Geometrical ambiguity of pair statistics: Point configurations

Y. Jiao

Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, New Jersey 08544, USA

F. H. Stillinger

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

S. Torquato*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA;
Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey 08544, USA;
Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA;
and Princeton Center for Theoretical Physics, Princeton University, Princeton, New Jersey 08544, USA

Point configurations have been widely used as model systems in condensed-matter physics, materials science, and biology. Statistical descriptors, such as the n-body distribution function g_n, are usually employed to characterize point configurations, among which the most extensively used is the pair distribution function g_2. An intriguing inverse problem of practical importance that has been receiving considerable attention is the degree to which a point configuration can be reconstructed from the pair distribution function of a target configuration. Although it is known that the pair-distance information contained in g_2 is, in general, insufficient to uniquely determine a point configuration, this concept does not seem to be widely appreciated and general claims of uniqueness of the reconstructions using pair information have been made based on numerical studies. In this paper, we present the idea of the distance space called the $\mathbb{D}$ space. The pair distances of a specific point configuration are then represented by a single point in the $\mathbb{D}$ space. We derive the conditions on the pair distances that can be associated with a point configuration, which are equivalent to the realizability conditions of the pair distribution function g_2. Moreover, we derive the conditions on the pair distances that can be assembled into distinct configurations, i.e., with structural degeneracy. These conditions define a bounded region in the $\mathbb{D}$ space. By explicitly constructing a variety of degenerate point configurations using the $\mathbb{D}$ space, we show that pair information is indeed insufficient to uniquely determine the configuration in general. We also discuss several important problems in statistical physics based on the $\mathbb{D}$ space, including the reconstruction of atomic structures from experimentally obtained g_2 and a recently proposed "decorrelation" principle. The degenerate configurations have relevance to open questions involving the famous traveling salesman problem.

I. INTRODUCTION

A collection of a finite or infinite number of points in d-dimensional Euclidean space $\mathbb{R}^d$ is called a *point configuration*. Point configurations are one of the most popular and widely used models for many-particle systems in various branches of modern science, including condensed-matter physics and materials science [1–5], statistical mechanics [6–8], discrete mathematics (packing problems) [9], astrophysics (distribution of galaxy clusters) [10,11], ecology (tree distributions in forests) [12], and biology (various cellular structures) [13]. Point configurations can exhibit a variety of degrees of disorder, from the most random Poisson distribution [2] to a perfectly ordered Bravais lattice [1]. The degrees of disorder can be quantified by discriminating order metrics [1,14], which, in their simplest forms, are scalars and normalized such that the most disordered system is associated with zero and the most ordered ones with unity.

In this paper, we focus on individual point configurations. In most circumstance, it is impossible and even unnecessary

to acquire detailed knowledge of all positions of the points in a single configuration. Instead, statistical descriptors, such as distribution functions, are typically employed to characterize point configurations. In particular, the n-body distribution function $g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is related to the probability of finding a generic configuration of n points at positions $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. It is well known that a set of n-body distribution functions $g_1, g_2, \ldots, g_n$ [2] is required to statistically characterize an n-point configuration completely. As $n \rightarrow \infty$ in the thermodynamic limit (e.g., the volume V which the n points occupy also increases to infinity such that the number of points per volume—number density $\rho = N/V$—is a well-defined finite number), the set contains an infinite number of correlation functions. For statistically homogeneous systems, which are the focus of this paper, g_n is translationally invariant and hence depends only on the relative displacements of the positions with respect to some chosen origin, say $\mathbf{x}_1$, i.e., $g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = g_n(\mathbf{x}_{12}, \mathbf{x}_{13}, \ldots, \mathbf{x}_{1n})$ with $\mathbf{x}_{ij} = \mathbf{x}_j - \mathbf{x}_i$. Thus, the one-body distribution function g_1 is just equal to the number density ρ. The important two-body quantity $g_2(\mathbf{x}_{12})$ is usually referred to as the pair distribution function. In the statistically isotropic case, g_2 is a radial function, i.e., $g_2(\mathbf{x}_{12}) = g_2(|\mathbf{x}_{12}|)$ and it is also called the radial distribution function. The radial distribution function, which is one of the

---

*torquato@electron.princeton.edu

most widely used structural descriptors, essentially provides the distribution of the point-pair separation distances and can be obtained experimentally via scattering of radiation [1]. The three-body function g_3 contains information about how the pair separations involved in g_2 are linked into triangles.

It is worth noting that by decorating the points in the system (e.g., letting equal-sized spheres be centered at each point), one can construct a two-phase random texture from a given point configuration. In general, there is an infinite number of ways to decorate a point configuration. In the characterization of random textures, the analog of the n-body distribution functions are the n-point correlation functions $S_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ [2], which give the probability of finding n points at positions $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in the phase of interest. In general, a complete statistical characterization of a continuum random texture requires an infinite set of S_n. Though under certain conditions, g_n of a point configuration and S_n of the associated decorated random texture might convey the same level of structural information (in fact the associated S_n can expressed as functional of g_n given the details of the decorating phase [2]), the former evidently reflects the essential geometrical features of the point configuration more directly.

An intriguing inverse problem that has been receiving considerable attention is the reconstruction (or construction) of configurations of a many-body system (essentially point configurations) that match the prescribed structural information of the system in the form of g_2 or S_2 obtained from either experiments or theoretical considerations. Examples include the reconstruction of random media [15–23] and colloidal suspensions [24], investigation of the iso-g_2 process [25] or g_2-invariant processes, and the realizability conditions of g_2 [26] as well as the more recent discovery of unusual disordered classical ground states [27]. X-ray scattering techniques have been an indispensable tool historically in the study of the structures of crystalline matter, and it has been generalized to probing disordered media [1]. In particular, the pair distribution function $g_2(\mathbf{r})$ is obtainable from the Fourier transform of the structure factor $S(\mathbf{k})$ [1], which is proportional to the scattering intensity (with the atomic structure function removed) and can be directly measured in experiments. With the obtained g_2, one can then employ various reconstruction techniques to generate realizations of the system of interest. Another related family of inverse problems is the reconstruction of pair interaction potential from a given radial distribution function $g_2(r)$ between particles, i.e., the inverse Monte Carlo problems [28]. It is worth mentioning that in quantum mechanics, the analog of the g_2-realizability condition is the representability condition for the two-body density matrix, i.e., the conditions the density matrix must satisfy in order to correspond to a physical state [29–32].

It is known that though the information contained in g_2 can be sufficient to completely characterize ordered point configurations in very special circumstances [33], it is generally devoid of crucial structural information to uniquely determine a disordered point configuration [34–36]. However, it seems that this aspect has not been widely appreciated and general claims of uniqueness of the reconstructions using g_2 or S_2 have been made based on numerical studies
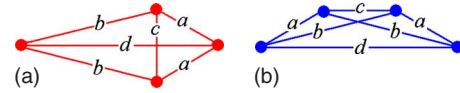

FIG. 1. (Color online) An example of two-dimensional four-point configurations possessing twofold degeneracy: (a) a "kite." (b) a "trapezoid." The specific distance sets are $a = (2x^2 - 3x + \frac{5}{4})^{1/2}$, $b = (2x^2 - x + \frac{1}{4})^{1/2}$, $c = 2x - 1$, $d = 1$, for $\frac{1}{2} < x < 1$. For $x > 1$, the outer boundary of the kite is no longer a quadrilateral but reduces to an isosceles triangle.

[18,23]. One aim of this paper is to show via a variety of examples the existence of distinct point configurations with identical pair-distance distributions (e.g., g_2), which implies the nonuniqueness of the reconstructions involving g_2 of these point configurations. Besides, general mathematical formalism to characterize the structural ambiguity of pair information is also devised.

Figure 1 shows two distinct configurations of four points in two dimensions with identical pair distances. In particular, one configuration (with the pair distances shown) resembles a "kite" and the other resembles a "trapezoid." In order to provide an in-depth presentation of the ambiguity of pair-distance distributions, it is necessary to examine the problem mathematically first and then discuss the physical implications. Some definitions are in order here. Two d-dimensional statistically homogeneous and isotropic n-point configurations $\Gamma_{d,n}^i$ and $\Gamma_{d,n}^j$ are identical if and only if they possess identical sets of k-body distribution functions g_k for $k = 1, 2, \ldots, n$. The configurations $\Gamma_{d,n}^i$ and $\Gamma_{d,n}^j$ are g_k distinct if and only if they process distinct n-body distribution functions for all $n \geq k$. A d-dimensional n-point configuration $\Gamma_{d,n}^1$ is k-fold degenerate if and only if there exist additional $(k-1)$ d-dimensional n-point configurations $\Gamma_{d,n}^i$ $(i = 2, \ldots, k)$ that are mutually g_3 distinct and also g_3 distinct from $\Gamma_{d,n}^1$, all of which possess the same two-body distribution function g_2. This definition of structural degeneracy rules out the possibility that two degenerate point configurations are trivially connected by translation, rotation, mirror reflection, or any of their combinations. Moreover, we consider that two-point configurations are equivalent (i.e., do not form a degenerate pair) if they are related by a trivial isotropic rescaling, which does not change the internal structure of the configuration. Thus, we see that the kite and the trapezoid are associated with the same set of pair distances (i.e., they are twofold degenerate), but the triangle information of the two is distinguishable [37]. It is worth noting the historically prominent Kirkwood superposition approximation of g_3, which replaces the three-body distribution function with a product of three pair distribution functions [38]. Because the separate members of our pair-distance degeneracy examples present distinct triangle (i.e., three-body) distributions, the conclusion must be that no functional of g_2 (Kirkwood or otherwise) can uniquely specify g_3.

It is clear that given g_2 associated with the degenerate point configurations, it is impossible even in principle to obtain a unique reconstruction, and each degenerate configuration should be recovered with equal probability. Therefore, an outstanding problem is to determine under what conditions the pair-distance information contained in g_2 could
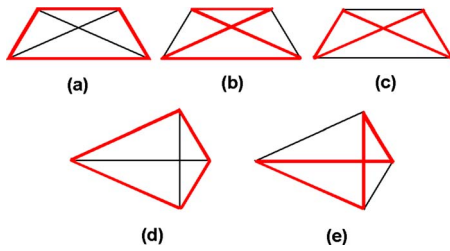
FIG. 2. (Color online) The three distinguishable circuits for the trapezoid (upper panel) and two distinguishable circuits for the kite (lower panel). The circuits are shown in thick lines. The circuit shown in (a) is the shortest route among all possibilities.

uniquely determine a point configuration, i.e., there is no associated structural degeneracy. A question with more practical importance is that how the point configurations would change when the measurement of g_2 is subject to slight imprecision, a common situation in experiments and numerical simulations.

In addition to their physical relevance, degenerate point configurations are also of mathematical interest. For example, open questions connected to the famous traveling salesman problem [39] can be raised. Given the degenerate configurations associated with the same set of pair distances, what are the optimal solutions of the traveling salesman problem for each configuration and are they unique? Are there special degenerate configurations whose solutions are identical? For the simplest "kite-trapezoid" example shown in Fig. 1, for the parameter values $x > 1/2$, the trapezoid has three distinguishable circuits and the kite has only two (see Fig. 2). The shortest route among all is presented by the trapezoid, i.e., a closed circuit visiting each vertex once and only once. For $x = 1/2$, both the kite and the trapezoid collapse onto a line segment, and in that limiting case all circuits have the same length. For more general and complicated degenerate configurations, such questions are notoriously difficult to solve; the problem belongs to the NP-complete class.

In this paper, we present the idea of the distance space (i.e., the $\mathbb{D}$ space), in which each dimension is associated with the separation distance between a given point pair. The pair-distance distribution of a particular point configuration is then presented by a single point in the $\mathbb{D}$ space. It is clear that not all the points in $\mathbb{D}$ space correspond to realizable configurations, i.e., the separation distances have to satisfy certain conditions such that they could be assembled into a point configuration. These conditions together define a (partially) bounded region in the $\mathbb{D}$ space. For example, for three-point configurations in $\mathbb{R}^2$ (i.e., triangles), the region of the feasible distances is an open "pyramid" in the three-dimensional $\mathbb{D}$ space, as shown in Fig. 3. When degeneracy exists, the region of the feasible distances is generally a complicated closed intersection of several such simple curved pyramids in high dimensions. The determination of the region of feasible distances is equivalent to obtaining the conditions of a realizable g_2, i.e., a pair distribution function that can be associated with a point configuration. We emphasize that our focus here is the pair-distance distribution of a *single* point configuration (or a *single* realization of a point pro-
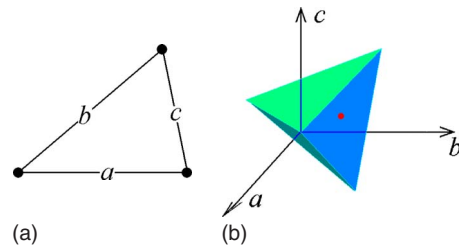

FIG. 3. (Color online) (a) A three-point configuration (i.e., a triangle) in $\mathbb{R}^2$. (b) The region of feasible distances in the $\mathbb{D}$ space (bounded by the blue planes). The three pair distances of the triangle shown in (a) is represented as a point (red spot) in the $\mathbb{D}$ space.

cess). Thus, the function g_2 we examine should be interpreted as the *volume-averaged* pair distribution [14]. Although in the infinite-volume limit, volume averaging is equivalent to ensemble averaging for ergodic systems, we do not explicitly consider this limiting case here. Discussions of the realizability conditions and the nonuniqueness issue for ensemble-averaged pair distributions can be found in Refs. [25,26,40,41]. Using the $\mathbb{D}$ space, we will answer various aspects of the aforementioned questions concerning the degenerate point configurations and the nonuniqueness issue of the reconstruction. We will show that the utility of the $\mathbb{D}$ space also improves our understanding of various important problems in statistical physics such as the recently proposed decorrelation principle in high-dimensional Euclidean space [42]. In a sequel to this paper, we will extend the present analysis to understand degeneracy issues pertaining to two–phase random media which is a larger classification than point configurations [43].

The rest of the paper is organized as follows. In Sec. II, we discuss the $\mathbb{D}$ space in detail and derive the conditions for feasible distances and for the occurrence of degeneracy, through which we show that in general degeneracy is rare. In Sec. III, we provide a variety of examples of degenerate point configurations and illustrate how the conditions derived in Sec. II could be employed to construct point configurations with specific degeneracy. In Sec. IV, we discuss several problems in statistical physics such as the reconstruction of atomic structures from experimentally obtained g_2 and the decorrelation principle, based on the idea of the $\mathbb{D}$ space. Finally, we make concluding remarks.

II. DISTANCE SPACE $\mathbb{D}$

In this section, we will discuss in detail the $\mathbb{D}$ space. In particular, we will derive the conditions under which the pair distances could be assembled into a point configuration, i.e., the feasibility conditions as well as the conditions under which the pair distances correspond to degenerate point configurations. We will first study a four-point configuration in $\mathbb{R}^2$ to illustrate the idea and then consider the general n-point configurations in $\mathbb{R}^d$.

A. Simple example: Four-point configuration in $\mathbb{R}^2$

Consider a four-point configuration $\Gamma_{2,4}$ in $\mathbb{R}^2$ (see Fig. 4) and the associated six-dimensional $\mathbb{D}$ space. We would like
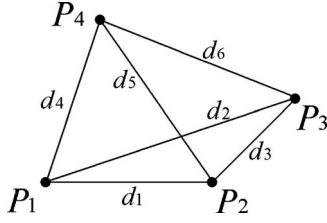
FIG. 4. A four-point configuration in $\mathbb{R}^2$.

to know the answers to the following two questions:

(i) What are the conditions the six pair separation distances must satisfy so that they correspond to a four-point configuration in $\mathbb{R}^2$?

(ii) What are the conditions the pair distances must satisfy so that they correspond to k-fold degenerate four-point configurations in $\mathbb{R}^2$?

To answer these questions, we consider a particular construction as follows. Suppose the six pair distances are elements of the set $\Omega = \{d_1, d_2, \ldots, d_6\}$, which can be further partitioned as $\Omega = \{\Omega_1, \Omega_2, \Omega_3, \Omega_4\}$, where $P_1 = \{\Phi\}$ (Φ is the null set), $P_2 = \{d_1\}$, $P_3 = \{d_2, d_3\}$, and $P_4 = \{d_4, d_5, d_6\}$. We will see that such a partition enables us to associate the pair distances with the corresponding points in a convenient way. Recall that from our definition (Sec. I), point configurations are considered identical if they are connected by translation, rotation, mirror reflection, and any of their combinations. Thus, we can put point P_1 at the origin of a Cartesian coordinate system and put point P_2 on one of the two orthogonal axes of the coordinate system separated from the origin (i.e., P_1) by a distance d_1. Note different choices of the position of P_1 and the orientation of the line segment $\overline{P_1 P_2}$ lead to point configurations that are identical up to translations and rotations. For point P_3, we can either let $\overline{P_1 P_3} = d_2$, $\overline{P_2 P_3} = d_3$ or $\overline{P_1 P_3} = d_3$, $\overline{P_1 P_2} = d_2$. The two choices correspond to two configurations connected by a mirror reflection, which are considered identical and either choice is acceptable. Without loss of generality, we choose $\overline{P_1 P_3} = d_2$, $\overline{P_2 P_3} = d_3$. Finally, we choose $\overline{P_1 P_4} = d_4$, $\overline{P_2 P_4} = d_5$, $\overline{P_3 P_4} = d_6$ for point P_4.

We see from the above construction that the positions of points P_3 and P_4 are determined with respect to the line segment defined by points P_1 and P_2 as a "reference" structure. Note that the line segment is a one-dimensional simplex. In $\mathbb{R}^2$, the position of a point is completely determined by specifying two distances from the point of interest to the vertices of a reference line segment, given that the distances involved satisfy the triangular inequality, i.e., the triangle formed by the point of interest and the two vertices of the reference line segment possess non-negative area. The area Δ of a triangle with edges a, b, c is related to the Cayley-Menger determinant [44], i.e.,

$$\Delta^2 = -\frac{1}{16} \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & a^2 & b^2 \\ 1 & a^2 & 0 & c^2 \\ 1 & b^2 & c^2 & 0 \end{vmatrix}. \qquad (1)$$

Thus, for point P_3, we obtain

$$d_1^4 + d_2^4 + d_3^4 - 2d_1^2 d_2^2 - 2d_1^2 d_3^2 - 2d_2^2 d_3^2 > 0, \qquad (2)$$

and for point P_4, we obtain

$$d_1^4 + d_4^4 + d_5^4 - 2d_1^2 d_4^2 - 2d_1^2 d_5^2 - 2d_4^2 d_5^2 > 0. \qquad (3)$$

Inequalities (2) and (3) define a partially bounded region in the six-dimensional $\mathbb{D}$ space, the lower-dimensional analog of which is the open pyramid shown in Fig. 3. The distance d_6 between points P_3 and P_4 is also completely determined by $d_1, d_2, \ldots, d_5$ via

$$\begin{vmatrix} -2d_1^2 & (d_3^2 - d_1^2 - d_2^2) & (d_5^2 - d_1^2 - d_4^2) \\ (d_3^2 - d_1^2 - d_2^2) & -2d_2^2 & (d_6^2 - d_2^2 - d_4^2) \\ (d_5^2 - d_1^2 - d_2^2) & (d_6^2 - d_2^2 - d_4^2) & -2d_5^2 \end{vmatrix} = 0, \qquad (4)$$

which results from the requirement that all the 3×3 minors of the Gram matrix [36] involving the distances possess zero determinant. We will discuss the Gram matrix in detail in Sec. II B. Equation (4) defines a curved hypersurface in the $\mathbb{D}$ space, whose intersection with the region defined by Eqs. (2) and (3) contains the feasible distances Ω that can be assembled into a four-point configuration in $\mathbb{R}^2$. We call Eqs. (2)–(4) the *feasibility* conditions. Note that for the four-point configuration in $\mathbb{R}^2$, only five pair distances can be chosen almost independently subject to the mild triangle inequality constraint. Thus, we define the *free dimension* of the $\mathbb{D}$ space to be the number of pair distances that are only constrained by inequalities, which total to five here. The free dimension is also the dimension of the region for the feasible distances, which is also referred to as the *feasible region*.

Now we can answer question (i) given at the beginning of this section easily. Suppose a list of distances is given, when any one permutation of these distances satisfies the feasibility conditions [Eqs. (2)–(4)], the pair distances correspond to a four-point configuration in $\mathbb{R}^2$. However, such a simple answer does not exist for (ii). For the pair distances to correspond to k-fold degenerate point configurations, a necessary condition is that the dimension of the intersection of the feasible regions for the k permutations of the pair distances is nonzero. In other words, each permutation of the pair distances is associated with a set of feasibility conditions, and a feasible region can be constructed. To obtain a k-fold degeneracy, all sets of the feasibility conditions need to be satisfied simultaneously, which is only possible when the intersection of the feasible regions is at least a single curve in the $\mathbb{D}$ space. For the two-dimensional four-point configuration of interest, the free dimension is five, which leads to an upper bound on the order of the degeneracy, i.e., $k_{max} = 5$. This condition is only a necessary one because there are certain permutations that lead to identical configurations, such as those that correspond to the permutations among the point indices, which do not change the structure of the configuration, since the points are indistinguishable. For example, $\Omega_1 = \{d_1, d_2, d_3, d_4, d_5, d_6\}$ and $\Omega_2 = \{d_1, d_4, d_5, d_2, d_3, d_6\}$ correspond to an exchange of P_3 and P_4, which possess the identical feasible regions and thus do not contribute to the degeneracy. No further conclusions can be made without knowing the details of how the distances are permuted. In

Sec. III, we will construct concrete examples of degenerate $\Gamma_{2,4}$, where the details of permutations are considered.

B. General formulation: Feasibility conditions

The generalization of the above formulation is straightforward. Note in the following discussion in this section, we assume $n > (d+1)$; the case when $n = d+1$ (i.e., the simplex configurations) is discussed in detail in Sec. III A and the case when $n < (d+1)$ is trivial. Consider an n-point configuration $\Gamma_{d,n}$ in $\mathbb{R}^d$, which possesses $m = n(n-1)/2$ pair distances $\Omega = \{d_1, d_2, \ldots, d_m\}$. The distances can be further partitioned, i.e., $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_n\}$, where $\Omega_1 = \{\Phi\}$, $\Omega_2 = \{d_1\}, \ldots$, $\Omega_i = \{d_{(i^2-3i+3)/2}, \ldots, d_{(i^2-i)/2}\}, \ldots$, $\Omega_n = \{d_{(n^2-3n+3)/2}, \ldots, d_{(n^2-n)/2}\}$. Following the same construction procedure prescribed in Sec. II A, the distances associated with the first d points, i.e., $\Omega_1, \Omega_2, \ldots, \Omega_d$, are assembled into a $(d-1)$-dimensional simplex as the reference structure. Each point $P_i (i > d)$ is associated with $(i-1)$ distances and the position of point P_i is completely determined by specifying the d distances from P_i to the vertices of the reference structure, given that the d-dimensional simplex formed by P_i and the vertices of the reference structure possesses a nonnegative volume. In particular, denote the $(d+1)$ vertices of the d-dimensional simplex by $\mathbf{v}_i$ $(i = 1, \ldots, d+1)$, we can define a $(d+1) \times (d+1)$ distance matrix M, i.e.,

$$M_{ij} = M_{ji} = \|\mathbf{v}_i - \mathbf{v}_j\|^2, \tag{5}$$

where $\|\cdot\|$ denotes the L^2 norm of a d-dimensional vector and M_{ij} (M_{ji}) is the squared distance between vertices i and j. The volume Δ of the simplex is then given by the Cayley-Menger determinant, i.e.,

$$\Delta_d^2 = \frac{(-1)^{d+1}}{2^d (d!)^2} |\hat{M}| \geq 0, \tag{6}$$

where $\hat{M}$ is a $(d+2) \times (d+2)$ matrix obtained from M by bordering M with a top row $(0, 1, \ldots, 1)$ and a left column $(0, 1, \ldots, 1)^T$. For example, Eq. (6) reduces to Eq. (1) in $\mathbb{R}^2$, and in $\mathbb{R}^3$ we obtain

$$\Delta^2 = \frac{1}{288} \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & M_{12} & M_{13} & M_{14} \\ 1 & M_{21} & 0 & M_{23} & M_{24} \\ 1 & M_{31} & M_{32} & 0 & M_{34} \\ 1 & M_{41} & M_{42} & M_{43} & 0 \end{vmatrix}. \tag{7}$$

The requirement that the constructed d-dimensional simplex possesses non-negative volume leads to higher-dimensional analogs of the well-known triangle inequalities in two dimensions, which we will refer to as *simplex inequalities*. In general, each set of the simplex inequalities associated with a point P_i $(i > d)$ defines a partially bounded region in the $\mathbb{D}$ space, the intersection of which is a high-dimensional analog of the open pyramid shown in Fig. 3(b).

It is clear from the above construction that a point configuration $\Gamma_{d,n}$ can be completely determined by only specifying $f = [\frac{1}{2} d(d-1) + (n-d)d]$ distances (e.g., "free" distances), satisfying the simplex inequalities. Thus, the free

dimension of the $\mathbb{D}$ space is f and the remaining $(m-f)$ pair distances (e.g., "constrained" distances) cannot be chosen freely but instead are completely determined by the f free distances. To obtain the relations between the constrained and free distances, we will employ the following theorem [36].

Theorem 1. For a set of n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in $\mathbb{R}^d$ $(n > d)$, let the Gram matrix be defined by $G_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$, where $\langle \cdot \rangle$ denotes the inner product. Then all $(d+1) \times (d+1)$ minors of G must have zero determinant.

The proof of the theorem is given in Ref. [36]. It is essentially another way of stating the fact that there are at most d linearly independent vectors among $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a d-dimensional Euclidean space. Without loss of generality, we can choose the origin at $\mathbf{v}_1$ and obtain

$$\hat{G}_{ij} = \langle \hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j \rangle = \langle \mathbf{v}_i - \mathbf{v}_1, \mathbf{v}_j - \mathbf{v}_1 \rangle. \tag{8}$$

Consider the identity [36],

$$\langle \mathbf{v}_i - \mathbf{v}_1, \mathbf{v}_j - \mathbf{v}_1 \rangle = \frac{1}{2} [\langle \mathbf{v}_i - \mathbf{v}_1, \mathbf{v}_i - \mathbf{v}_1 \rangle + \langle \mathbf{v}_j - \mathbf{v}_1, \mathbf{v}_j - \mathbf{v}_1 \rangle - \langle \mathbf{v}_i$$
$$- \mathbf{v}_j, \mathbf{v}_i - \mathbf{v}_j \rangle], \tag{9}$$

we obtain that

$$\hat{G}_{ij} = \frac{1}{2} (\hat{d}_{i1}^2 + \hat{d}_{j1}^2 - \hat{d}_{ij}^2), \tag{10}$$

where $\hat{d}_{ij}$ is the distance between the two points i and j. Thus, we see the requirement that all $(d+1) \times (d+1)$ minors of $\hat{G}$ have zero determinant (denoted by $M_{(d+1)}$), i.e.,

$$|M_{(d+1)}(\hat{G})| = 0 \tag{11}$$

leads to fourth-order algebraic equations, involving the $(m-f)$ constrained distances. It is clear that each constrained distance can be explicitly expressed as a function of the free distances alone. For four-point configuration in $\mathbb{R}^2$, Eq. (11) gives Eq. (4). Note these equalities define curved hypersurfaces in the $\mathbb{D}$ space. The intersection of these curved hypersurfaces as well as the partially bounded regions defined by Eq. (6) gives the feasible region of the $\mathbb{D}$ space, i.e., when any permutation of the $m = n(n-1)/2$ pair distances lies within the feasible region, these distances can be assembled into an n-point configuration in $\mathbb{R}^d$.

C. General formulation: Necessary conditions of degeneracy

For the distances Ω to correspond to k-fold degenerate point configurations $\Gamma_{d,n}^1, \Gamma_{d,n}^2, \ldots, \Gamma_{d,n}^k$ $(k \leq f)$, the feasibility conditions for k distinct permutations of Ω should be satisfied simultaneously. The feasibility conditions associated with any particular permutation of Ω include a set of equalities, which would reduce the dimension of the feasible region in the $\mathbb{D}$ space. Suppose that for each distinct permutation, only one additional equality constraint is introduced. Then we can obtain an upper bound on the order of the degeneracy, i.e., $k_{\max} = f$, which corresponds to a feasible region that has been reduced to a single curve (with one free

dimension). That is, only one distance can be chosen arbitrarily. However, different choices of the single free distance correspond to trivial isotropic rescaling of the entire configuration, which leads to no degeneracy based on our definition. Note that if the permutation does not introduce new feasibility conditions, it corresponds to a permutation of the point indices, which does not affect the structure of the configuration.

The properties of the feasible regions have important implications. As we have seen, the structural degeneracy would reduce the dimension of the feasible regions, the volume of which is proportional to the number of feasible distance sets. For a particular n-point configuration, we could in principle identify all feasible distance sets by exploring the whole feasible region in the $\mathbb{D}$ space point by point. However, the distance sets associated with degeneracies can only lie on a hypersurface with lower dimensions than the feasible region. The volume ratio of the hypersurface to the feasible region, which is also the number ratio of the distance sets associated with degeneracies to those without degeneracies, is vanishingly small. In other words, although degeneracies exist they are extremely rare. This might explain why perfect reconstructions (identical match of the pair distances and the configurations up to translations, rotations and mirror reflections) can be obtained numerically [18,23]. However, the general conclusion that pair statistics alone would uniquely determine the configurations could not be made only based on those numerical results, as we will show in the next section via a variety of examples of degeneracy.

III. EXAMPLES OF DEGENERATE POINT CONFIGURATIONS

In this section, we construct a variety of degenerate point configurations using the general scheme developed in Sec. II. In particular, we study the degeneracies of simplices in $\mathbb{R}^d$, four-point configurations in $\mathbb{R}^2$, and specific n-point configurations in $\mathbb{R}^d$ possessing twofold degeneracy.

A. Degenerate d-dimensional simplices

A simplex in $\mathbb{R}^d$ is the convex hull of a set of $(d+1)$ points $\Gamma_{d,d+1}$ that do not all lie on the same $(d-1)$-dimensional hyperplane. A simplex in $\mathbb{R}^2$ is a triangle and in $\mathbb{R}^3$ is a tetrahedron. Simplices in $\mathbb{R}^d$ $(d \geq 4)$ can be considered to be d-dimensional generalizations of the three-dimensional tetrahedron. The simplex is so named because it represents the simplest possible polyhedron in the given dimension. The volume a d-dimensional simplex is given by Cayley-Menger determinant (6).

A unique feature of simplex configurations $\Gamma_{d,d+1}$ is that their feasibility conditions only include the simplex inequalities. These inequalities define a partially bounded region possessing the same dimensions as the $\mathbb{D}$ space. In other words, the free dimension of the feasible region is not reduced due to degeneracy. Thus, one should expect that it is much easier to obtain highly degenerate simplices than other point configurations.

Suppose we have a distance set $\Omega = \{d_1, d_2, \ldots, d_m\}$ $(m = d(d+1)/2)$. It is clear that if we choose $d_i = \bar{d} + \delta_i$, where

δ_i $(i = 1, 2, \ldots, m)$ are mutually distinct small numbers. they will satisfy all the simplex inequalities and correspond to a point in the vicinity of the centroid of the feasible region. The maximum magnitude of the δ's depends on the boundaries of the feasible region, which we need not to worry about for the moment, as long as the δ's are sufficiently small and mutually distinct.

In $\mathbb{R}^2$ $(d=2)$, the three distances can be assembled into a unique triangle, i.e., we have $k_{\max}^{(2)} = 1$. This can also be seen from the following argument. Since configurations connected by translations and rotations are considered identical, we could pick any one of the three distances along one of the coordinate axes starting from the origin and require the same for the corresponding distance of all possible degenerate configurations. In this way, we rule out translations and rotations in a plane. There are only two distances left, which can be assembled into a triangle in two ways. However, the two resulting triangles are mirror image of each other. Thus, we have $k_{\max}^{(2)} = 2!/2 = 1$, where "!" indicates factorial.

In $\mathbb{R}^3$ $(d=3)$, we similarly choose one of the six distances as the reference distance, and the remaining five are assigned to different edges of a tetrahedron, which results in 5! tetrahedra. However, among these tetrahedra there are pairs that are connected by mirror reflection, which has to be excluded. Two mirror reflection plane can be identified: one perpendicular to the reference distance and the other contains the reference distance. This further reduces the number of distinct tetrahedra by a factor of 1/4. Thus, we obtain $k_{\max}^{(3)} = 5!/(2 \times 2) = 30$.

Generally, in d dimensions when one of the $m = d(d+1)/2$ distances is chosen as the reference distance, there are $(m-1)!$ ways to assemble the remaining $(m-1)$ distances into a simplex in $\mathbb{R}^d$. However, $(d-1)$ hyperplanes (among which one contains the reference distance and the others are perpendicular to it) can be identified that are mirror reflection hyperplanes of the simplex. Each mirror reflection reduces the number of distinct simplices by a factor of 1/2. Thus, we have

$$k_{\max}^{(d)} = \frac{(m-1)!}{2^{(d-1)}} = \frac{[d(d+1)/2 - 1]!}{2^{(d-1)}}. \qquad (12)$$

We can see that for simplex configurations in $\mathbb{R}^d$ with $d \geq 3$, $k_{\max}^{(d)}$ is significantly larger than the dimension of the feasible region $f = d(d+1)/2$, which indeed implies a high level of degeneracies associated with these configurations.

B. Two-dimensional four-point configurations

We show here how the conditions determining the feasible region in $\mathbb{D}$ space can be employed to construct four-point configurations $\Gamma_{2,4}$ in $\mathbb{R}^2$ with k-fold degeneracy. As pointed out in Sec. II, the feasibility conditions are only necessary and the details of how the distances are permuted must be considered.

The relations of the six distances of $\Gamma_{2,4}$ are given by Eq. (4) for the particular order $\Omega = \{d_1, d_2, d_3, d_4, d_5, d_6\}$. For a permutation Ω^*, the variable d_i in Eq. (4) should be replaced by the ith element of Ω^*, which generally would lead to a different equation for the six distances. As mentioned in the
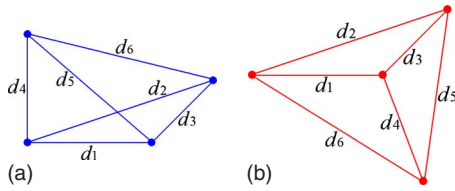
FIG. 5. (Color online) An example of twofold-degenerate four-point configurations in $\mathbb{R}^2$. The distances are given by $d_1=1$, $d_2=1.58114...$, $d_3=0.70710...$, $d_4=0.87228...$, $d_5=1.32698...$, $d_6=1.54551...$.

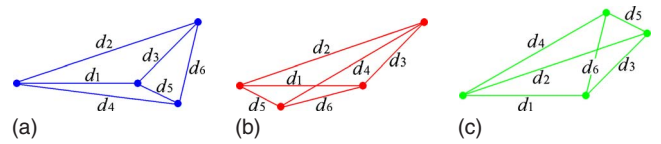FIG. 6. (Color online) An example of threefold-degenerate four-point configurations in $\mathbb{R}^2$. The distances are given by $d_1=1$, $d_2=1.581144...$, $d_3=0.70710...$, $d_4=1.34371...$, $d_5=0.37267...$, $d_6=0.68718...$.

last section, we could choose d_1 as the reference distance to rule out translation and rotation and only consider the permutations of the remaining five distances, which gives $5!=120$ distinct equations. Without loss of generality, we could also choose $d_1=1$, which corresponds to a trivial isotropic rescaling of the entire point configuration.

In principle, a k-fold degeneracy ($k \leq k_{max}=5$) could be constructed by requiring that the k equations for the five distances corresponding to k distinct permutations hold simultaneously. However, we find that high level degeneracies (those with k close to k_{max}) are difficult to realize. In particular, when k is large the equations for the distances possess roots that are algebraically multiple, i.e., Ω contains two or more equal valued distances, which leads to configurations connected by rotations and mirror reflections. Thus, the number of distinct configurations associated with the distances is smaller than k. For n-point configurations, the largest k that we have realized is $\hat{k}=n-1$. Due to space limitation, we could not exhaust all degeneracies for each k (i.e., about C_{120}^k cases) in this paper and only provide a few specific examples.

For $k=2$, requiring the distance permutations $\Omega_1 = \{d_1,d_2,d_3,d_4,d_5,d_6\}$ and $\Omega_2 = \{d_1,d_2,d_3,d_6,d_4,d_5\}$ to hold simultaneously yields

$$D(\Omega_1) = D(d_1,d_2,d_3,d_4,d_5,d_6) = 0,$$

$$D(\Omega_2) = D(d_1,d_2,d_3,d_6,d_4,d_5) = 0, \quad (13)$$

where $D(x_1,x_2,x_3,x_4,x_5,x_6)$ is the multinomial given by

$$
D(x_1,\ldots,x_6)
$$
$$
= \begin{vmatrix} -2x_1^2 & (x_3^2-x_1^2-x_2^2) & (x_5^2-x_1^2-x_4^2) \\ (x_3^2-x_1^2-x_2^2) & -2x_2^2 & (x_6^2-x_2^2-x_4^2) \\ (x_5^2-x_1^2-x_4^2) & (x_6^2-x_2^2-x_4^2) & -2x_5^2 \end{vmatrix}.
$$
$$(14)$$

Equations (13) reduce the free dimensions of the $\mathbb{D}$ space from five to four. Without loss of generality, we choose $d_1=1$, $d_2=1.58114...$, $d_3=0.70710...$, $d_4=0.87228...$, and solve Eqs. (13) to obtain $d_5=1.32698...$, $d_6=1.54551...$. The twofold-degenerate configurations are shown in Fig. 5. It should be noted in passing that the kite-trapezoid example shown earlier in Fig. 1 is a special case of this four-point twofold degeneracy, for which the shapes each have a reflection symmetry. If we require the permutations $\Omega_3 = \{d_1,d_2,d_3,d_4,d_6,d_5\}$ and $\Omega_4 = \{d_1,d_2,d_3,d_5,d_4,d_6\}$ to hold

simultaneously, the same degeneracy can be obtained because the apparently different groups of distance permutations (Ω_1,Ω_2) and (Ω_3,Ω_4) correspond to the permutation of indistinguishable points.

Similarly, for $k=3$, we choose $\Omega_1=\{d_1,d_2,d_3,d_4,d_5,d_6\}$, $\Omega_2=\{d_1,d_2,d_3,d_5,d_6,d_4\}$, and $\Omega_3=\{d_1,d_2,d_3,d_4,d_6,d_5\}$ to hold simultaneously, which reduces the free dimensions to three. By choosing $d_1=1$, $d_2=1.581144...$, $d_3=0.70710...$, equations $D(\Omega_i)=0$ ($i=1,2,3$) can be solved to yield $d_4=1.34371...$, $d_5=0.37267...$, $d_6=0.68718...$. The threefold degenerate configurations are shown in Fig. 6.

C. d-dimensional n-point configurations with twofold degeneracy

In general, the feasible region of k-fold degenerate n-point configurations in $\mathbb{R}^d$ can be obtained by carrying out a similar calculation used in the previous section, which would be extremely tedious. However, when the point configurations possess certain symmetries, particular degeneracies can readily be constructed. Here we provide constructions of twofold-degenerate n-point configurations in $\mathbb{R}^d$ by taking advantage of their symmetries.

Consider a *centrally symmetric* n_1-point configuration $\Gamma_{d,n_1}^{(1)}$ in $\mathbb{R}^d$, i.e., there exists a center O_1 such that for every point $P_i^{(1)}$ in $\Gamma_{d,n_1}^{(1)}$ there exists a point $P_j^{(1)}$, for which the line segment $\overline{P_i^{(1)}P_j^{(1)}}$ passing O_1 is bisected by O_1 (note that $i=j$ is allowed), i.e., $P_i^{(1)}$ and $P_j^{(1)}$ are points of inversion symmetry about O_1. Consider another centrally symmetric point configuration $\Gamma_{d,n_2}^{(2)}$, in which all the n_2 points are distributed symmetrically on a one-dimensional line $l^{(2)}$ embedded in $\mathbb{R}^d$. Denote the symmetry center of $\Gamma_{d,n_2}^{(2)}$ by O_2. We require that the line segment $\overline{O_1O_2}$ is perpendicular to $l^{(2)}$. Finally, consider the centrally symmetric point configuration $\Gamma_{d,2n_3}^{(3)}$, the $2n_3$ points of which are also distributed symmetrically on a one-dimensional line $l^{(3)}$ that is parallel to $l^{(2)}$ with the symmetry center coinciding with O_1. $\Gamma_{d,2n_3}^{(3)}$ can be further partitioned into two subsets: Λ_{d,n_3}^p which contains n_3 points in $\Gamma_{d,2n_3}^{(3)}$ such that no two points in Λ_{d,n_3}^p are symmetric about O_1 (i.e., they are "primary" points), and Λ_{d,n_3}^d, which contains the remaining n_3 points of $\Gamma_{d,2n_3}^{(3)}$ (i.e., the "dual" points). It is clear that

$$\Gamma_{d,(n_1+n_2+n_3)}^p = \Gamma_{d,n_1}^{(1)} \cup \Gamma_{d,n_2}^{(2)} \cup \Lambda_{d,n_3}^p,$$

$$\Gamma_{d,(n_1+n_2+n_3)}^d = \Gamma_{d,n_1}^{(1)} \cup \Gamma_{d,n_2}^{(2)} \cup \Lambda_{d,n_3}^d, \quad (15)$$

form a degenerate pair, i.e., the distances from the n_3 primary points to the remaining (n_1+n_2) points in $\Gamma_{d,(n_1+n_2+n_3)}^p$ are
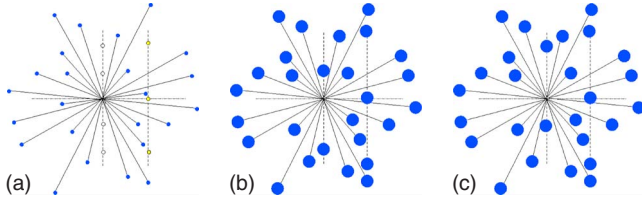
FIG. 7. (Color online) An example of twofold-degenerate configurations in $\mathbb{R}^2$ constructed as described in the text. (a) The points in $\Gamma_{d,n_1}^{(1)}$ are shown in blue (dark gray), the points in $\Gamma_{d,n_1}^{(1)}$ are shown in yellow along the right most dashed vertical line, and the points in $\Gamma_{d,2n_3}^{(3)}$ are shown as void circles. (b) and (c) show the degenerate configuration pair.

identical to those from the n_3 dual points to the remaining (n_1+n_2) points in $\Gamma_{d,(n_1+n_2+n_3)}^d$, while the two resulting configurations are not connected by translation, rotation, mirror reflection, or any of their combinations. Specific twofold degeneracy examples in $\mathbb{R}^2$ and $\mathbb{R}^3$ are shown in Figs. 7 and 8, respectively.

IV. DISCUSSION

The $\mathbb{D}$ space concept can be applied to reconcile a variety of problems in statistical physics, such as the reconstruction of atomic structures from experimentally obtained g_2, and the decorrelation principle, which we will discuss in the ensuing subsections.

A. Reconstruction of atomic structures from experimentally obtained g_2

As mentioned in Sec. I, a knowledge of atomic structures of condensed matter can be obtained via x-ray scattering experiments. In particular, the two-body distribution function $g_2(r)$ is related to the Fourier transform of the structure factor $S(k)$, which is proportional to the scattering intensities (with the atomic structure function removed). For ideal crystalline structures (without any thermal agitation of the atomic centers), g_2 consists of a series of Dirac delta functions at specific distances. For disordered structures (lack of long-range order), g_2 is generally a continuous damped oscillating function that decays to its long-range value very quickly. Interestingly, it seems that though the pair information contained in g_2 of the crystalline matter would determine the structures
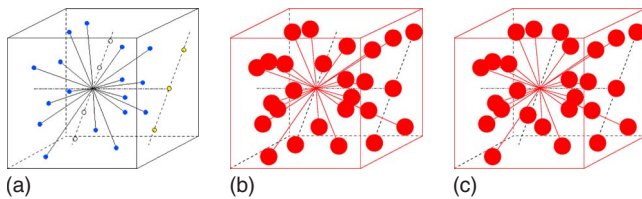

FIG. 8. (Color online) An example of twofold-degenerate configurations in $\mathbb{R}^3$ constructed as described in the text. (a) The points in $\Gamma_{d,n_1}^{(1)}$ are shown in blue (dark gray), the points in $\Gamma_{d,n_1}^{(1)}$ are shown in yellow along the dashed line on the right surface of the cubic box, and the points in $\Gamma_{d,2n_3}^{(3)}$ are shown as void circles. (b) and (c) show the degenerate configuration pair.

to high accuracy, it is not the case for disordered structures.

The reason can be easily seen if we consider the $\mathbb{D}$ space. For an ordered point configuration (i.e., a lattice), there are strong dependencies among the distances besides those required by the feasibility conditions. For example, consider a d-dimensional Bravais lattice whose basis vectors are $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d$. The vector connecting any two points in the lattice can then be expressed as

$$\mathbf{d} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + \cdots + n_d\mathbf{a}_d, \tag{16}$$

where n_i $(i=1,\ldots,d)$ are integers. Thus, the distance d between any two lattice points is given by

$$d^2 = \sum_{i=1}^{n} n_i^2 \langle \mathbf{a}_i, \mathbf{a}_i \rangle + \sum_{i \neq j}^{n} n_i n_j \langle \mathbf{a}_i, \mathbf{a}_j \rangle, \tag{17}$$

where $\langle , \rangle$ denotes the inner product of two vectors. Note that $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \frac{1}{2}(\langle \mathbf{a}_i, \mathbf{a}_i \rangle + \langle \mathbf{a}_j, \mathbf{a}_j \rangle - \langle \mathbf{a}_i - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j \rangle)$. Thus, Eq. (17) implies that every distance of an ordered point configuration can be obtained if the lengths of the basis lattice vectors and the distances between the end points of different basis lattice vectors are specified. In other words, Eq. (17) further reduces the free dimensions of the $\mathbb{D}$ space of the ordered point configuration in $\mathbb{R}^d$ to $f = d(d+1)/2$. The additional conditions given by Eq. (17) significantly reduce the number of feasible permutations of the distances. A unique feature of the distances for lattices is that the basis lattice vectors are associated with the smallest distances. To completely reconstruct the lattice configuration, the smallest $d(d+1)/2$ distances are selected to be assembled into a simplex in $\mathbb{R}^d$ defined by the common origin and the end points of all the lattice vectors, which in turn determines the fundamental cell of the lattice. In $\mathbb{R}^2$, three feasible distances uniquely determine a triangle and, thus, the rhombical fundamental cell. In $\mathbb{R}^3$, there are maximally 30 ways that the six distances could be assembled into a tetrahedron. However, even for a twofold degeneracy, the number of equality constraints introduced by Eq. (17) [i.e., for two permutations of the 6 distances, Eq. (17) should hold simultaneously] is much larger than the free dimensions of the system, which generally rules out all nontrivial solutions. Indeed, it is known that for $d \leq 3$, pair distances are sufficient to uniquely determine Bravais lattices. However, in high dimensions, degeneracies of Bravais lattices can be constructed [9].

For a disordered structure, Eq. (17) does not hold and the values of distances would form a continuous spectrum in the infinite-volume limit. We consider the idealized case that there are a finite number of well-defined distances and try to reconstruct the configuration from them. An important point is that no matter how carefully experiments might be carried out, there would still be small but finite errors associated with the distances, i.e., $d_i = \hat{d}_i + \epsilon_i$, where $\hat{d}_i$ denotes the real value of the distance and ϵ_i denotes the error. Thus, instead of a single point in the $\mathbb{D}$ space, the distances correspond to a small uncertainty region with same dimensions as the $\mathbb{D}$ space. As we have pointed out, the presence of degeneracies corresponds to a feasible region with reduced free dimension and thus has vanishing "volume" compared with the feasible region free of degeneracies, which leads us to the conclusion

that degeneracies are rare in general. However, due to the uncertainties of the measured distances, we see that the feasible regions now are "finite" in size compared with those free of degeneracies. This explains why in the reconstructions it is hard to exactly recover the target configurations, i.e., all configurations associated with the distances corresponding to the points in the feasible region should be considered with equal probability for a "fair" reconstruction procedure.

B. Decorrelation principle

Recently, Torquato and Stillinger [42] proposed a decorrelation principle concerning the disordered hard-sphere packings in high-dimensional Euclidean space $\mathbb{R}^d$. In particular, the decorrelation principle states that unconstrained spatial correlations vanish for disordered packings as the spatial dimension becomes large. In other words, as d increases, the short-ranged order beyond contact that exists in low dimensions must diminish. This principle has been explicitly observed in a variety of disordered packings in high dimensions [45].

The centroids of the hard spheres completely determine a packing, which can be considered as a point configuration in $\mathbb{R}^d$, in which there is a minimum value of pair separation distances D (i.e., the diameter of the spheres) due to the nonoverlapping condition. The decorrelation principle amounts to the following statement concerning the $\mathbb{D}$ space of the configuration: the requirement that the distances cannot be smaller than D does not affect the occurrence frequency of distances with values greater than D in very high dimensions. Note the above should be true for any disordered packings, including both dilute and jammed packings. It is known that in low dimensions, $g_2 = H(r-D)$ can only be maintained for packings with densities less than a critical value [26] and for disordered jammed packings g_2 shows strong short-ranged oscillations [46], which is the manifestation of local spatial correlations due to the nonoverlapping constraint. In other words, for the jammed disordered packings, the requirement that a desired number of distances of value D must be realized in the configuration strongly constrains the possible values of other distances in low dimensions, especially those on the same magnitude of D. In high dimensions, the above requirement becomes less significant in determining the local arrangements of points. Consider the construction used in deriving the feasibility conditions to completely determine the position of a point in $\mathbb{R}^d$, a "reference" structure containing at least d points is used. The positions of the points in the reference structure can be chosen almost freely subject to the mild constraint that no two points can be closer than D. As d increases, larger local structures (containing more points) can be constructed before the constraints on the separation distances between the points begin to play an important role. In addition, there are $(d-m)$ ways to arrange a point that has fixed distances to m ($m < d$) points in $\mathbb{R}^d$. Thus, as $d \to \infty$, the constraints on the pair-distance values imposed by the requirement that a desired number of distances with value D must be realized become insignificant, which is consistent with the decorrelation principle.

C. Additional structural information

As we have shown, pair-distance statistics in general is not sufficient information to completely determine the point configuration. A natural question is what additional information could be used to further reduce the compatible configurations associated with identical radial distribution functions. A conventional choice is the three-body correlation function g_3 [6,8], which provides information how the pair distances should be linked into triangles. Though in certain circumstances g_3 could provide additional information on the point configuration, its determination requires additional effort either theoretically or computationally.

It has been suggested in Ref. [47] that instead of incorporating information contained higher-order versions of g_2, namely, g_3, g_4, etc., one might be better served to seek other descriptors at the two-point level, which can be both manageably measured and yet reflect nontrivial higher-order structural information. One such quantity is the pair-connectedness function P_2 [2], (i.e., the connectedness contribution to g_2), which contains nontrivial topological connectedness information of the point configuration. Note the "connectedness" in a point configuration can be defined in many ways, e.g., one could circumscribe spheres around each of the points and then define that two points are connected if the two associated spheres are either contacting or overlapping, for example. Connectedness information contained in P_2 is distinct from the "triangular" information embodied in g_3, e.g., P_2 is sensitive to clustering effects, whereas g_3 is not.

D. Generalization to two-phase media

As pointed out in Sec. I, two-phase media can be constructed by decorating point configurations. For example, one can construct sphere packings by assigning to each point a sphere centered at the point with diameter equal to the minimal distance in the configuration. In this sense, two-phase media are more general than point configurations. The degeneracy of discrete point configurations implies the existence of degenerate two-phase media. The corresponding pair-distance information for two-phase media is the two-point correlation functions S_2 [2]. The degeneracy of two-phase media and the nonuniqueness issue of their reconstruction will be discussed in a sequel (Part II). Here we only provide an example of a twofold-degenerate two-phase medium constructed from the kite-trapezoid example given in Sec. I.

As shown in Fig. 9, suppose we have two large solid circles, in which small circular holes are made. One large circle contains the kite holes and the other contains the trapezoid holes. Since initially the two solid large circles are characterized by identical infinite distance set and the same subset of distances are then removed to make the holes, the remaining sets of distances for the two large circles with holes are still identical.

V. CONCLUDING REMARKS

In this paper, we discussed various aspects of the geometrical ambiguity of pair-distance statistics associated with
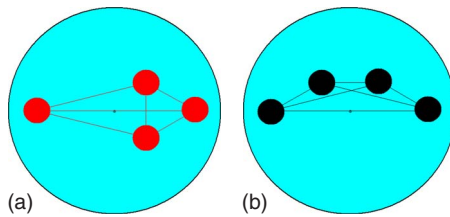
FIG. 9. (Color online) An example of twofold-degenerate continuum media based on the "kite-trapezoid" example given in Sec. I. The longest distance in the kite and trapezoid is symmetrically placed on the large circle diameter.

general point configurations in $\mathbb{R}^d$. In particular, we introduced the idea of the $\mathbb{D}$ space and derived the feasibility conditions of the distances, which are equivalent to the realizability conditions of g_2 and the necessary conditions for degeneracy. We applied the conditions to construct explicit examples of degenerate point configurations and showed that

though degeneracies are rare, one could not exclude their existence merely based on numerical reconstruction studies. We also applied the $\mathbb{D}$ space to problems in statistical physics, such as the reconstruction of atomic structures from experimentally obtained g_2, and the decorrelation principle.

As pointed out in Sec. IV D, the degeneracy of point configurations implies the existence of degenerate random media and a simple example is provided there. In a sequel to this paper [43], we will study the structural degeneracy of general random media and the nonuniqueness issue in the reconstruction of heterogeneous materials [2,3,5,48].

[1] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, New York, 2000).

[2] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer, New York, 2002).

[3] M. Sahimi, *Heterogeneous Materials* (Springer, New York, 2003).

[4] S. F. Edwards, in *Granular Matter*, edited by A. Metha (Springer, New York, 2003).

[5] T. I. Zohdi, Mech. Mater. **38**, 969 (2006).

[6] S. A. Rice and P. Gray, *The Statistical Mechanics of Simple Liquids* (John & Wiley, New York, 1966).

[7] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).

[8] J. P. Hansen and I. R. McDonald, *Theory of Simple of Liquids* (Academic, New York, 2006).

[9] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups* (Springer, New York, 1987).

[10] P. J. E. Peebles, *Principles of Physical Cosmology* (Princeton University Press, Princeton, 1993).

[11] A. Gabrielli, F. Sylos Labini, M. Joyce, and P. Pietronero, *Statistical Physics for Cosmic Structures* (Springer, New York, 2005).

[12] A. Pommerening and D. Stoyan, Can. J. Res. **38**, 1110 (2008).

[13] J. L. Gevertz and S. Torquato, PLOS Comput. Biol. **4**, e1000152 (2008).

[14] T. M. Truskett, S. Torquato, and P. G. Debenedetti, Phys. Rev. E **62**, 993 (2000); S. Torquato and F. H. Stillinger, *ibid.* **68**, 041113 (2003).

[15] C. L. Y. Yeong and S. Torquato, Phys. Rev. E **57**, 495 (1998).

[16] C. L. Y. Yeong and S. Torquato, Phys. Rev. E **58**, 224 (1998).

[17] N. Sheehan and S. Torquato, J. Appl. Phys. **89**, 53 (2001).

[18] M. G. Rozman and M. Utz, Phys. Rev. Lett. **89**, 135501 (2002).

[19] K. Wu, M. I. J. Dijke, G. D. Couples, Z. Jiang, J. Ma, K. S. Sorbie, J. Crawford, I. Young, and X. Zhang, Transp. Porous

Media **65**, 443 (2006).

[20] H. Kumar, C. L. Briant, and W. A. Curtin, Mech. Mater. **38**, 818 (2006).

[21] Y. Jiao, F. H. Stillinger, and S. Torquato, Phys. Rev. E **76**, 031110 (2007).

[22] Y. Jiao, F. H. Stillinger, and S. Torquato, Phys. Rev. E **77**, 031135 (2008).

[23] D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, Acta Mater. **56**, 942 (2008).

[24] M. D. Rintoul and S. Torquato, J. Colloid Interface Sci. **186**, 467 (1997).

[25] F. H. Stillinger, S. Torquato, J. M. Eroles, and T. M. Truskett, J. Phys. Chem. B **105**, 6592 (2001); F. H. Stillinger and S. Torquato, Mol. Phys. **103**, 2943 (2005).

[26] S. Torquato and F. H. Stillinger, J. Phys. Chem. B **106**, 8354 (2002); J. Crawford, S. Torquato, and F. H. Stillinger, J. Chem. Phys. **119**, 7065 (2003); O. U. Uche, F. H. Stillinger, and S. Torquato, Physica A **360**, 21 (2006).

[27] R. D. Batten, F. H. Stillinger, and S. Torquato, J. Appl. Phys. **104**, 033504 (2008).

[28] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995), and references therein.

[29] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1756 (1964).

[30] C. Garrod, M. V. Mihailovic, and M. Rosina, J. Math. Phys. **16**, 868 (1975).

[31] M. Rosina and C. Carrod, J. Comput. Phys. **18**, 300 (1975).

[32] R. M. Erdahl, C. Garrod, B. Golli, and M. Rosina, J. Math. Phys. **20**, 1366 (1979).

[33] For Bravais lattices in two and three dimensions, it is known that pair-distance information uniquely determines the lattice, see A. Schiemann, Math. Ann. **308**, 507 (1997).

[34] A. L. Patterson, Phys. Rev. **65**, 195 (1944).

[35] R. Hosemann and S. N. Bagchi, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **7**, 237 (1954).

[36] M. Boutin and G. Kemper, Adv. Appl. Math. **32**, 709 (2004).

[37] D. K. Stillinger, F. H. Stillinger, S. Torquato, T. M. Truskett,

and P. G. Debenedetti, J. Stat. Phys. **100**, 49 (2000); F. H. Stillinger, D. K. Stillinger, S. Torquato, T. M. Truskett, and P. G. Debenedetti, J. Chem. Phys. **113**, 10186 (2000).

[38] J. G. Kirkwood, J. Chem. Phys. **10**, 394 (1942).

[39] R. Bellman, J. ACM **9**, 61 (1962).

[40] R. V. Ambartsumian and H. S. Sukiasian, Acta Appl. Math. **22**, 15 (1991).

[41] T. Kuna, J. L. Lebowitz, and E. R. Speer, J. Stat. Phys. **129**, 417 (2007); T. Kuna, J. Lebowitz, and E. Speer, e-print arXiv:0910.1710.

[42] S. Torquato and F. H. Stillinger, Exp. Math. **15**, 307 (2006).

[43] Y. Jiao, F. H. Stillinger, and S. Torquato (unpublished).

[44] D. M. Y. Sommerville, *An Introduction to the Geometry of n Dimensions* (Dover, New York, 1958).

[45] S. Torquato and F. H. Stillinger, Phys. Rev. E **73**, 031106 (2006); M. Skoge, A. Donev, F. H. Stillinger, and S. Torquato, *ibid.* **74**, 041127 (2006); S. Torquato, O. U. Uche, and F. H. Stillinger, *ibid.* **74**, 061308 (2006).

[46] A. Donev, S. Torquato, and F. H. Stillinger, Phys. Rev. E **71**, 011105 (2005).

[47] Y. Jiao, F. H. Stillinger, and S. Torquato, Proc. Natl. Acad. Sci. U.S.A. **106**, 17634 (2009).

[48] S. Torquato, Int. J. Solids Struct. **37**, 411 (2000).