MOSCOW MATHEMATICAL JOURNAL
Volume 10, Number 4, October–December 2010, Pages 713–727

COMBINATORIAL SPECIES AND CLUSTER EXPANSIONS

WILLIAM G. FARIS

Dedicated to the memory of Roland L. Dobrushin

Abstract. This paper will survey recent progress on clarifying the connection between enumerative combinatorics and cluster expansions. The combinatorics side concerns species of combinatorial structures and the associated exponential generating functions. Cluster expansions, on the other hand, are supposed to give convergent expressions for measures on infinite dimensional spaces, such as those that occur in statistical mechanics. There is a dictionary between these two subjects that sheds light on each of them. In particular, it gives insight into convergence results for cluster expansions, including a well-known result of Roland Dobrushin. Furthermore, the species framework provides a context for recent results of Fernández–Procacci and of the author.

2000 Math. Subj. Class. Primary: 60K35, 82B05, 05A15; Secondary: 82B20, 05C30.

Key words and phrases. Combinatorial species, species of structures, exponential generating function, equilibrium statistical mechanics, grand partition function, cluster expansion.

1. Introduction

There is a something akin to an isomorphism between a framework used in combinatorics (species of structures and their associated exponential generating functions) and a framework used in equilibrium statistical mechanics (grand partition functions, etc.). To the scholars in each community one of the frameworks looks obvious and familiar, and the other one looks alien and bizarre. However they are rather close to being the same. The purpose of this paper is to explore the connection.

The plan of this paper is to begin with the combinatorical side. There are various operations on species that translate to operations on exponential generating functions. These operations may be used to formulate equations for species. These translate to exponential generating function equations.

Received January 7, 2010; in revised form February 23, 2010.
Paper presented at Dobrushin Conference, Institute for Problems of Information Transmission, Moscow, July 15–20, 2009.

Two examples that are fundamental in physics are the species of rooted trees and the species of rooted connected graphs. We shall see that the species of rooted trees satisfies an equation that only uses nice operations. However the corresponding equation for rooted connected graphs satisfies an equation that uses a rather unpleasant operation. This explains in a fundamental and precise way why dealing with rooted connected graphs is more difficult. Furthermore, it motivates the attempt to related solutions of the rooted connected graph equation with the simpler solutions of the rooted tree equation. These results are species formulations of ideas that belong to the statistical physics literature.

The next part of the paper deals with the convergence of cluster expansions. In the simplest context, this is the theory of an equilibrium gas system. The density of the gas is given by the solution of the equation for rooted connected graphs. There is a classical result of Kotecký–Preiss that reduces convergence for solutions of this equation to those for the rooted tree equation. This connected is made explicit by noting that the Kotecký–Preiss condition is equivalent to finiteness of the solution of the tree equation.

The paper continues with a brief description of recent results of Fernández–Procacci and of the author. These results are stronger, since they reduce convergence of the rooted connected graph equation to an enriched rooted tree equation with better convergence properties. In particular, the Fernández–Procacci type result may be extended to soft interactions and to the continuum. These more recent results are not described in detail, but they are in the spirit of the species analysis. More information may be found in the references.

2. Enumerative Combinatorics

2.1. Colored sets. The combinatorial structures under consideration are built over colored sets. These are defined as follows. There is a set $\mathcal{P}$ that serves as a fixed palette of colors. A *colored set* is a function $a\colon U \to \mathcal{P}$, where U is a finite set. If j is a point in U, then $a(j)$ is the color of j. The colored sets form the objects of a category $\mathbf{B}$. The morphisms in this category are bijections of the underlying sets that preserve the colors.

In combinatorics the underlying set U is a set of labels, and the coloring is an additional structure that is imposed on a label set. This structure also occurs in physics. In this case $\mathcal{P}$ is a fixed set of locations. A set U is a set of particles, and a function $a\colon U \to \mathcal{P}$ is a particle configuration, that is, an assignment of particles to locations.

2.2. Weighted sets. We need another category $\mathbf{E}$ with objects that consist of weighted sets. There is a fixed commutative ring R; for instance this could be the real numbers or the complex numbers. A *weighted set* is a finite set A together with a weight function $\mathrm{wt}\colon A \to R$. A morphism of weighted sets is a bijection that preserves the weights. The basic requirement on the category is that the weight function behaves well on disjoint unions and on cartesian products. The weight function on a disjoint union must agree with the weights on the individual parts. The weight function on a cartesian product must assign to each ordered tuple the product of the weights of the components.

If A is a weighted set, then the total weight of A is the sum of the weights of the points in A. The total weight of a sum (disjoint union) is obviously the sum of the total weights of the parts. The total weight of a product (cartesian product) is the product of the total weights of the factors.

There are many examples of weighted sets, but the one of most use in the following is a set of graphs over a colored set. There is a given function t that assigns to each ordered pair of colors p, q in $\mathcal{P}$ an element $t(p, q)$ in the ring R. It is required to be symmetric. Consider a colored set $a\colon U \to \mathcal{P}$ with underlying set U. We think of U as a vertex set. Then a graph g with vertex set U is identified with a set of two-element subsets $\{i, j\}$ of U. These are the edges of the graph g. The weight of a graph g is

$$\mathrm{wt}(g) = \prod_{\{i,j\}\in g} t(a(i), a(j)). \tag{1}$$

In the physics application the palette of colors $\mathcal{P}$ represents a set of locations. The colored set $a\colon U \to \mathcal{P}$ represents a particle configuration. A graph g is a collection of two-element sets of particles that are regarded as interacting. The interaction between two particles i, j depends on their locations $a(i)$, $a(j)$ and is given by $t(a(i), a(j))$. The interaction for the entire collection of pairs is the product of these individual pair interactions, that is, it is the weight of the graph. The total weight of a set of graphs is the sum of the weighs of the individual graphs.

2.3. Combinatorial species. The colored set and weighted set concepts lead to a framework for constructing combinatorial structures in a systematic way. Each instance of such a construction is called a "species" of structures. The theory is explained in detail in the book of Bergeron, Labelle, and Leroux [1]; here we can only give an outline.

A combinatorial *species* is a functor F from the category $\mathbf{B}$ of colored sets to the appropriate category $\mathbf{E}$ of weighted sets. Thus for every colored set $a\colon U \to \mathcal{P}$ there is a corresponding weighted set $F[a]$. There are several species of interest to us. Here are some basic ones:

The graph species G. The species G of *graphs* associates to each colored set $a\colon U \to \mathcal{P}$ the set $G[a]$ of all $2^{\binom{n}{2}}$ graphs with vertex set U. The weight of a graph is as given above.

The connected graph species C. Even more important is the species C of *connected graphs*. A graph is connected if it has a non-empty vertex set that cannot be partitioned into two or more parts that are not connected by edges. A connected graph on an n element vertex set has at least $n - 1$ edges. The number of connected graphs in $C[a]$ is somewhat smaller but almost as large as the number of graphs in $G[a]$.

The tree species T. A final example in this series is the species T of *trees*. A graph is a tree if it is a minimal connected graph. A tree on an n element vertex set has $n - 1$ edges. The number of trees in $T[a]$ is only n^{n-2}.

The set indicator species E. Sometimes it is useful to have a monochromatic species, one that is a functor from the category of finite sets (without colors) to an appropriate category of weighted sets. One such species is the set indicator

(ensemble) species E that assigns to each finite set U a single point $\{U\}$ with weight one. While this does not seem to be a very interesting species in itself, it occurs as a building block in various important constructions.

2.4. Operations on species. There are various important operations on species. Here is a list of some of the most important ones. They are chosen so that they lead to relatively simple computations, except for the last one.

Distinguished point. If F is a species, then for each color p there is another species $F^\bullet_p$. Then $F^\bullet_p[a]$ for $a\colon U \to \mathcal{P}$ consists of all ordered pairs consisting of a point in U of color p and an element of $F[a]$. Thus this species incorporates a *distinguished point* of color p. As examples we have $G^\bullet_p$ and $C^\bullet_p$ and $T^\bullet_p$. These are the species of rooted graphs, rooted connected graphs, and rooted trees.

Combinatorial convolution. Another important operation is the *combinatorial convolution*. The convolution of two species F, G is defined as a disjoint union of cartesian products by

$$(F * G)[a] = \sum_{\langle V,W\rangle} F[a_V] \times G[a_W]. \tag{2}$$

Here $\langle V, W\rangle$ denotes an ordered pair of complementary sets with union U. A notation such as a_V denotes the restriction $a\colon U \to \mathcal{P}$ to the subset V. In other words, one splits the underlying set in all possible ways. This gives a disjoint union of a cartesian product corresponding to the two subsets in the splitting.

Here is an example. We have

$$G^\bullet_p = C^\bullet_p * G. \tag{3}$$

This says that a graph together with a distinguished point of color p corresponds to a connected graph with a distinguished point of color p on a subset together with a graph on the complement.

Scalar combinatorial composition. If F is a monochromatic species, and if H is a species with $H[\varnothing] = \varnothing$, then $F \circ H$ is a new species. The value of this species on $a\colon U \to \mathcal{P}$ is given as a disjoint union of cartesian products by

$$(F \circ H)[a] = \sum_{\Gamma} F[\Gamma] \times \prod_{V\in\Gamma} F[a_V]. \tag{4}$$

Here Γ ranges over partitions of U into disjoint non-empty sets V. The colored set $a_V : V \to \mathcal{P}$ is given by restriction.

The species C of connected graphs satisfies $C[\varnothing] = \varnothing$, since the set of connected graphs on an empty vertex set is empty. This allows the formulation of one of the most famous examples of the combinatorial exponential:

$$G = E \circ C. \tag{5}$$

This says that for every graph on U there is a partition Γ of U with a connected graph on each set V in the partition. The set indicator species E inserts a trivial one-point factor in the cartesian products.

Combinatorial composition. This operation has a generalization to a more general *combinatorial composition*. If F is a species, and if H_p is a species with $H_p[\varnothing] = \varnothing$ for each p in $\mathcal{P}$, then $F \circ H$ is a new species. The value of this species on $a\colon U \to \mathcal{P}$ is given by a disjoint union of cartesian products by

$$(F \circ H)[a] = \sum_{\Gamma} \sum_{c\colon \Gamma \to \mathcal{P}} F[c] \times \prod_{V \in \Gamma} F_{c(V)}[a_V]. \tag{6}$$

Here Γ ranges over partitions of U into disjoint non-empty sets V, and c ranges over colorings of Γ. The colored set $a_V : V \to \mathcal{P}$ is given by restriction.

Cartesian product. The preceding operators are nice, in a sense that will be explained in the next subsection. There is another operation that is important but somewhat unpleasant. The *cartesian product* of two species F, G is defined by

$$(F \times G)[a] = F[a] \times G[a]. \tag{7}$$

2.5. Exponential generating functions. A key idea in the theory of species is the *exponential generating function*. A species is a functor F from the category of colored sets to a category of weighted sets. If a is a colored set, then we write the corresponding weighted set of combinatorial structures as $F[a]$. The sum of weights of $F[a]$ is written $f(a)$. The exponential generating function is a function of many variables, one for each possible color. Thus we use a variable w_p for each color $p \in \mathcal{P}$. For each n let U_n be a set with n points. The exponential generating function is written

$$F(w) = \sum_{n=0}^{\infty} \sum_{a\colon U_n \to \mathcal{P}} \frac{1}{n!} f(a) \prod_{i \in U_n} w_{a(i)}. \tag{8}$$

There is an alternative expression for such an exponential generating function using multi-indices. A multi-index is a function N defined on $\mathcal{P}$ with natural number values. Each colored set a induces a corresponding multi-index that counts how many times each color occurs. The sum of weights $f(a)$ only depends on N, and so it may be written $f(N)$. Define $N! = \prod_p N(p)!$ and define $w^N = \prod_p w_p^{N(p)}$. With these notations the exponential generating function becomes

$$F(w) = \sum_N \frac{1}{N!} f(N) w^N. \tag{9}$$

For each combinatorial operation there is a corresponding operation on exponential generating functions.

Euler derivative. The operation of choosing a distinguished point has a simple expression in terms of exponential generating functions. It is given by a differential operator according to

$$F_p^{\bullet}(w) = w_p \frac{\partial}{\partial w_p} F(w). \tag{10}$$

The effect on the coefficients is to replace $f(N)$ by $N(p)f(N)$.

Product. The combinatorial convolution is easy; in this case it is the product

$$(F * G)(w) = F(w)G(w). \tag{11}$$

Scalar composition. The operation of taking the scalar combinatorial composition is also simple. In this case the species F is monochromatic, so its exponential generating function is a function of a single variable. The composition is

$$(F \circ H)(w) = F(H(w)). \tag{12}$$

As an example, note that the set indicator species has exponential generating function $E(w) = \exp(w)$. Since the graph and connected graph species are related by $G = E \circ C$, we have $G(w) = \exp(C(w))$. Furthermore, $G_p^\bullet(w) = w_p(\partial/\partial w_p)G(w) = C_p^\bullet(w)G(w)$, as one would expect from the combinatorial convolution.

Composition. The combinatorial composition of F with the family of species H_p leads to the composition

$$(F \circ H)(w) = F(H(w)). \tag{13}$$

Pointwise product of coefficients. The cartesian product $F \times G$ has an exponential generating function with coefficients $f(a)g(a)$.

2.6. Set indicator species. In graph constructions it is useful to have species that serve as building blocks. The simplest of these are indicator species. These are species constructions that produce as output either the empty set of structures or a set with only one element. Here are three basic examples that indicate inputs with one-point sets.

One-point set of color p indicator species X_p. The *one-point set of color p indicator* species X_p on the colored set $a\colon U \to \mathcal{P}$ has the value $\varnothing$ unless U consists of a one-point set $U = \{j\}$ with $a(j) = p$. In that case it has value a single point $\{U\}$ with weight one. It has exponential generating function $X_p(w) = w_q$.

One-point set indicator species X. The *one-point set indicator* species X on the colored set $a\colon U \to \mathcal{P}$ has the value $\varnothing$ unless U consists of a one-point set. In that case it has value a single point $\{U\}$ with weight one. It has exponential generating function $X(w) = \sum_q w_q$.

Edge to one-point set indicator spacies X^p. The *edge to one-point set indicator* species X^p on the colored set $a\colon U \to \mathcal{P}$ has the value $\varnothing$ unless U consists of a one-point set $U = \{j\}$. In that case it has value a single point $\{U\}$ with weight $t(p, a(j))$. It represents a contribution of an edge from a single external point of color p to the internal point j. It has exponential generating function $X^p(w) = \sum_q t(p, q)w_q$.

Here is a list of indicator species that indicate more general set inputs. They are used in the power set constructions of Section 2.7.

Set indicator species $\hat{E}$. The *set indicator* species $\hat{E} = E \circ X$ on the colored set $a\colon U \to \mathcal{P}$ has the value a single point with weight one. It has exponential generating function given by the scalar composition $\hat{E}(w) = \exp(\sum_q w_q)$.

Edges to set indicator species $\hat{E}^p$. The *edges to set indicator* species $\hat{E}^p = E \circ X^p$ on the colored set $a\colon U \to \mathcal{P}$ has the value a single point with weight $\prod_{j \in U} t(p, a(j))$. It represents a contribution of edges from a single

external point of color p to all internal points in U of whatever color. It has exponential generating function given by the scalar composition $\hat{E}^p(w) = \exp(\sum_q t(p, q)w_q)$.

Edges to non-empty set indicator species $\hat{E}^p_+$. The *edges to non-empty set indicator* species $\hat{E}^p_+$ on the colored set $a\colon U \to \mathcal{P}$ has the value $\varnothing$ if $U = \varnothing$. Otherwise it returns a single point with weight $\prod_{j\in U} t(p, a(j))$. It represents a contribution of edges from a single external point of color p to all internal points in U of whatever color. It has exponential generating function given by $\hat{E}^p_+(w) = \exp(\sum_q t(p, q)w_q) - 1$.

2.7. Edges to subsets species. A subset U with n points has 2^n subsets and $2^n - 1$ non-empty subsets. If U is a set of vertices, then one can identify a subset with a collection of edges from an external point to the subset. This idea leads to the following species constructions. They are ingredients in the graph and connected graph equations.

Edges to subset species P^p. The *edges to subsets* species is the combinatorial convolution $P^p = \hat{E} * \hat{E}^p$. Its value on the colored set $a\colon U \to \mathcal{P}$ is the set of all subsets of U (the power set). The weight of a subset $W \subseteq U$ is $\prod_{j\in W} t(p, a(j))$. It represents a contribution of edges from a single external point of color p to all internal points in W of whatever color. The exponential generating function is the product $P^p(w) = \exp(\sum_q w_q)\exp(\sum_q t(p, q)w_q) = \exp(\sum_q (1 + t(p, q))w_q)$.

Edges to non-empty subsets species P^p_+. The *edges to non-empty subsets* species is the combinatorial convolution $P^p = \hat{E} * \hat{E}^p_+$. Its value on the colored set $a\colon U \to \mathcal{P}$ is the set of all non-empty subsets of U. The weight of a non-empty subset $W \subseteq U$ is $\prod_{j\in W} t(p, a(j))$. It represents a contribution of edges from a single external point of color p to all internal points in W of whatever color. The exponential generating function is the product $P^p_+(w) = \exp(\sum_q w_q)(\exp(\sum_q t(p, q)w_q) - 1) = \exp(\sum_q(1 + t(p, q))w_q) - \exp(\sum_q w_q)$.

3. Combinatorial Fixed Points

3.1. The rooted tree fixed point equation. The next topic is combinatorial fixed point equations. One case where this is straightforward is for rooted trees. Let X_p be the species that indicates one point sets of color p. In other words, it produces a single point for each such set, and the empty set otherwise. The rooted tree equation is actually a family of equations, one for each color p. Thus we should think of $T^\bullet_p$ as a family of species. The following two propositions are standard combinatorics.

Proposition 1. *The species fixed point equation for rooted trees is*

$$T^\bullet_p = X_p * (\hat{E}^p \circ T^\bullet). \tag{14}$$

This says that a rooted tree with root of color p consists of a single point of color p, together with a structure on the complement of this point. This structure consists of a partition of the complement into disjoint non-empty sets V. On each

set V in the partition there is a tree with a root of some color q. These trees get the additional weight $t(p, q)$ from the edges to set construction.

Proposition 2. *The fixed point equation for the exponential generating function for rooted trees is*

$$T_p^\bullet(w) = w_p \exp\Big(\sum_q t(p, q) T_q^\bullet(w)\Big). \tag{15}$$

Let us look at the rooted tree fixed point equation (15) in more detail. Take each vertex weight $w_p \geqslant 0$. Let the edge weights $0 \leqslant t(p, q)$ be positive. The rooted tree fixed point equation for $z_p = T_p^\bullet(w)$ is

$$z_p = w_p \exp\Big(\sum_q t(p, q) z_q\Big). \tag{16}$$

The *Kotecký–Preiss condition* is that there exists $0 \leqslant x_p < \infty$ such that

$$w_p \exp\Big(\sum_q t(p, q) x_q\Big) \leqslant x_p. \tag{17}$$

Proposition 3. *The rooted tree fixed point equation* (16) *has a least finite solution z if and only if the Kotecký–Preiss condition* (17) *holds. In that case $z_p \leqslant x_p$ for all p.*

This follows from the Knaster–Tarski fixed point theorem. The complete analysis of the rooted tree fixed point equation is due to the fact that the equation may be analyzed for fixed values of the w_p parameters.

3.2. Dobrushin trees. Dobrushin [2] introduced a condition for convergence of cluster expansions. It turns out that this condition is equivalent a certain enriched rooted tree fixed point equation. It seems reasonable to call a tree satisfying this condition a Dobrushin tree. Thus a *Dobrushin tree* is a rooted tree for which the fiber over each vertex satisfies the condition that no two points can have the same color. In order to formulate this condition, we need yet one more indicator species.

Edges to polychromatic set indicator species $\hat{E}_\star^p$. The *edges to polychromatic set indicator* species $\hat{E}_\star^p$ on the colored set $a\colon U \to \mathcal{P}$ with injective a has the value a single point with weight $\prod_{j \in U} t(p, a(j))$. Otherwise it has the value $\varnothing$. It represents a contribution of edges from a single external point of color p to internal points in U that are prescribed to have all different colors. It has exponential generating function given by $\hat{E}_\star^p(w) = \prod_q (1 + t(p, q) w_q)$.

Proposition 4. *The species fixed point equation for Dobrushin trees is*

$$D_p^\bullet = X_p * (\hat{E}_\star^p \circ D^\bullet). \tag{18}$$

This says that a Dobrushin tree with root of color p consists of a single point of color p, together with a structure on the complement of this point. This structure consists of a partition of the complement into disjoint non-empty sets V. On each set V in the partition there is a Dobrushin tree with a root of some color q. The colors are all different. The trees get the additional weight $t(p, q)$ from the edges to set construction.

Proposition 5. *The fixed point equation for the exponential generating function of Dobrushin trees is*

$$D_p^\bullet(w) = w_p \prod_q (1 + t(p, q) D_q^\bullet(w)). \tag{19}$$

Take each vertex weight $w_p \geqslant 0$. Let the edge weights $0 \leqslant t(p, q)$ be positive. The Dobrushin tree fixed point equation (19) written for $z_p = D_p^\bullet(w)$ is

$$z_p = w_p \prod_q (1 + t(p, q) z_q). \tag{20}$$

The *Dobrushin condition* is that there exists $0 \leqslant x_p < \infty$ such that

$$w_p \prod_q (1 + t(p, q) x_q) \leqslant x_p. \tag{21}$$

Proposition 6. *The Dobrushin tree fixed point equation* (20) *has a least finite solution z if and only if the Dobrushin condition* (21) *holds. In that case $z_p \leqslant x_p$ for all p.*

Dobrushin [2, Section 6] formulated his condition is a somewhat different way. Suppose that $t(p, p) = 1$ for each p. Let $1 + t(p, q) x_q = \exp(w_q b_q)$. Then the condition becomes

$$w_p \exp\Big(\sum_q w_q b_q\Big) \leqslant \exp(w_p b_p) - 1, \tag{22}$$

or

$$1 - w_p \exp\Big(\sum_{q \neq p} w_q b_q\Big) \geqslant \exp(-w_p b_p), \tag{23}$$

which is essentially Dobrushin's equation (6.1). In this form it might be difficult to recognize as related to a fixed point equation for rooted trees.

3.3. The rooted graph equation. This subsection presents a species equation for rooted graphs and also the corresponding exponential generating function equation. The result for the exponential generating function may be interpreted in the sense of formal power series. This material is elementary; it is preparation for the result for rooted connected graphs in the next subsection.

Proposition 7. *The species equations for rooted graphs is*

$$G_p^\bullet = X_p * (P^p \times G). \tag{24}$$

This says that every graph with a designated point of color p consists of a point of that color plus a structure on the complement. This structure consists of a graph together with subset of vertices that are onnected by edges to the original designated point.

The cartesian product is somewhat unpleasant. Let t_p be the vector with components $t(p, q)$ for q in $\mathcal{P}$. The exponential generating function for the edge to subsets species has the exponential generating function

$$P^p(w) = \exp(1 \cdot w) \exp(t_p \cdot w) = \exp((1 + t_p) \cdot w) = \sum_N \frac{1}{N!} (1 + t_p)^N w^N. \tag{25}$$

The exponential generating function for the cartesian product is given by pointwise multiplication of the coefficients, so it ultimately produces the scaling

$$(P^p \times G)(w) = \sum_N \frac{1}{N!}(1+t_p)^N g(N) w^N = G((1+t_p)w). \tag{26}$$

The following result is an elementary consequence of the corresponding species equation. Unfortunately, it involves an unpleasant scaling, due to the fact that the species equation involves cartesian product.

Proposition 8. *The equation for the exponential generating function for rooted graphs is*

$$G_p^\bullet(w) = w_p G((1+t_p)w). \tag{27}$$

This equation has an explicit solution. It leads to the recursion $g(N+\delta_p) = g(N)(1+t_p)^N$ for the coefficients of $G(w)$. This recursion has solution $g(N) = (1+t)^{\mathrm{Pair}(N)}$. Here $\mathrm{Pair}(N)$ is a multi-index for pairs $\{p, q\}$ of colors with the value $N(p)N(q)$ for $p \neq q$ and the value $\binom{N(p)}{2}$ for $p = q$. It represents the number of edges connecting vertices of these colors. The coefficients of $G_p^\bullet(w)$ are $N(p)g(N)$.

3.4. The rooted connected graph fixed point equation. The objects of central interest in the physics applications are connected graphs and rooted connected graphs. The equation for rooted connected graphs is more complicated. This subsection presents a species equation for rooted connected graphs and the corresponding exponential generating functions equation.

The exponential generating function is for the moment to be considered in the sense of formal power series. The relevant expansions are

$$C(w) = \sum_N \frac{1}{N!} c(N) w^N \tag{28}$$

for connected graphs and

$$C_p^\bullet(w) = \sum_N \frac{1}{N!} N(p) c(N) w^N \tag{29}$$

for rooted connected graphs. Here the multi-index N is a *cluster* and $c(N)$ is the corresponding *cluster coefficient*. Such a formula is called a *cluster expansion*. The convergence question is addressed in a later section.

Proposition 9. *The combinatorial equation for the species of rooted connected graphs is*

$$C_p^\bullet = X_p * (E \circ (P_+^p \times C)). \tag{30}$$

Recall that P_+^p is the *edges to non-empty subsets* species. This equation thus says that every connected graph with a designated point of color p consists of a point of that color plus a structure on the complement. The complement is partitioned into disjoint non-empty subsets V. On each such subset V there is a connected graph and a non-empty subset W of V. The edges to non-empty subsets species also incorporates the contribution of the edges from the original point to the points in W.

The exponential generating function for the edges to non-empty subsets species is the product

$$P_+^p(w) = \exp(1 \cdot w)(\exp(t_p \cdot w) - 1) = \sum_N \frac{1}{N!}((1+t_p)^N - 1)w^N. \tag{31}$$

The exponential generating function for the cartesian product is given by pointwise multiplication of the coefficients, so it ultimately produces the scaling difference

$$(P_+^p \times C)(w) = \sum_N \frac{1}{N!}((1+t_p)^N - 1)c(N)w^N = C((1+t_p)w) - C(w). \tag{32}$$

The following proposition is known in many special cases; however the general formulation may be less familiar. The nice feature is that it exhibits the relation to the rooted tree equation rather clearly.

Proposition 10. *The rooted connected graph exponential generating function equation is*

$$C_p^\bullet(w) = w_p \exp(C((1+t_p)w) - C(w)). \tag{33}$$

The scaling in the equation means that one has to consider the function $C^\bullet(w)$ for varying values of the parameters w_p. An integration converts this into an equation for a rooted connected graph fixed point. We use

$$\frac{d}{ds}C((1+st_p)w) = \sum_q t(p,\, q)w_q \frac{\partial}{\partial w_q}C((1+st_p)w) = \sum_q t(p,\, q)C_q^\bullet((1+st_p)w). \tag{34}$$

This gives the main result of this section, which is the following (functional) fixed point equation.

Proposition 11. *The fixed point equation for the exponential generating function for rooted connected graphs is*

$$C_p^\bullet(w) = w_p \exp\biggl(\sum_q t(p,\, q)\int_0^1 C_q^\bullet((1+st_p)w)\, ds\biggr). \tag{35}$$

4. Cluster Expansions

4.1. The equilibrium discrete particle gas model. In the application to a discrete particle gas the terminology is somewhat different. The color palette $\mathcal{P}$ is a fixed set of particle locations. The vertex set U is a finite set of particles. The colored set $a\colon U \to \mathcal{P}$ is a particle configuration. The color variable $w_p \geqslant 0$ in the exponential generating function is the weight for particles at $p \in \mathcal{P}$ (the *activity*). Finally, there is a quantity $0 \leqslant 1 + t(p,\, q) \leqslant 1$ which is the *Boltzmann* factor for pair of particles at locations $p,\, q \in \mathcal{P}$. Thus $-1 \leqslant t(p,\, q) \leqslant 0$ is a measure of the interaction between the pair of particles.

The *grand partition function*

$$G(w) = \sum_{n=0}^\infty \sum_{a\colon U_n \to \mathcal{P}} \frac{1}{n!} g(a) \prod_{i \in U_n} w_{a(i)} \tag{36}$$

is the exponential generating function for graphs. The coefficient

$$g(a) = \prod_{\{i,j\}} (1 + t(a(i),\, a(j)))$$

is the product of Boltzmann factors corresponding to all unordered pairs of distinct particles. When this is expanded, it give a sum of $2^{\binom{n}{2}}$ terms, each of which is a contribution from some graph. As before, define a multi-index $\mathrm{Pair}(N)$ for unordered pairs $\{p,\, q\}$ of locations. Thus $\mathrm{Pair}(N)$ is the number of pairs of distinct particles at the locations, which is $N(p)N(q)$ for distinct locations $p \neq q$ and the binomial coefficient $\binom{N(p)}{2}$ for the same location $p = q$. The coefficient may then be written as $g(N) = (1 + t)^{\mathrm{Pair}(N)}$. The grand partition function in the multi-index notation is

$$G(w) = \sum_N \frac{1}{N!} (1 + t)^{\mathrm{Pair}\, N} w^N. \tag{37}$$

A particularly convenient quantity for convergence results is the *density* (expected number of particles at a location). The density at p is

$$n(p) = \frac{1}{G(w)}\, w_p \frac{\partial}{\partial w_p} G(w) = \frac{1}{G(w)} G^\bullet_p(w) = C^\bullet_p(w). \tag{38}$$

Thus the density $n(p)$ at p of the gas, regarded as a function of the local activity variables w, is the exponential generating function $C^\bullet_p(w)$ for rooted connected graphs with root of color p. This fundamental relation is at the heart of statistical physics.

4.2. Fixed points and convergence in the gas model. One classic version of the cluster expansion theorem is the Kotecký–Preiss result. Their result was confined to the hard-core interaction case where the only interaction factors are $1 + t(p,\, q) = 0,\, 1$ and $1 + t(p,\, p) = 0$. However it is also true with a more general soft interaction with $0 \leqslant 1 + t(p,\, q) \leqslant 1$. This condition means that the edge weights satisfy $-1 \leqslant t(p,\, q) \leqslant 0$. The graph weights can have both positive and negative signs, leading to a remarkable amount of cancelation. It is this cancelation that is responsible for the convergence.

Write $|t_p| \cdot x$ for $\sum_q |t(p,\, q)| x_q$. The following Kotecký–Preiss type theorem is by now a rather standard result.

Theorem 1. *Suppose the activities and Boltzmann factors satisfy $w_p \geqslant 0$ and $0 \leqslant 1 + t(p,\, q) \leqslant 1$. If there are $x_p \geqslant 0$ such that for each p we have the rooted tree contribution inequality*

$$w_p \exp(|t_p| \cdot x) \leqslant x_p, \tag{39}$$

then the cluster expansion for rooted connected graphs (the density) $C_p(w)$ in powers of w converges absolutely.

The exponential on the left hand side is a partition function for a variable number of particles that only interact with a particle at location p. From a combinatorial point of view the inequality is related to a rooted tree fixed point equation. Thus the theorem may be understand as a comparison of the rooted tree fixed point equation (15) with the rooted connected graph fixed point equation (35).

The cluster expansion theorem presented here is an extension of a relatively recent result of Fernández and Procacci. Their result was also for the hard-core interaction case. See [4] for this work and for references to earlier work. The proof in [3] relies on an interpolation identity that relates rooted trees and rooted connected graphs. For the formulation introduce the notation $|t_p|x$ for the vector whose coordinates are $|t(p, q)|x_q$. The following theorem is the Fernández–Procacci result extended to the soft interaction case.

Theorem 2. *Suppose the activities and Boltzmann factors satisfy* $w_p \geqslant 0$ *and* $0 \leqslant 1 + t(p, q) \leqslant 1$. *If there are* $x_p \geqslant 0$ *such that for each* p *we have the enriched rooted tree contribution inequality*

$$w_p G(|t_p|x) \leqslant x_p, \tag{40}$$

then the cluster expansion for rooted connected graphs (the density) $C_p(w)$ *in powers of* w *converges absolutely.*

The sum on the left hand side is a partition function for a variable number of particles that interact with a particle at location p and also with each other. From a combinatorial point of view the inequality is related to an enriched rooted tree fixed point equation. Thus this Fernández–Procacci type theorem may be understand as a comparison of the enriched rooted tree fixed point equation with the rooted connected graph fixed point equation. The iteration function in the Fernández–Procacci condition is majorized by the iteration function in the Kotecký–Preiss condition. This because the factors $1 + t(p, q)$ are between 0 and 1. If one drops the product $(1 + t)^{\mathrm{Pair}(N)}$ with these factors, the expression in the condition can only become larger. The result is

$$w_p \sum_N \frac{1}{N!}(1+t)^{\mathrm{Pair}(N)}|t_p|^N x_p^N \leqslant w_p \sum_N \frac{1}{N!}|t_p|^N x^N = w_p \exp(t_p \cdot x). \tag{41}$$

It follows that the hypothesis for this Fernández–Procacci result is weaker than that for the Kotecký–Preiss hypothesis. As a consequence, the conclusion of the Fernández–Procacci type theorem is stronger than the conclusion of the corresponding Kotecký–Preiss type theorem.

4.3. Dobrushin's theorem. A system satisfies *hard-core self-repulsion* if for each p the Boltzmann factor $1 + t(p, p) = 0$. The following theorem is a slight reformulation of a result of Dobrushin [2]. The connection with the Fernández–Procacci result was noted in [4].

Theorem 3. *Suppose the activities and Boltzmann factors satisfy* $w_p \geqslant 0$ *and* $0 \leqslant 1 + t(p, q) \leqslant 1$. *Suppose also that the hard-core self-repulsion condition* $1 + t(p, p) = 0$ *is satisfied. If there are finite* $x_p \geqslant 0$ *with*

$$w_p \prod_q (1 + |t(p, q)|x_q) \leqslant x_p, \tag{42}$$

then the cluster expansion for the density $C_p(w)$ *converges absolutely.*

The hypothesis of this theorem is the Dobrushin tree finite fixed point condition. The relation to the other conditions is the following. It is clear that

$$w_p \prod_q (1 + |t(p,\, q)| x_q) \leqslant w_p \prod_q \exp(|t(p,\, q)| x_q) = w_p \exp\Big(\sum_q |t(p,\, q)| x_q\Big). \tag{43}$$

Thus the Kotecký–Preiss condition rooted tree condition implies the Dobrushin tree condition, so for hard-core self-repulsion the Dobrushin result implies the Kotecký–Preiss result. For a comparison with the Fernández–Procacci result, notice that with hard-core self-repulsion $(1+t)^{\mathrm{Pair}(N)} \neq 0$ implies $0 \leqslant N \leqslant 1$. So

$$w_p G(|t|_p x) \leqslant w_p \sum_{0 \leqslant N \leqslant 1} |t_p|^N x^N = w_p \prod_q (1 + t(p,\, q) x_q). \tag{44}$$

It follows that for hard-core self-repulsion the Dobrushin tree condition implies the Fernández–Procacci enriched rooted tree condition. Turning this around, the Fernández–Procacci result implies the Dobrushin result.

4.4. Particles in the continuum. This type of convergence result may be extended to the continuum [3]. Let Λ be a measurable space, corresponding to possible locations of the particles. For each n take U_n to be a finite set with n elements. Write $\boldsymbol{a}_n = (a_1, \ldots, a_n)$ for a function from U_n to Λ, that is, for a point in Λ^n. This is a particle configuration. The goal is to assign probabilities to sets of particle configurations.

The first ingredient in the physical model is a potential energy function $0 \leqslant V \leqslant +\infty$, taken to be a symmetric Borel measurable function on Λ^2. The second ingredient is a finite measure μ defined on Borel subsets of Λ. This represents activity, proportional to prior probability for finding particles in a particular region of space. Then the equilibrium probability measure for n particles is given by

$$d\,\mathrm{prob}_n(\boldsymbol{a}_n) = \frac{1}{\Xi}\frac{1}{n!} \exp\Big(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)\Big)\, d\mu(a_1) \cdots d\mu(a_n). \tag{45}$$

The normalization constant is the usual partition function

$$\Xi = \sum_{n=0}^{\infty} \frac{1}{n!} \int_\Lambda \cdots \int_\Lambda \exp\Big(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)\Big)\, d\mu(a_1) \cdots d\mu(a_n). \tag{46}$$

If E is a Borel subset of Λ, then the number of particles in E is given by the function $h_n(\boldsymbol{a}_n) = 1_E(a_1) + \cdots + 1_E(a_n)$. So the expected number of particles in E is obtained by integrating this with respect to the probability measure. This gives

$$n_E = \frac{1}{\Xi} \sum_{n=1}^{\infty} \frac{1}{n!} \int_\Lambda \cdots \int_\Lambda h_n(\boldsymbol{a}_n) \exp\Big(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)\Big)\, d\mu(a_1) \cdots d\mu(a_n), \tag{47}$$

where the particle indices range over numbers from 1 to n.

Permutation symmetry and a shift of index $n = m + 1$ lead to an equivalent expression

$$n_E = \frac{1}{\Xi} \sum_{m=0}^{\infty} \frac{1}{m!} \int_E \int_\Lambda \cdots \int_\Lambda \exp\Big(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)\Big)\, d\mu(a_0) \cdots d\mu(a_m). \tag{48}$$

Here the particle indices range from 0 to m.

The fundamental quantity to be bounded is the density

$$\rho(a_0) = \frac{1}{\Xi} \sum_{m=0}^{\infty} \frac{1}{m!} \int_\Lambda \cdots \int_\Lambda \exp\Big(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)\Big)\, d\mu(a_1) \cdots d\mu(a_m). \tag{49}$$

The particle indices range from 0 to m, but there is no integration over the index 0 variable. This is the expectation of the Boltzmann factor $\exp(-\beta \sum_{i=1}^{m} V(a_i,\, a_0))$ corresponding to an external particle fixed at location a_0. Since

$$n_E = \int_E \rho(a_0)\, d\mu(a_0), \tag{50}$$

it is the density of expected particle number with respect to activity.

Again we introduce the interaction factor $t(a_i,\, a_j) = \exp(-\beta V(a_i,\, a_j)) - 1$. Then $\exp(-\beta \sum_{\{i,j\}} V(a_i,\, a_j)) = \prod_{\{i,j\}}(1 + t(a_i,\, a_j))$, and the distributive law gives the expansion in graphs. The following theorem is a Fernández–Procacci type result extended to the continuum.

Theorem 4. *Consider the continuum gas model with interaction factor satisfying* $-1 \leqslant t(a,\, b) \leqslant 0$*, and with an activity measure* μ*. Suppose that there exists a finite measurable function* $y \geqslant 1$ *such that*

$$\sum_{n=0}^{\infty} \frac{1}{n!} \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{n} |t(a,\, a_i)| \prod_{i \neq j} (1 + t(a_i,\, a_j)) \prod_{j=1}^{n} y(a_j)\, d\mu(a_1) \cdots d\mu(a_n) \leqslant y(a). \tag{51}$$

Then the expansion for the density $\rho(a)$ *converges absolutely and has absolute value bounded by* $y(a)$*.*

References

[1] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, Encyclopedia of Mathematics and its Applications, vol. 67, Cambridge University Press, Cambridge, 1998. MR 1629341.

[2] R. L. Dobrushin, *Perturbation methods of the theory of Gibbsian fields*, Lectures on probability theory and statistics (Saint-Flour, 1994), Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, pp. 1–66. MR 1600880

[3] W. G. Faris, *A connected graph identity and convergence of cluster expansions*, J. Math. Phys. **49** (2008), no. 11, 113302, 14. MR 2468534

[4] R. Fernández and A. Procacci, *Cluster expansion for abstract polymer models. New bounds from an old approach*, Comm. Math. Phys. **274** (2007), no. 1, 123–140. MR 2318850

Department of Mathematics, University of Arizona, Tucson, AZ 85721 USA
E-mail address: `faris@math.arizona.edu`