such as the Wilson action [2]. However, there are large numbers of abelian theories of statistical (and perhaps field theoretic) interest that are somewhat more complicated and do not fall into the above categories. Statistical theories with next nearest neighbor as well as nearest neighbor interactions, or field theories with propagators which have behavior other than the usual  $k^{-2}$  are examples of such systems. It is the purpose of this paper to generalize the formalism of duality transformations to include these other theories.

Actually, the generalization is quite straightforward, but the mathematical structure that emerges is cute and there are some important lessons to be learned from the investigation. In our more general case we find that duality transformations still map high to low temperature regions, and vice versa. On the other hand, we learn that the form of the dual theory does not depend on the number of space (or, for a field theory space-time) dimensions. Rather, the form of the dual theory depends only on the number of independent fields and the number of different kinds of interactions in the original Hamilton. (This latter concept will be made more precise in the next section.) In most standard theories heretofore studied, the number of different kinds of interactions (for instance, the number of terms in a kinetic energy laplacian) was a simple function of the dimension and led to the understanding that the form of a dual theory depended on dimension. (For example, a spin theory is dual to a spin theory in two dimensions and dual to a gauge theory in three dimensions.) In the more general theories discussed in this paper this is not necessarily the case, and the form of the dual theory is not a direct function of dimension. (For instance, ref. [3] discusses a three-dimensional self-dual scalar theory in some detail.)

To arrive at the dual form of a general abelian theory we must solve a set of  $m$  linear homogenous equations in  $n$  unknowns with  $n \geq m$ , where  $n$  is the number of kinds of interactions and  $m$  is the number of independent fields in the original hamiltonian. The solution is expressed in terms of  $n - m$  independent functions which become the dual fields. Since the equations are linear and homogeneous there is a good deal of freedom in the form of the solution which one is allowed to choose. Our solutions are normalized in such a way that the dual form of the hamiltonian contains no explicit inverses of the operators which appear in the original hamiltonian. Thus, if the original hamiltonian contains only short-range interactions, then so does the dual hamiltonian. Furthermore, for  $n - m \geq 2$ , the solutions to the equations cry out for the addition of certain dependent fields to render the solutions symmetric and aesthetic. The dual form of the theory then has the form of a generalized gauge theory with a local gauge symmetry. When we apply our formalism to the standard theories studied previously we recover the usual results.

Because of the very general nature of our results, much of the paper is, unfortunately, rather formal. At first sight the formalism appears somewhat complex. However, it really has quite a simple structure. Moreover, in specific problems the results usually simplify considerably and the content and usefulness of the procedure

becomes more transparent. In the paper we briefly describe a couple such examples. Another specific application of our results can be found in ref. [3].

The rest of the paper is organized as follows: Sect. 2 first describes the general class of theories we shall consider and outlines the strategy for carrying out the duality transformations. After a brief discussion of some simple examples, the general solution of the linear equations needed to complete the duality transformation is presented. A discussion of the gauge invariance of the solutions follows and the section concludes with a brief mathematical digression. Sect. 3 addresses the problem of the geometric interpretation of the dual theory. Unlike in the more garden variety theories, in the general case this interpretation is not always so straightforward. Sect. 4 contains comments and a summary of our results.

## 2. General structure of the dual form

### 2.1. STRATEGY AND DEFINITION OF THE THEORIES

Consider a statistical system described by a hamiltonian of the form

$$-\beta\mathcal{H} = \sum_x \sum_{i=1}^n V_i \left( \sum_{j=1}^m \bar{R}_i^j \phi_j(x) \right). \quad (2.1)$$

That is, there are  $n$  different kinds of potentials in the hamiltonian each one of which is a function of a linear combination of a set of  $m$  different fields,  $\{\phi\}$ . These fields may be, for example, scalar fields, vector or tensor fields with various directional indices or some combination thereof. The sum over  $x$  represents a sum over  $d$ -dimensional (euclidean) space (or space-time). The  $\bar{R}_i^j$  are a set of operators acting on the  $\phi$ 's. As will become clear later, the  $R$ 's should commute, certain combinations of them should be invertible, and they should also satisfy a certain "integration by parts" criterion. We have in mind that the  $\bar{R}$ 's are short-range derivative-like or difference operators. However, this may not be necessary and many of our results should, in principle, generalize to systems with explicit long-range interactions.

The partition function for the system (2.1) is

$$Z = \sum_{\{\phi\}} e^{-\beta\mathcal{H}}. \quad (2.2)$$

We assume that our system is abelian. There are well-known difficulties in applying duality transformations to most non-abelian theories, and these difficulties will be present also in non-abelian versions of the cases we wish to discuss.

Two comments should be made about the set of fields  $\{\phi\}$ . First, their measure depends on the symmetry of the theory. For abelian theories they may be either

discrete or continuous fields in  $(-\infty, \infty)$  or angles. We will discuss this point further in a moment. Second, the set of fields,  $\{\phi\}$ , for our purposes does not include the auxiliary fields which are sometimes introduced into a theory to enforce a certain symmetry. An example of these auxiliary fields are the link variables which appear in the Villain approximation to the  $x$ - $y$  model [4]. Any auxiliary fields of this type can be taken into account by modifying (2.1) and (2.2) to read

$$-\beta\mathcal{H} = \sum_x \sum_{i=1}^n V_i \left( \sum_{j=1}^m \bar{R}_i^j \phi_j(x) + t_i \right), \quad (2.1a)$$

$$Z = \sum_{\{t\}} \sum_{\{\phi\}} e^{-\beta\mathcal{H}}, \quad (2.2a)$$

where the measure of the  $t_i$ 's depends on the symmetry of the theory. From the point of view of duality transformations, these auxiliary fields have the effect of modifying the measure of the dual variables. They do not affect the form of the dual hamiltonian which is our main concern in this paper. Therefore, in what follows we shall not explicitly include them. For some specific examples of the effect of these auxiliary fields, the reader should consult refs. [1, 3, 4].

To derive the dual form of (2.2) we introduce a Fourier conjugate variable,  $l_i(x)$  for each term in the hamiltonian (2.1) and write

$$Z = C \sum_{\{l\}} \sum_{\{\phi\}} \prod_x \prod_{i=1}^n \exp \left[ \tilde{V}_i(l_i) + il_i \sum_{j=1}^m \bar{R}_i^j \phi_j \right], \quad (2.3)$$

where

$$e^{V_i(s)} \propto \sum_l e^{\tilde{V}_i(l) + ils}, \quad (2.4)$$

and  $C$  denotes a collection of field-independent factors which are irrelevant for our purposes. The measure of the  $l_i$  is determined by the symmetry of the  $V_i$ 's in the usual way for Fourier transforms. Typically the  $l$ 's will be continuous or discrete valued in  $(-\infty, \infty)$  or will be continuous or discrete value angles.

We now integrate by parts the last terms in (2.3). We assume that the  $\bar{R}_i^j$  are operators such that

$$\sum_x \sum_{i=1}^n \sum_{j=1}^m l_i \bar{R}_i^j \phi_j = \sum_x \sum_{j=1}^m \phi_j \sum_{i=1}^n R_i^j l_i. \quad (2.5)$$

The  $R$ 's will usually be simply related to the  $\bar{R}$ 's, such as differing from them by a sign (as would be the case for a simple difference.) But more complex relationships are also allowed. For instance, if  $\Delta_x$  is a difference operator and if  $\bar{R}_i^j = \alpha \Delta_x^2 + \beta \Delta_x$ ,

then  $R_i^j = \alpha \Delta_x^2 - \beta \Delta_x$ . In writing (2.5) we have assumed certain boundary conditions such that no surface terms appear. The specific nature of the necessary boundary conditions depends on the set of operators,  $R$ . The general solutions to (2.7) which we shall present below also imply certain conditions on the behavior of the  $l_i$ 's at infinity, and our choice of boundary condition should also be consistent with this behavior. A specific simple example is discussed near the end of sect. 3. (See also refs. [3, 1] and references therein for more examples.)

Having found a set  $\{R_i^j\}$  satisfying (2.5), we rewrite (2.3) as

$$\begin{aligned} Z &= C \sum_{\{l\}} \prod_x \exp \left[ \sum_{i=1}^n \tilde{V}_i(l_i) \right] \prod_{j=1}^m \sum_{\phi_j(x)} \exp \left[ i \phi_j(x) \sum_{i=1}^n R_i^j l_i \right] \\ &= C \sum_{\{l\}} \exp \left[ \sum_x \sum_{i=1}^n \tilde{V}_i(l_i) \right] \prod_x \prod_{j=1}^m \delta \left[ \sum_{i=1}^n R_i^j l_i \right], \end{aligned} \tag{2.6}$$

where the delta functions in (2.6) are the ones appropriate to the group structure of the theory. For instance, if the  $V$ 's were periodic functions of continuous valued angles,  $\phi$ , then the  $l$ 's would be integers and the delta functions would be Kronecker delta functions. Clearly, in this example it would also be advantageous [for simplicity in (2.6)] if the operators  $R_i^j$  were such that the linear combination  $\sum_i R_i^j l_i$  were also integer valued. Such things can usually be arranged.

We now want to find a representation for the  $l_i$  in terms of another set of variables which automatically satisfies the delta functions in (2.6). In what follows we will consider the case in which  $l$ 's can range continuously in  $(-\infty, \infty)$  and the delta functions in (2.6) are ordinary Dirac delta functions. In most other abelian cases in which the  $l_i$ 's have a different measure the form of the solution to (2.7) which we shall present should also be applicable. (See, for example, ref. [3].) However, depending on the  $R_i^j$ 's this may not always be the case, and modifications may be necessary in particular cases.

With this caveat in mind, we seek a solution of the set of linear homogenous equations

$$\sum_{i=1}^n R_i^j l_i = 0, \quad j = 1, \dots, m, \quad (m < n). \tag{2.7}$$

Since we have  $m$  constraints and  $n$  unknowns, we expect that the solution will be expressible in terms of  $n - m$  arbitrary independent fields. (From our point of view, the case  $n = m$  is trivial; (2.2) then contains no dynamics.) Furthermore, because the equations are homogeneous we may multiply any solution by a uniform factor to obtain another solution. Now, we shall often be faced with a system in which the  $R$ 's are short range, and so, to avoid explicit long-range interactions in the dual theory, we shall seek a solution in which the  $l$ 's are expressed in a form which contains no

factors of inverses of the  $R$ 's. Finally, the solutions we find will, when  $n - m > 1$ , have various kinds of gauge symmetries. For example, when  $n = m + 2$  our solution will be expressed in terms of a three-component "vector" field with a local gauge symmetry so that there are only two independent degrees of freedom.

Our procedure for solving (2.7) is as follows: We use the equation with  $j = k$  to eliminate  $l_k$  in the remaining equations starting with  $k = 1$  and proceeding sequentially to  $k = m - 1$ . We then solve the remaining one equation in  $n - m + 1$  unknowns by regarding it as a kind of generalized divergenceless condition and expressing  $(l_n, l_{n-1}, \dots, l_m)$  as a generalized curl. This solution is substituted in the other equations and a solution for  $l_1, l_2, \dots, l_{m-1}$  is obtained.

## 2.2. SOME SIMPLE EXAMPLES

Before writing down the general solution, it may be useful to present a couple of simple examples. Suppose first that  $n = 2$ ,  $m = 1$ . Then we have

$$R_1^1 l_1 + R_2^1 l_2 = 0. \quad (2.8)$$

Assuming the appropriate boundary conditions, that  $(R_1^1)^{-1}$  exists, and that the  $R$ 's commute, we have

$$\begin{aligned} l_1 &= - (R_1^1)^{-1} R_2^1 l_2 = R_2^1 \left[ - (R_1^1)^{-1} l_2 \right] \equiv R_2^1 A, \\ l_2 &\doteq - R_1^1 A. \end{aligned} \quad (2.9)$$

A lightning review of the steps leading to (2.6) indicates that in this case the theory has a self-dual character in that the original hamiltonian was a function of  $(R_1^1 \phi, R_2^1 \phi)$ , while the dual hamiltonian is a function of  $(R_2^1 A, R_1^1 A)$ . Note that this feature is absolutely independent of the number of spatial dimensions but depends only on the fact that the original hamiltonian has two terms involving one field.

As a second example, consider the case  $n = 4$ ,  $m = 2$ . If  $(R_1^1)^{-1}$  exists, we can use the  $j = 1$  equation to eliminate  $l_1$  in the  $j = 2$  equation and we have

$$\sum_{i=2}^4 R_{1i}^{12} l_i = 0, \quad (2.10)$$

where  $R_{ij}^{\alpha\beta} \equiv R_i^\alpha R_j^\beta - R_j^\alpha R_i^\beta$ . Assuming appropriate behavior at infinity, (2.10) is then solved by

$$l_i = -\frac{1}{2} \varepsilon_{ijk} R_{1j}^{12} A_k, \quad i, j, k = 2, 3, 4, \quad (2.11a)$$

where  $\varepsilon_{234} = +1$  and changes sign under an interchange of the indices. This is easily

recognized as a generalization of the usual expression for a gauge theory field strength tensor in three dimensions in terms of the vector potential with  $R_{ij}^{12}$  playing the role of the covariant derivative. Finally, using (2.11a) and the  $j = 1$  equation we find

$$l_i = \frac{1}{2} \epsilon_{ijk} R_{ij}^{12} A_k, \quad i, j, k = 2, 3, 4. \quad (2.11b)$$

(2.11a) and (2.11b) can be combined into a more symmetric form which can also be generalized to arbitrary  $m$  and  $n$ . Before discussing that, however, we point out that the solutions (2.11) are invariant under a gauge transformation on the  $A_i$ , namely

$$A_i \rightarrow A_i + R_{i1}^{12} \Lambda, \quad (2.12)$$

where  $\Lambda$  is an arbitrary scalar function. For later purposes it is important to note that we may also define a fourth component of  $A_i$ ,  $A_1$  which we will take to be zero. Then,  $A_1$  remains zero under the transformation (2.12) since  $R_{11}^{12} = 0$ .

### 2.3. GENERAL SOLUTION TO THE CONSTRAINT EQUATIONS

We proceed now to describe a general (gauge invariant) solution to the set of equations (2.7). This section contains no proofs. But, after sufficient rumination it should be fairly clear that the form we present does constitute a correct solution. First some definitions are in order. We define a sequence of operators with  $2^P$  superscripts and subscripts composed of the  $R_i^j$ :

$$\begin{aligned} R_{ij}^{\alpha\beta} &\equiv R_i^\alpha R_j^\beta - R_j^\alpha R_i^\beta, \\ R_{ijkl}^{\alpha\beta\gamma\delta} &\equiv R_{ij}^{\alpha\beta} R_{kl}^{\gamma\delta} - R_{kl}^{\alpha\beta} R_{ij}^{\gamma\delta}, \\ R_{ijklmnop}^{\alpha\beta\gamma\delta\epsilon\phi\rho\sigma} &= R_{ijkl}^{\alpha\beta\gamma\delta} R_{mnop}^{\epsilon\phi\rho\sigma} - R_{mnop}^{\alpha\beta\gamma\delta} R_{ijkl}^{\epsilon\phi\rho\sigma}, \quad \text{etc.} \end{aligned} \quad (2.13)$$

Next we define a sequence of  $2^P$  integers, which we shall call the “standard sequence”, as follows: The standard sequence of  $2^P$  integers is generated from the standard sequence with  $2^{P-1}$  integers by first listing the sequence with  $2^{P-1}$  integers, then repeating the list adding one to the final integer. The standard sequence with  $P = 0$  is 1. The first few standard sequences are

$$\begin{aligned} &1 \\ &12 \\ &12 \ 13 \\ &12 \ 13 \ 12 \ 14 \\ &12 \ 13 \ 12 \ 14 \ 12 \ 13 \ 12 \ 15 \\ &\text{etc.} \end{aligned} \quad (2.14)$$

Finally, the  $R$  whose  $2^P$  superscripts appear in the standard sequence will be defined

as  $S$ . For example,

$$S_i = R_i^1, \quad S_{ij} = R_{ij}^{12}, \quad S_{ijkl} = R_{ijkl}^{1213},$$

$$S_{ijklmnop} = R_{ijklmnop}^{12131214}, \quad \text{etc.} \tag{2.15}$$

An important property of the  $S$ 's (and the  $R$ 's) which can easily be gleaned from (2.13) is the following. Partition the  $2^P$  subscripts into groups of  $2^{P-1}, 2^{P-2}, \dots, 2$ . An example for the case of 16 subscripts is shown in fig. 1.  $S$  is antisymmetric under the interchange of any pair of  $2^{k-1}$  consecutive strings of indices which are enclosed in the same box of length  $2^k$ . So for example,  $S_{i_1 \dots i_{16}} \rightarrow -S_{i_1 \dots i_{16}}$  under the interchange of  $i_5 \leftrightarrow i_6$  or  $(i_9 i_{10}) \leftrightarrow (i_{11} i_{12})$ , or  $(i_1 i_2 i_3 i_4) \leftrightarrow (i_5 i_6 i_7 i_8)$ , but not under the interchange of  $i_6 \leftrightarrow i_7$  or  $(i_{11} i_{12}) \leftrightarrow (i_{13} i_{14})$ .

We are now in a position to present the general solution. First consider the case  $m = n - 1$ . The solution can be written

$$l_i = 2^{-(n-2)(n-3)/2} \sum_{\text{perms}} (-1)^P S_{\sigma(i,n)} A, \quad i = 1, \dots, n, \tag{2.16}$$

where  $\sigma(i, n)$  is a set of  $2^{n-2}$  subscripts and  $A$  is an arbitrary function (which plays the role of a scalar field). The first factor in (2.16) is for simple normalization. The sum is over all permutations of the  $2^{n-2}$  subscripts and  $P$  is a permutation factor. The set  $\sigma(i, n)$  may be determined as follows:  $\sigma(n, n)$  is just the set of integers contained in the standard sequence of length  $2^{n-2}$ . The set  $\sigma(n-1, n)$  has the same elements as the set  $\sigma(n, n)$  except that the integer  $(n-1)$  is replaced by the integer  $n$ .  $\sigma(n-2, n)$  has the same elements as  $\sigma(n-1, n)$  except that the integer  $(n-2)$  is replaced by the integer  $(n-1)$ . In general,  $\sigma(n-j, n)$  has the same elements as  $\sigma(n-j+1, n)$  except that one of the integers with value  $(n-j)$  is replaced by the integer  $(n-j+1)$ . For example, for  $n = 5$ , we have

$$\begin{aligned} \sigma(5,5) &= \{1, 1, 1, 1, 2, 2, 3, 4\}, \\ \sigma(4,5) &= \{1, 1, 1, 1, 2, 2, 3, 5\}, \\ \sigma(3,5) &= \{1, 1, 1, 1, 2, 2, 4, 5\}, \\ \sigma(2,5) &= \{1, 1, 1, 1, 2, 3, 4, 5\}, \\ \sigma(1,5) &= \{1, 1, 1, 2, 2, 3, 4, 5\}. \end{aligned} \tag{2.17}$$

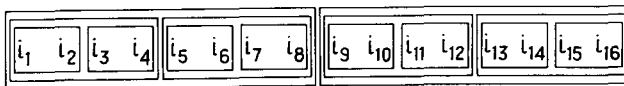


Fig. 1. An example of 16 subscripts and their partitioning which illustrates the antisymmetry properties of the operators,  $S$ . [See the discussion following eq. (2.15).]



We now must determine the permutation index,  $P$ . Consider some definite ordering of the elements of  $\sigma(i, n)$ . Define a process which takes this sequence of integers into the standard sequence of length  $2^{n-2}$ . If that process involves an even (odd) number of operations then  $P = 0$  (1). There are two types of operations allowed in the process: (i) Interchange in a box. Two subsequences of length  $2^{k-1}$  contained in a box of length  $2^k$  (see e.g. fig. 1) can be interchanged. (ii) Replacement. A sequence of length  $2^k$  (wholly contained in a box of length  $2^k$  [including  $k = 0$ ]) which differs from the standard sequence of length  $2^k$  only in its last digit can be replaced by the standard sequence of length  $2^k$ . For each value of  $k$  this need be done at most once. Moreover, one should not produce any intermediate sequences  $\rho$ , such that  $S_\rho \equiv 0$  due to the symmetry discussed following (2.15).

These rules are harder to express in prose than they turn out to be in practice as the following simple example will illustrate. Consider the case  $n = 5, i = 2$ . Using (2.16) and (2.17), our rules for determining  $P$ , and recalling the symmetry properties of the  $S$ 's, we have

$$I_2 = (S_{12131415} + S_{12141513} + S_{12151314})A. \tag{2.18}$$

To show that the sign for, say, the third term in (2.18) is correct consider the sequence of operations

$$12151314 \rightarrow 12151214 \rightarrow 12131214 \tag{2.19a}$$

or

$$12151314 \rightarrow 12151312 \rightarrow 12151213 \rightarrow 12131215 \rightarrow 12131214. \tag{2.19b}$$

Both (2.19a) and (2.19b) involve an even number of legal operations so that  $S_{12151314}$  appears with a plus sign.

Next consider the case  $m = n - 2$ . The solution can be written

$$I_i = 2^{-(n-3)(n-4)/2} \sum_{\text{perms}} (-1)^P [SA]_{\tau(i, n)}. \tag{2.20}$$

Here  $A$  is a three-component "vector" field,  $A_\beta$ , with  $\beta = n, n - 1, n - 2$ . (Alternatively, we may consider  $A_\beta$  an  $n$ -component field with  $A_1 = A_2 = \dots = A_{n-3} = 0$ .)  $S$  is an operator with  $2^{n-2}$  indices, and  $\tau(i, n)$  is a set of  $2^{n-3} + 1$  integers. The sum is over all possible permutations of these  $2^{n-3} + 1$  integers among the  $2^{n-3}$  subscript positions of  $S$  and the one of  $A$ . As before  $P = 0, 1$  is a permutation index. The set  $\tau(n, n)$  has as members the members of the set  $\sigma(n - 1, n - 1)$  and, in addition, the integer  $n$ . As for the case of the  $\sigma$ 's,  $\tau(n - j, n)$  has the same elements as  $\tau(n - j + 1, n)$  except that one of the integers with value  $(n - j)$  is replaced by the integer

$(n - j + 1)$ . For example, for  $n = 6$ , we have

$$\begin{aligned}
 \tau(6,6) &= \{1, 1, 1, 1, 2, 2, 3, 4, 5\}, \\
 \tau(5,6) &= \{1, 1, 1, 1, 2, 2, 3, 4, 6\}, \\
 \tau(4,6) &= \{1, 1, 1, 1, 2, 2, 3, 5, 6\}, \\
 \tau(3,6) &= \{1, 1, 1, 1, 2, 2, 4, 5, 6\}, \\
 \tau(2,6) &= \{1, 1, 1, 1, 2, 3, 4, 5, 6\}, \\
 \tau(1,6) &= \{1, 1, 1, 2, 2, 3, 4, 5, 6\}.
 \end{aligned}
 \tag{2.21}$$

The permutation index,  $P$ , for each term is determined by a set of rules only slightly more complicated than those used in the previous case ( $m = n - 1$ ). The additional complication arises from the fact that our solution (2.20) is expressed in terms of three non-zero fields rather than one as in (2.16). To determine the sign of a given term in (2.20) we note the sequence of subscripts associated with the operator  $S$  in that term. Using rules (i) and (ii) described following eq. (2.17) (with one minor change stated below) we alter this sequence until we reach a modified standard sequence (MSS). In the case  $m = n - 2$  there are three MSS's corresponding to the three non-zero components of  $A$ . To find the MSS corresponding to a given  $A_j$  we first find an even permutation of the sequence  $(n - 2, n - 1, n)$  such that  $j$  is in the last position. The MSS corresponding to  $A_j$  is then the standard sequence of length  $2^{n-3}$ , but with the last digit replaced by the integer immediately to the left of  $j$  in our evenly permuted sequence. So, for example, for  $n = 6$  the MSS corresponding to  $A_4$  is 12131216 whereas the MSS corresponding to  $A_5$  is just the standard sequence.

For a term in (2.20) involving  $A_j$ , we determine  $P$  by counting the number of legal operations necessary to reach this MSS. We must make one change in the rules following (2.17) which is the following addition to rule (ii): When  $k = n - 3$  the sequence must be replaced by the relevant MSS (rather than by the standard sequence). Moreover, it may be so replaced only if it differs from the MSS only in its last digit. If, using rules (i) and (ii) modified with this addition, we have performed an even (odd) number of operations to arrive at the MSS then  $P = 0$  (1) for that term.

It is now simple to state the general solution for arbitrary  $m < n$ . It is

$$l_i = \frac{1}{(n - m - 1)!} 2^{-(m-1)(m-2)/2} \sum_{\text{perms}} (-1)^P [SA]_{\nu(i, n, n-m)}. \tag{2.22}$$

$\nu(i, n, n - m)$  is a set of  $2^{m-1} + n - m - 1$  integers.  $\sigma(i, n) = \nu(i, n, 1)$  and  $\tau(i, n) = \nu(i, n, 2)$ .  $\nu(n, n, n - m)$  has as members the integers in the standard sequence of

length  $2^{m-1}$  plus the integers  $m + 1, m + 2, \dots, n - 1$ . To obtain  $\nu(n - j, n, n - m)$  from  $\nu(n - j + 1, n, n - m)$  one integer with the value  $n - j$  in  $\nu(n - j + 1, n, n - m)$  is replaced by  $n - j + 1$ .

$A$  is a “tensor” field with  $n - m - 1$  indices and is antisymmetric under the interchange of any two indices.  $A$  is non-zero only if each of its indices,  $\mu$ , has a value  $n \geq \mu \geq n - m + 1$ . To determine  $P$  for a given term in (2.22) we first note the set of integers which are the subscripts of  $A$  for that term. Then we find an even permutation of the sequence  $m, m + 1, \dots, n$  such that the  $n - m - 1$  integers comprising the subscripts of the  $A$  occupy the right most  $n - m - 1$  positions and are in increasing order from left to right. The integer immediately to the left of these  $n - m - 1$  indices (i.e. the one in the second position from the left of the entire sequence of  $n - m + 1$  integers) should be used to replace the last digit in the standard sequence of length  $2^{m-1}$ . This is the MSS for this term.  $P$  is determined by counting the number of operations necessary to take the indices of  $S$  into the MSS (as described for eq. (2.20) and adding to it the number of interchanges necessary to render the subscripts on  $A$  in ascending order from left to right. If the sum of the operations on the subscripts of  $S$  and permutations on the subscripts of  $A$  is even (odd) then  $P = 0$  (1).

#### 2.4. GAUGE INVARIANCE AND INDEPENDENT DEGREES OF FREEDOM

The solutions (2.22) can now be inserted into (2.6), the delta functions dropped, and the sum over the  $l_i$ 's replaced by a sum over the fields,  $A$ . However, solutions (2.22) contain, in general, a gauge symmetry which is a generalization of (2.12). In order to faithfully represent (2.6) in terms of the fields  $A$ , one should choose a gauge before evaluating  $Z$ . Since we are dealing with abelian theories, failure to choose a gauge will simply result in spurious uniform factors of infinity which will cancel in the computation of correlation functions (either of the original variables, or of gauge-invariant combinations of the dual variables). Nevertheless, it is interesting to expose the gauge symmetry of (2.22). Indeed, choosing a gauge may sometimes lead to a simpler or more useful representation of (2.6) than its gauge-invariant form.

The solutions (2.22) are invariant under the transformation

$$A_{\mu_1 \dots \mu_{n-m-1}} \rightarrow A_{\mu_1 \dots \mu_{n-m-1}} + C \sum_{\text{perms}} (-1)^P [S\Lambda]_{\gamma(m; \mu_1, \dots, \mu_{n-m-1})},$$

$$C = \frac{1}{(n - m - 2)!} + 2^{-(m-1)(m-2)/2}, \tag{2.23}$$

where  $\Lambda_{\alpha_1 \dots \alpha_{n-m-2}}$  is an arbitrary gauge function, antisymmetric under the interchange of any two indices, and non-zero only if each  $\alpha_k$  has a value  $n \geq \alpha_k \geq n - m + 1$  ( $k = 1, \dots, n - m - 2$ ). The set  $\gamma(m; \mu_1, \dots, \mu_{n-m-1})$  has as elements the integers of the standard sequence of length  $2^{m-1}$  except for the final integer (i.e., the integer  $m$ )

plus the integers  $\mu_1 \cdots \mu_{n-m-1}$ . The sum is over all permutations of the integers of  $\gamma$  and  $P$  changes from zero to one or one to zero under an interchange of any pair of  $\mu$ 's as well as under an "interchange in a box" of the  $2^{m-1}$  subscripts of  $S$  [see rule (i) following eq. (2.17)]. Note that the antisymmetry of the  $A$ 's is preserved and that any of the  $A$ 's with an index not in the range  $[n-m+1, n]$  remains zero even after a gauge transformation.

After having chosen a gauge we expect that the number of independent degrees of freedom of the theory, when expressed in terms of the dual variables, is the same as when expressed in terms of the original variables. Consider, for example, the case  $n=6$  and  $m=3$ . We then expect to have  $6-3=3$  independent dual functions. According to our solution (2.22) the  $l_i$ 's can be expressed as functions of the antisymmetric  $A_{\alpha\beta}$  with  $\alpha, \beta = 6, 5, 4, 3$ . However, we have four gauge functions at our disposal for gauge fixing. The number of independent degrees of freedom therefore appears to be  $4 \times 3/2 - 4 = 2$ . But this is incorrect. The point is that there are effectively only three independent gauge functions, not four. To see this note that the gauge transformation associated with a set of gauge functions  $\{\Lambda_\delta\}$  is the same as the gauge transformation associated with the set  $\{\Lambda_\delta + S_{121\delta}\Omega\}$  where  $\Omega$  is an arbitrary function. Therefore, there are really only three independent gauge degrees of freedom so that the number of independent degrees of freedom of the theory is  $4 \times 3/2 - 3 = 3$ , as it should be. The generalization of this observation to other values of  $m$  and  $n$  is straightforward.

2.5. MATHEMATICAL DIGRESSION: SUFFICIENT CONDITIONS FOR THE SOLUTION (2.22)

Having formally stated a solution (2.22) to (2.7) we now ask for a set of sufficient conditions which will assure us that this solution is sensible. To address this question we rewrite (2.7) in an inhomogenous form. Write

$$\sum_{i=m+1}^n R_i^j l_i \equiv -T^j, \tag{2.24}$$

so that

$$\sum_{i=1}^m R_i^j l_i = T^j, \quad j = 1, \dots, m. \tag{2.25}$$

For the moment let us regard the  $T^j$  as given. Then (2.25) is a set of  $m$  equations in  $m$  unknowns. For a given set of  $T^j$  it is a straightforward matter to show that (2.25) will have a solution if  $S_1^{-1}, S_{12}^{-1}, S_{1213}^{-1}, \dots, S_e^{-1}$  all exist, where  $e$  is the standard sequence of length  $2^m$ .

Now let us suppose that we can write the  $l_i$  for  $i = m+1, \dots, n$  in the form dictated by the solution (2.22). If we imagine having chosen a gauge, this just involves writing  $n-m$  functions in terms of  $n-m$  other functions. Suppose we choose the gauge

which leaves non-zero only those  $n - m$   $A$ 's whose subscripts have the largest totals. For instance for  $n = 6, m = 3$  the non-zero  $A$ 's in this gauge would be  $A_{65}, A_{64}$  and  $A_{54}$  while  $A_{63}, A_{53}$  and  $A_{43}$  would all be zero. It is then clear that if  $S_e^{-1}$  exists we can invert (at least in this gauge) our presumed solution for  $l_i, i = m + 1, \dots, n$  to find the required non-zero  $A$ 's in terms of  $l_{m+1}, \dots, l_n$ . Hence, if the  $S_e^{-1}$  exist, we can express  $l_{m+1}, \dots, l_n$  in the form of (2.22). Finally, we can use (2.25) with the sufficient conditions listed thereafter to write the remaining  $l_i$ 's in the form (2.22). Note that this sufficient condition is actually quite flexible since we may label the operators  $\{R_i^j\}$  in any convenient order.

### 3. Geometric interpretation of the dual form

The dual form of the usual abelian gauge and spin models heretofore studied had a straightforward interpretation, particularly when the theory was defined on a simple hypercubic lattice. However, the geometric interpretation in our more general case is not so straightforward. This is because the nature of the appropriate dual lattice on which the dual variables reside is not obvious. For example, in a theory with nearest neighbor and next nearest neighbor interactions the dual lattice should be such as to have a lattice element (e.g. on a cubic lattice a site, link, plaquette or cube) associated with each interaction term. In general dimensions the connection of such a lattice with the original real space lattice is not always simple or elegant. Thus each particular case must be addressed separately.

Nevertheless, a simple example can help to illustrate many of the general features which one might expect in a concrete, geometric realization of the dual form. For this purpose consider the trivial one-dimensional gaussian model with nearest neighbor (nn) and next nearest neighbor (nnn) interactions defined by

$$Z = \int D\phi \exp \left[ \sum_i -\alpha (R_1\phi_i)^2 - \beta (R_2\phi_i)^2 \right],$$

where

$$R_1\phi_i \equiv \phi_{i+1} - \phi_i, \quad R_2\phi_i \equiv \phi_{i+2} - \phi_i. \quad (3.1)$$

The model is illustrated in fig. 2. Introducing Fourier variables  $l_1(i)$  and  $l_2(i)$  (conjugate to  $R_1\phi_i$  and  $R_2\phi_i$ , respectively) and proceeding in the usual way (dropping overall inessential factors) we have

$$Z \propto \int Dl_1 Dl_2 \exp \left[ \sum_i -\frac{1}{4\alpha} l_1^2(i) - \frac{1}{4\beta} l_2^2(i) \right] \prod_j \delta(R_1 l_1(j) + R_2 l_2(j)). \quad (3.2)$$

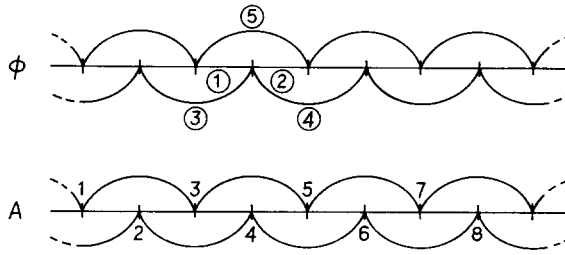


Fig. 2. One-dimensional model with nearest neighbor and next nearest neighbor interactions ( $\phi$ ) and its dual ( $A$ ) discussed in sect. 3.

The solution to the delta function constraints is [see eq. (2.9)]

$$l_1 = R_2 A, \quad l_2 = -R_1 A, \quad (3.3)$$

where  $A$  is a continuous scalar field,  $-\infty < A < \infty$ . Using (3.3) in (3.2) we easily see that this theory is self-dual (albeit trivial) so that

$$Z(\alpha, \beta) \propto Z\left(\frac{1}{4\beta}, \frac{1}{4\alpha}\right). \quad (3.4)$$

Since the theory is self-dual we expect that the dual theory can also be interpreted as a one-dimensional gaussian model with nn and nnn interactions. Using (3.3) we can try to construct such a geometric structure. In fig. 2 we show the original lattice ( $\phi$ ) and a dual lattice ( $A$ ). From (3.3) we note that  $l_1$ 's, those associated with nearest neighbor interactions on the  $\phi$  lattice are to be represented as next nearest neighbor differences of the  $A$  field on the dual lattice, and conversely for the  $l_2$ 's. We therefore need a prescription for associating geometrically an  $l$  corresponding to a given interaction link of the  $\phi$  lattice with an appropriate difference of  $A$  fields. Our prescription will be as follows: Consider a nearest neighbor link on the lattice  $\phi$ . Find the site on the lattice  $A$  which corresponds to the site on the  $\phi$  lattice which is at the left end of the  $\phi$  link in question. Call this dual site  $L$ . The  $l_1$  associated with this  $\phi$ -link is then  $A_{L+2} - A_L$ . Now consider a next nearest neighbor link on the  $\phi$  lattice. Find the site on the dual lattice ( $A$ ) which corresponds to the site on the  $\phi$  lattice which is at the right side of the nnn link in question. Call this dual ( $A$ ) site  $R$ . Then, the  $l_2$  associated with nnn  $\phi$  interaction is  $A_{R-1} - A_R$ .

In fig. 2 we have labelled 5 interaction lines in the  $\phi$  lattice with circled numbers and have labelled the sites of the  $A$  lattice. Following the above prescriptions the  $l$ 's associated with the labelled links are given by

$$\begin{aligned} l_{\textcircled{1}} &= A_5 - A_3, & l_{\textcircled{2}} &= A_6 - A_4, & l_{\textcircled{3}} &= A_3 - A_4, \\ l_{\textcircled{4}} &= A_5 - A_6, & l_{\textcircled{5}} &= A_4 - A_5. \end{aligned} \quad (3.5)$$

Now, if we imagine that the  $l$ 's represent currents, then the delta functions in (3.2) can be interpreted as enforcing a kind of divergenceless condition at each site of the  $\phi$  lattice. For example, in fig. 2 the delta function enforces the constraint that

$$l_{\textcircled{2}} - l_{\textcircled{1}} + l_{\textcircled{4}} - l_{\textcircled{3}} = 0. \quad (3.6)$$

It is interesting to see how (3.6) is satisfied in terms of the  $A$ 's. Using (3.5) we find that the divergenceless condition (3.6) corresponds to taking differences of  $A_i$ 's on the dual lattice around a closed path defined by the circuit (the numbers refer to site labels)  $5 \rightarrow 3 \rightarrow 4 \rightarrow 6 \rightarrow 5$ . That is, geometrically a condition of divergenceless (of the  $l$ 's) on the original lattice is expressed by a curlless property of the (single valued)  $A$ 's on the dual lattice. Such a diverge  $\leftrightarrow$  curl correspondence is not surprising in view of the nature of dual transformations. In other cases a similar divergence  $\leftrightarrow$  curl relationship will hold although it may be, geometrically, considerably more complex depending on the specific theory.

Before leaving this example we note that there is another type of closed circuit on the dual lattice about which we could have taken differences of  $A$ 's. Consider for instance the path  $5 \rightarrow 3 \rightarrow 4 \rightarrow 5$ . Using (3.5) this corresponds to the condition  $l_{\textcircled{1}} + l_{\textcircled{3}} + l_{\textcircled{5}} = 0$ . Pursuing our current analogy, this tells us that not only is the  $l$  current divergenceless at the sites of the  $\phi$  lattice but there is also no net  $l$  current flowing through the lattice. This result is related to the choice of boundary conditions, about which we have been cavalier in (3.1)–(3.3). It is clear that we can add in (3.3) arbitrary lattice independent constants to the expressions for  $l_1$  and  $l_2$  and still satisfy the delta functions in (3.2). Eq. (3.3) corresponds to boundary conditions in which  $l_1$  and  $l_2$  are set equal to zero at, say, the left end of the lattice. This can also be accomplished by a suitable choice of boundary conditions on  $\phi$  in the original lattice. This is an example of the point discussed in sect. 2 concerning the importance of fixing boundary conditions for defining the admissible solutions of (2.7). In addition to altering somewhat the solutions (3.3), another choice of boundary conditions might have resulted in a different net  $l$ -current flowing through the lattice, but, of course, this current would still be conserved.

#### 4. Summary and comments

In this paper we have derived a dual form for a very wide class of abelian theories which includes theories with more general types of interactions (such as nearest neighbor plus next nearest neighbor) as well as the more usual kinds of statistical and field theoretic models. Under the duality transformation theories are mapped into dual theories in such a way that the high (low) temperature regime of the original form of the theory maps into the low (high) temperature regime of the dual form of the theory. The dual form of the hamiltonian is determined by the number of different kinds of terms in the original hamiltonian and the number of different

independent fields. In particular, the form of the dual theory does not directly depend on the space or space-time dimensionality of the theory. Furthermore, the dual forms we have presented in general have a gauge symmetry as discussed in subsect. 2.4.

Because the form of the dual theory depends on finding a solution to a set of linear homogeneous equations [see (2.7)], there are actually a variety of functional forms possible for the dual theory. Our solution was chosen to satisfy the criterion that in the dual hamiltonian there be no inverses of the (differential) operators that appear in the original form of the hamiltonian. However, for certain problems it may be advantageous to choose a different solution to eqs. (2.7). Even in that case, though, the dual form of the theory will retain the usual property of relating high (low) temperature regions of the original theory to low (high) temperature regions of its dual.

When duality transformations have been used to discuss ordinary spin and gauge theories, it has usually been a simple matter to use the dual theory to obtain the form and interactions of the topological excitations of the original degrees of freedom of the model. This was generally accomplished by using the Poisson summation formula to rewrite the dual theory in a somewhat different form. For the members of the very general class of theories which we treat here one can rewrite the dual theory using the Poisson summation trick, but in general it is not clear that the fields that thereby appear can be interpreted as topological excitations of the original theory. This inability to easily identify the topological excitations stems largely from the problem of geometrically interpreting the dual hamiltonian, as discussed in sect. 3. Indeed, even in the case of a relatively simple theory, such as that of ref. [3], it is not obvious how to do this. A better understanding of this problem would certainly be helpful.

The techniques and results presented in this paper should be a useful tool for studying many theories of statistical and field theoretic interest. As with any tool, the more skillful the user the better will be the result.

The work in this paper was motivated in large part by a very pleasant and fruitful collaboration on a related subject with D. Amit, S. Elitzur, and E. Rabinovici. I thank them for their stimulating and helpful remarks. I also thank them and the Hebrew University for the warm hospitality extended to me while part of this work was being done. This work was supported by the United States-Israel Binational Science Foundation and by the National Science Foundation under grant no. PH78-08426.

## References

- [1] R. Savit, *Rev. Mod. Phys.* 52 (1980) 453, and references therein
- [2] K. Wilson, *Phys. Rev. D* 10 (1974) 2445
- [3] D. Amit, S. Elitzur, E. Rabinovici and R. Savit, in preparation
- [4] V. Berezinskii, *ZhETF (USSR)* 59 (1970) 907; 61 (1972) 1144 [*JETP (Sov. Phys.)* 32 (1971) 493; 34 (1972) 610];  
J. Villain, *J. Phys. C* 36 (1975) 581