On the number of solutions of the discretizable molecular distance geometry problem

LEO LIBERTI[1], BENOÎT MASSON[2], JON LEE [3], CARLILE LAVOR [4], AND ANTONIO MUCHERINO[5]

[1] LIX, École Polytechnique, 91128 Palaiseau, France
liberti@lix.polytechnique.fr
[2] IRISA, INRIA, Campus de Beaulieu, 35042 Rennes, France
benoit.masson@inria.fr
[3] Dept. of Mathematical Sciences, IBM T.J. Watson Research Center, PO Box 218, Yorktown Heights, NY 10598, USA, jonlee@us.ibm.com
[4] Department of Applied Mathematics (IMECC-UNICAMP), State University of Campinas, C.P. 6065, 13081-970, Campinas - SP, Brazil, clavor@ime.unicamp.br
[5] CERFACS, Toulouse, France mucherino@cerfacs.fr

Abstract. The Discretizable Molecular Distance Geometry Problem is a subset of instances of the distance geometry problem that can be solved by a combinatorial algorithm called "Branch-and-Prune". It was observed empirically that the number of solutions of YES instances is always a power of two. We perform an etensive theoretical analysis of the number of solutions for these instances and we prove that this number is a power of two with probability one.
Keywords: distance geometry, symmetry, Branch-and-Prune, power of two.

1 Introduction

We consider the following problem arising in the analysis of Nuclear Magnetic Resonance (NMR) data for general molecules.

MOLECULAR DISTANCE GEOMETRY PROBLEM (MDGP).
Given a simple undirected graph $G = (V, E)$ and a function $d : E \to \mathbb{R}$, decide whether there is an embedding $x : V \to \mathbb{R}^3$ such that

$$\forall \{u, v\} \in E \quad (||x_u - x_v|| = d_{uv}) \tag{1}$$

The MDGP is a mixed-combinatorial optimization problem; it can be cast as the global optimization problem $\min \sum_{\{u,v\}\in E}(||x_u - x_v||^2 - d_{uv}^2)^2$ in continuous variables, which is generally solved using continuous search techniques [1, 2]. The generalization of the MDGP to arbitrary dimensions asks for an embedding of G in $\mathbb{R}^K$ satisfying (1) and is called the DISTANCE GEOMETRY PROBLEM (DGP). The DGP is strongly **NP**-hard [3]; it is related to the Euclidean Distance Matrix Completion Problem (EDMCP) [4] (whose complexity status is currently unknown), the difference being that in the EDMCP the dimension K of the embedding space is part of the output rather than part of the input.

Finding a Euclidean embedding of a weighted graph has two main applications: to molecular conformation [5] and to sensor networks [6, 7]. The results of this paper were inspired by the application to the conformation of proteins: in particular, chemical analysis and NMR experiments can help identify a subset of inter-atomic distances [8]. The motivation is that the function of a protein is determined by its 3D structure [9]. Since proteins are a strict subset of molecules, it makes sense to ask whether there the restriction of the MDGP to proteins might yield more efficient methods than those developed for the MDGP applied to general molecules. In 2005 two of the authors of this paper (CL and LL) started working on a discrete algorithm which exploits two observations: (i) proteins are organized in a *backbone* and some *side chains*, which can be embedded separately, once the backbone embedding is known [10]; (ii) the distances between any atom of the backbones, seen as as a total order on the set of atoms, to its three immediate predecessors are generally known (and by applying certain technical devices to the order can be assumed to be precise [11]). This algorithm, called Branch-and-Prune (BP), is based on the *Sphere Intersection Property* (SIP): the intersection of K spheres in $\mathbb{R}^K$ generally consists of either 0 or 2 points. Here the term *generally* has a definite significance: it means that the set K-tuples of spheres for which the SIP does not hold has Lebesgue measure 0 in the set of all possible K-tuples of spheres.

In the following, we identify atoms with the set V of vertices of a given graph G, whose edge set E includes the pairs of atom for which a distance is known. The weight of each edge $\{u, v\} \in E$ is the value of the distance d_{uv}, and an order on the vertices (the backbone order in the case of proteins) is given. BP exploits the SIP by performing a binary search in the space of embeddings: under the hypothesis that *for each vertex of rank $> K$ in the order, the distances to its K immediate predecessors are known*, the BP places a vertex v in both of the positions guaranteed by the SIP, verifies whether these are compatible with the distances to *all* adjacent predecessors of v, and then accordingly recurses the search to the successor of v. This yields a worst-case exponential behaviour, occurring when the set of adjacent predecessors of each vertex v is equal to the set of its K immediate predecessors. In practice, however, the BP outperforms its continuous search competitors in both efficiency and reliability [12]. One particularly useful feature of BP is that, because the search is complete, it finds the set X of all incongruent embeddings for a given graph. In a sequence of papers (the main ones being [13, 12, 14–18]) we developed this idea in a number of directions. In particular, we defined a new optimization problem, the Discretizable MDGP (DMDGP) [12] as the class of all DGP instances that satisfy the conditions required by the BP: the existence of a vertex order such that the K immediate predecessors of each vertex v of rank $> K$ are adjacent to v in G, and the fact that d satisfies strict simplex inequalities [19, 15].

In all our computational tests on DMDGP instances, we observed that the number of incongruent embeddings is a power of two: this comes to no surprise in the exponential worst case mentioned above, but there is no apparent reason why this should be the case when adjacent predecessors also include other vertices than the K immediate predecessors (and, indeed, in Sect. 6 we exhibit a set of counterexamples to the conjecture that for all YES instances of the DMDGP $\exists h \in \mathbb{N}\ (|X| = 2^h)$). Yet, the computational trend remained unexplained. The contribution of this paper is a proof that the set of YES instances of the DMDGP such that $|X|$ is a power of two has Lebesgue measure 1 in the set of all YES instances of the DMDGP. The statement is based on the assumption that we consider solutions (i.e. graph embeddings) whose components range in the uncountable set $\mathbb{R}^K$. Our result is nontrivial, and accordingly the proof, which consists of several lemmata, propositions and theorems, is long, technical and difficult: because of the page limit, all proofs are in the appendix. The result is nonetheless very important insofar as it explains the behaviour of a practically useful solution method.

The rest of this paper is organized as follows. We give a formal description of the DMDGP in arbitrary dimensions (Sect. 2) and of the BP algorithm and some of its theoretical properties (Sect. 3); we then study some geometrical aspects of the BP tree (Sect. 4), and prove that the number of solutions of YES instances of the DMDGP is a power of two with probability one (Sect. 5). We exhibit a (zero measure) family of counterexamples to the "power of two" conjecture in Sect. 6.

2 The formal definition of the Discretizable Molecular Distance Geometry Problem

For a set $U = \{x_i \in \mathbb{R}^K \mid i \le K+1\}$ of points in $\mathbb{R}^K$, let D be the symmetric matrix whose (i,j)-th component is $\|x_i - x_j\|^2$ for all $i, j \le K+1$ and let D' be D bordered by a left $(0,1,\dots,1)^\top$ column and a top $(0,1,\dots,1)$ row (both of size $K+2$). Then the Cayley-Menger formula states that the volume $\Delta_K(U)$ of the K-simplex on U is given by $\Delta_K(U) = \sqrt{\frac{(-1)^{K+1}}{2^K(K!)^2}|D'|}$. The strict simplex inequalities are given by $\Delta_K(U) > 0$. For $K = 3$, these reduce to strict triangle inequalities. We remark that only the distances of the simplex edges are necessary to compute $\Delta_K(U)$, rather than the actual points in U; the needed information can be encoded as a complete graph $\mathbf{K}_{K+1}$ on $K+1$ vertices with edge weights as the distances.

Let $n = |V|$ and $m = |E|$. For all $v \in V$, let $N(v) = \{u \in V \mid \{u, v\} \in E\}$ be the star of vertices around v (also called the adjacencies of v); for a directed graphs (V, A), where $A \subseteq V \times V$, we denote the outgoing star by $N^+(v) = \{u \in V \mid (v, u) \in A\}$. For an order $<$ on V, let $\gamma(v) = \{u \in V \mid u < v\}$ be the set of predecessors of v, and let $\rho(v) = |\gamma(v)| + 1$ be the rank of v in $<$. For $V' \subseteq V$, we denote by $G[V']$ the subgraph of G induced by V'. For a finite set M, let $\mathcal{P}(M)$ be its power set. We call an embedding x of G *valid* if (1) holds for G. For a sequence $x = (x_1, \dots, x_n)$ and a subset $U \subseteq \{1, \dots, n\}$ we let $x[U]$ be the subsequence of x indexed by U. If x is an initial subsequence of y, then y is an *extension* of x. For each $v \in V$ with $\rho(v) > K$ we let U_v be the set of the K immediate predecessors of v, and remark that $U_v \subseteq N(v) \cap \gamma(v)$.

THE GENERALIZED DMDGP. Given an undirected graph $G = (V, E)$, an edge weight function $d : E \to \mathbb{R}_+$, an integer $K > 0$, a subset $V_0 \subseteq V$ with $|V_0| = K$, a partial embedding $\bar{x} : V_0 \to \mathbb{R}^K$ valid for $G[V_0]$, and a total order $<$ on V such that:

$$\{v \in V \mid \rho(v) \leq K\} = V_0; \tag{2}$$

$$\forall v \in V \quad (\rho(v) > K \to |N(v) \cap \gamma(v)| \geq K); \tag{3}$$

$$\forall v \in V \smallsetminus V_0 \; (G[U_v] = \mathbf{K}_K \wedge \Delta_{K-1}(U_v) > 0), \tag{4}$$

decide whether there is a valid extension $x : V \to \mathbb{R}^K$ of $\bar{x}$.

Conditions (2-4) allow the search for the Euclidean position of vertex v to only depend on the K vertices of rank preceding $\rho(v)$, as x_v is the intersection of at least K spheres centered at x_u and with radius d_{uv} for all $u \in N(v) \cap \gamma(v)$. This, in particular, implies that the predecessors of v are placed before v, so that all of the distances between all predecessors are known when placing v. Thus, we can also solve instances for which $G[U_v]$ is not the full K-clique, although they are not formally in the generalized DMDGP.

We remark that the SIP is independent of U_v, so that we could simply replace U_v with any subset of $N(v) \cap \gamma(v)$ with cardinality K. This actually yields a larger instance set called DISCRETIZABLE DISTANCE GEOMETRY PROBLEM (DDGP), or DDGP_K if K is fixed and not part of the input, discussed in [14]. We shall see, however, that the assumption that U_v contains the K *immediate* predecessors of v will be crucial in the following (this, by the way, also explains why the generalized DMDGP is not called "DDGP" in analogy with MDGP→DGP). In the rest of the paper we use the acronym DMDGP to actually mean the *generalized* DMDGP, and we use the name DMDGP_3 to name the original DMDGP in $\mathbb{R}^3$. Complexity-wise, a polynomial reduction from SUBSET-SUM to the DMDGP_3 [12] shows that the DMDGP is **NP**-hard.

3 Sphere intersections and reflections

The BP algorithm for the DMDGP_3, presented in [13], can easily be extended to the DMDGP. As mentioned above, once the vertices of U_v have been embedded in $\mathbb{R}^K$, the known distances from vertices in U_v to a given v will enforce the position of v as the intersection of K spheres. Under strict simplex inequalities, this intersection consists of at most two distinct points. The BP exploits this fact to recursively generate a binary search tree of height at most n where a node at level i represents a possible placement in $\mathbb{R}^K$ of the vertex of G with rank i in $<$. Paths of length n correspond to valid embeddings.

Let G be a DMDGP instance. Consider $v \in V$ with rank $\rho(v) = i > K$, let $G^v = G[\gamma(v) \cup \{v\}]$ and x be a valid embedding of $G[\gamma(v)]$. We characterize the number of extensions of x valid for G^v in the following lemmata (which also hold for the DDGP). Lemmata 3.1 and 3.2 essentially state that $G[\{v\} \cup (N(v) \cap \gamma(v))]$ are rigid and, respectively, uniquely rigid graphs.

In the following, we assume that the probability of any point of $\mathbb{R}^K$ belonging to any given subset of $\mathbb{R}^K$ having Lebesgue measure zero is equal to zero. Based on this assumption, when we state "$(\forall p \in P \ F(p))$ with probability 1" for a certain well-formed formula F with a free variable ranging over an uncountable set P, we really mean that there exists a Lebesgue measurable subset $Q \subseteq P$ with Lebesgue measure 1 in P such that $\forall q \in Q \ F(p)$. For example, the statement of Lemma 3.1 should be read as follows: the set of DMDGP instances and partial embeddings x for which the result does not hold has Lebesgue measure 0 in the set of all DMDGP instances and partial embeddings. We remark that this is different from the usual genericity notion employed in rigidity theory [20], which requires distances to be algebraically independent over $\mathbb{Z}$. Since our instances come from experimental measurements over existing structures, the distances may not be independent. One consequence is the validity of Lemma 3.2, which would not hold with the stronger genericity requirement (the intersection of $K+1$ "generic spheres" in $\mathbb{R}^K$ is empty).

Lemma 3.1. *If $|N(v) \cap \gamma(v)| = K$ then there are at most two distinct extensions of x that are valid for G^v. If one valid extension exists, then with probability 1 there are exactly two distinct valid extensions.*

Lemma 3.2. *If $|N(v) \cap \gamma(v)| > K$ then, with probability* 1*, there is at most one extension of x.*

Lemma 3.3. *With the notation of Lemma 3.1, if $\bar{x}$ is a valid embedding for $G[U_v]$, then z'' is a reflection of z' with respect to the hyperplane through the K points of $\bar{x}$.*

Reflections with respect to hyperplanes are isometries, and can therefore be represented by linear operators. If $a \in \mathbb{R}^K$ is the unit normal vector to a hyperplane H containing the origin, then the reflection operator R_0 w.r.t. H can be expressed in function of the standard basis by the matrix $I - 2aa^\top$, where I is the $K \times K$ identity matrix [21]. Let H be a hyperplane with equation $a^\top x = a_0$ (with $a_0 \neq 0$) and a_i, for some $1 \leq i \leq K$, be the nonzero coefficient of smallest index in a. Then, the reflection operator R acting on a point $p \in \mathbb{R}^K$ w.r.t. H is given by $R(p) = R_0(p - \frac{a_0}{a_i} e_i) + \frac{a_0}{a_i} e_i$, where $e_i \in \mathbb{R}^K$ is the unit vector with 1 at index i and 0 elsewhere: we first we translate p so that we can reflect it using R_0 w.r.t. the translation of H containing the origin, then we perform the inverse translation of the reflection.

3.1 Branch-and-Prune

A formal description of the BP algorithm for the DMDGP is given in Alg. 1. It builds a binary search tree $\mathcal{T} = (\mathcal{V}, \mathcal{A})$, directed from the root to the leaves, whose nodes are triplets $\alpha = (x(\alpha), \lambda(\alpha), \mu(\alpha))$. For $\alpha \in \mathcal{T}$ we denote by $\mathsf{p}(\alpha)$ the unique path from the root node r of $\mathcal{T}$ to α; $x(\alpha)$ is an extension of the embedding x^- found on $\mathsf{p}(\alpha^-)$, where α^- is the unique parent node of α. The symbol $\lambda(\alpha) \in \{0,1\}$ distinguishes whether α is a "left" or a "right" subnode of α^-. More precisely, let α be a node at level i in $\mathcal{T}$, $v = \rho^{-1}(i)$, $\bar{x}$ be a partial embedding of $G[U_v]$, and $a_v^\top x = a_{v0}$ be the equation of the $((K-1)$-dimensional by (4)) hyperplane through the points of $\bar{x}$. Assuming $u = \rho^{-1}(i-1)$, $a_v \in \mathbb{R}^K$ is oriented so that $a_v \cdot a_u \geq 0$; then:

$$\lambda(\alpha) = \begin{cases} 0 & \text{if } a_v^\top x(\alpha)_i \leq a_{v0} \\ 1 & \text{if } a_v^\top x(\alpha)_i > a_{v0}. \end{cases} \tag{5}$$

Lastly, $\mu(\alpha) = \boxplus$ if x is a valid extension of x^-, in which case the node is said to be *feasible*, and $\mu = \boxminus$ otherwise. This allows us to retrieve the set X of all valid embeddings of G by simply traversing $\mathcal{T}$ backwards from the leaf nodes marked $\boxplus$ up to r.

We remark that Alg. 1 differs from the original BP formulation [13] because it applies to K dimensions and explicitly stores several details of the binary search tree.

Lemma 3.4. *At termination of Alg. 1, X contains all valid embeddings of G extending $\bar{x}$.*

We now partition $\mathcal{V}$ in pairwise disjoint subsets $\mathcal{V}_1, \ldots, \mathcal{V}_n$ where for all $i \leq n$ the set $\mathcal{V}_i$ contains all the nodes of $\mathcal{V}$ at level i of the tree $\mathcal{T}$.

Proposition 3.5. *With probability 1, there is no level $i \leq n$ having two distinct feasible nodes $\beta, \theta \in \mathcal{V}_i$ such that $|\{\alpha \in N^+(\beta) \mid \mu(\alpha) = \boxplus\}| = 1$ and $|\{\alpha \in N^+(\theta) \mid \mu(\alpha) = \boxplus\}| = 2$.*

We remark that Prop. 3.5 also holds for the DDGP provided U_v is chosen in Alg. 1 as any subset of $N(v) \cap \gamma(v)$ satisfying the constraints of Eq. (4).

4 Geometry in BP Trees

The most important result of this section is that, for any valid embedding $y \in X$, if the BP tree branches at level $i = \rho(v)$ on the path to y and both branches continue to the last level, then the embedding obtained by reflecting all the points of y past the $(i-1)$-th vertex through the hyperplane defined by $y[U_v]$ is also valid with probability 1. We remark that the results in this section only apply to the DMDGP (not to the DDGP, as shown in the counterexample of Fig. 3).

We need to emphasize those BP branchings which carry on to feasible leaf nodes along both branches. For $y \in X$ and a vertex $v \in V \smallsetminus V_0$ we denote $\Upsilon(y, v)$ the following property:

> $\Upsilon(y, v)$: *there are feasible leaf nodes $\beta, \beta' \in \mathcal{V}_n$ such that $x(\beta) = y$, $\mathsf{p}(\beta) \cap \mathcal{V}_{\rho(v)-1} = \mathsf{p}(\beta') \cap \mathcal{V}_{\rho(v)-1}$ and $\mathsf{p}(\beta) \cap \mathcal{V}_{\rho(v)} \neq \mathsf{p}(\beta') \cap \mathcal{V}_{\rho(v)}$.*

Algorithm 1 The Branch and Prune algorithm.

Require: Partial embedding $\bar{x}$ of first K vertices of G
Ensure: Set X of valid embeddings of G

Let $\alpha = (\bar{x}_1, 0, \boxplus)$ and $\alpha' = (\bar{x}_1, 1, \boxminus)$
Initialize $\mathcal{V} = \{\alpha, \alpha'\}$ and $\mathcal{A} = \{(\mathsf{r}, \alpha), (\mathsf{r}, \alpha')\}$
for $1 < i \le K$ **do**
    Let $\alpha = (\bar{x}_i, 0, \boxplus)$, $\alpha' = (\bar{x}_i, 1, \boxminus)$, $\beta = (\bar{x}_{i-1}, 0, \boxplus)$
    Let $\mathcal{V} \leftarrow \mathcal{V} \cup \{\alpha, \alpha'\}$ and $\mathcal{A} \leftarrow \mathcal{A} \cup \{(\beta, \alpha), (\beta, \alpha')\}$
end for
BranchAndPrune($K+1$, $(\bar{x}_K, 0, \boxplus)$)
Let $X = \{x(\theta) \mid \theta \in \mathcal{V} \wedge |N^+(\theta)| = 0 \wedge \mu(\theta) = \boxplus\}$
stop

function BranchAndPrune(i, β):
if $i > n \vee \mu = \boxminus$ **then**
    return
end if
Let $v = \rho^{-1}(i)$
Compute the equation $a_v^\top x = a_{v0}$ of the hyperplane through $x[U_v]$
Let $Z = \{z', z''\}$ be extensions of $x(\beta)$ to v, and $Z' = Z$
for $z \in Z$ **do**
    if $\exists \{u, v\} \in E \; \|x(\beta)_u - z\| \ne d_{uv}$ **then**
        Let $Z = Z \smallsetminus \{z\}$
    end if
end for
if $Z = \{z', z''\}$ **then**
    if $a_v^\top z' \le a_{v0}$ **then**
        Let $\alpha = (z', 0, \boxplus)$, $\alpha' = (z'', 1, \boxplus)$
    else
        Let $\alpha = (z'', 0, \boxplus)$, $\alpha' = (z', 1, \boxplus)$
    end if
else if $Z = \{z\}$ **then**
    if $a_v^\top z \le a_{v0}$ **then**
        Let $\alpha = (z, 0, \boxplus)$, $\alpha' = (Z' \smallsetminus \{z\}, 1, \boxminus)$
    else
        Let $\alpha = (z, 1, \boxplus)$, $\alpha' = (Z' \smallsetminus \{z\}, 0, \boxminus)$
    end if
else
    return
end if
Let $\mathcal{V} \leftarrow \mathcal{V} \cup \{\alpha, \alpha'\}$ and $\mathcal{A} \leftarrow \mathcal{A} \cup \{(\beta, \alpha), (\beta, \alpha')\}$
for $\theta \in N^+(\beta)$ such that $\mu(\theta) = \boxplus$ **do**
    BranchAndPrune($i+1$, θ)
end for
return

If $\Upsilon(y,v)$ holds, it is easy to show that $\mathsf{p}(\beta)\cap\mathcal{V}_{\rho(v)-1}$ contains a single feasible node with two feasible subnodes. With $\Upsilon(y,v)$ true, we let R^v be the Euclidean reflection operator with respect to the hyperplane through $y[U_v]$ (as discussed in p. 5). Define $\tilde{R}^v = I^{\rho(v)-1}\times(R^v)^{n-\rho(v)}$, i.e. $\tilde{R}^v y = (y_1,\ldots,y_{i-1},R^v y_i,\ldots,R^v y_n)$. This is a *partial reflection* of y which only acts on vertices past rank $i-1$.

We emphasize that for all $\ell\in\{i,\ldots,n\}$ and for all $\alpha\in\mathcal{V}_\ell$ the set $\mathsf{p}(\alpha)\cap\mathcal{V}_i$ has a unique element, as it contains the unique node at level i on the path from α to the BP tree root node.

The following is a corollary to Lemma 3.3.

Corollary 4.1. *Let $\alpha\in\mathcal{V}_{i-1}$ for some $i>1$, $v=\rho^{-1}(i)$ and $N^+(\alpha)=\{\eta,\beta\}$ with $\mu(\eta)=\mu(\beta)=\boxplus$. Then $x(\eta)_v = R^v x(\beta)_v$.*

Remark 4.2. If $\Upsilon(y,v)$ holds for some $y\in X$ and $v\in V\smallsetminus V_0$, then by definition there are feasible leaf nodes in the BP tree, which implies that the considered DMDGP instance is YES.

An important consequence of Remark 4.2 is that all statements assuming $\Upsilon(y,v)$ and claiming a result with probability 1 implicitly also assume that the probability is conditional to the event of the DMDGP instance being a YES one. In particular, since the instance is YES, certain points *must* be placed at certain distances with probability 1, for otherwise the instance would be NO. This is evident in Prop. 4.4, Cor. 4.6, Cor. 4.7, and Thm. 4.9, where we state that certain real scalars and vectors must belong to certain finite sets with probability 1: the sense of these assertions, in this context, is that the Lebesgue measure of the set of YES instances not satisfying the result is zero in the set of all YES instances.

Lemma 4.3. *Let $\alpha\in\mathcal{V}_{i-1}$ for some $i>1$ such that $N^+(\alpha)=\{\eta',\beta'\}$, $u=\rho^{-1}(i)$; $v>u$ with $\rho(v)=\ell$, and consider two feasible nodes $\eta,\beta\in\mathcal{V}_\ell$ such that $\eta'=\mathsf{p}(\eta)\cap\mathcal{V}_i$ and $\beta'=\mathsf{p}(\beta)\cap\mathcal{V}_i$. Then, with probability 1, the following statements are equivalent:*

(i) $\forall\ i\le j\le\ell$, $x(\beta'')_w = R^u x(\eta'')_w$, *where* $\eta''=\mathsf{p}(\eta)\cap\mathcal{V}_j$, $\beta''=\mathsf{p}(\beta)\cap\mathcal{V}_j$, *and* $w=\rho^{-1}(j)$;
(ii) $\forall\ i\le j\le\ell$, $\lambda(\eta'')=1-\lambda(\beta'')$, *with* $\eta''=\mathsf{p}(\eta)\cap\mathcal{V}_j$ *and* $\beta''=\mathsf{p}(\beta)\cap\mathcal{V}_j$.

Proposition 4.4. *Consider a subtree $\mathcal{T}'$ of $\mathcal{T}$ consisting of $K+2$ consecutive levels $i-K-1,\ldots,i$ (where $i\ge 2K+1$), rooted at a single node η and such that all nodes at all levels are marked $\boxplus$. Let $p=2^{K+1}$ and consider the set $Y'=\{y_j\mid j\le p\}$ of partial embeddings of G at the leaf nodes $\{\alpha_j\mid j\le p\}$ of $\mathcal{T}'$. Let $u=\rho^{-1}(i-K-1)$ and $v=\rho^{-1}(i)$. Then with probability 1 there are two distinct positive reals r,r' such that $\|y_j(\alpha_j)_u - y_j(\alpha_j)_v\|\in\{r,r'\}$ for all $j\le p$.*

Fig. 1 shows a graphical proof sketch of Prop. 4.4 for $K=2$. Prop. 4.4 is useful in order to show that certain configurations of nodes within $\mathcal{T}$ can only occur with probability 0.

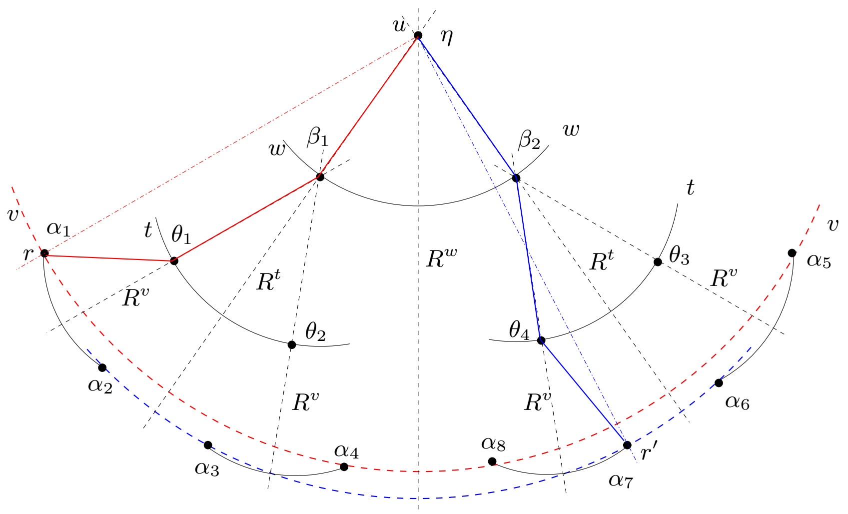


Fig. 1. Proof of Prop. 4.4 in $\mathbb{R}^2$. The arrangement of three segments gives rise, in general, to two distances r, r' between root and leaves.

Example 4.5. Consider a subtree $\mathcal{T}'$ of $\mathcal{T}$ like the one in Fig. 1 embedded in $\mathbb{R}^2$, and suppose that all nodes at level u, w, t are marked $\boxplus$, and further that only one node within α_1, α_2 is feasible, only one node within α_3, α_4 is feasible, only one node within α_7, α_8 is feasible, and α_5, α_6 are both infeasible. This must be due to a distance $d_{u'v}$ with $u' \leq u$. Consider now a circle C completely determined by its center at $y_1(\alpha_1)_{u'}$ and its radius $d_{u'v}$; if C also contains the points at the nodes $\alpha_1, \alpha_4, \alpha_8$ or the points at the nodes $\alpha_2, \alpha_3, \alpha_7$ then we must have $u' = u$, in which case also one of α_5, α_6 will be feasible (against the hypothesis). And the probability that C should contain the points at the nodes $\alpha_1, \alpha_3, \alpha_8$ or $\alpha_2, \alpha_4, \alpha_7$ is zero. Hence $\mathcal{T}'$ can only occur with probability 0. □

We now exploit a generalization of Prop. 4.4 to build up towards the main result of this section, i.e. that partial reflections map valid embeddings to valid embeddings (Thm. 4.9).

Corollary 4.6. *Consider a subtree $\mathcal{T}'$ of $\mathcal{T}$ consisting of $K+q+1$ consecutive levels $i-K-q, \ldots, i$ (where $i \geq 2K+q$ and $q \geq 1$), rooted at a single node η and such that all nodes at all levels are marked $\boxplus$. Let $p = 2^{K+q}$ and consider the set $Y' = \{y_j \mid j \leq p\}$ of partial embeddings of G at the leaf nodes $\{\alpha_j \mid j \leq p\}$ of $\mathcal{T}'$. Let $u = \rho^{-1}(i-K-q)$ and $v = \rho^{-1}(i)$. Then with probability 1 there is a set $H^{uv} = \{r_j \mid j \leq 2^q\}$ of 2^q distinct positive reals such that $\|y_i(\alpha_i)_u - y_i(\alpha_i)_v\| \in H^{uv}$ for all $i \leq p$.*

The next corollary shows that distances spanning more than K vertices must all belong to certain finite sets of values for YES instances.

Corollary 4.7. *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. If $\{u,w\} \in E$ with $u < v < w$ and $\rho(w) - \rho(u) > K$ then $d_{uw} \in H^{uw}$ with probability 1.*

Corollary 4.8. *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. If $u \in V$ with $u > v$ then $R^v y_u$ belongs to a valid extension of $y[U_v]$.*

Finally, we state the main result of the section: if a DMDGP instance has a valid embedding y and v is a vertex where a "valid branching" (in the sense of the $\Upsilon(y,v)$ assumption) takes place in the BP algorithm, then the partial reflection of y with respect to v is also a valid embedding. We remark that the $\Upsilon(y,v)$ assumption only says that at v there is a BP search tree branching one of whose branch eventually leads to y, whilst the other ends up at any other valid embedding. Thm. 4.9 states that in this case the partial reflection of y w.r.t. v is also valid.

Theorem 4.9. *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. Then $\tilde{R}^v y \in X$ with probability 1.*

5 Symmetry and Number of Solutions

Our strategy for proving that YES instances of the DMDGP have power of two solutions with probability 1 is as follows. We map embeddings $y \in X$ to binary sequences $\chi \in \{0,1\}^n$ describing the "branching path" in the tree $\mathcal{T}$. We define a symmetry operation on χ by flipping its tail from a given component i (this operation is akin to branching at level i). We show that the cardinality of the group of all such symmetries is a power of two by bijection with a set of binary sequences. Finally we prove that the cardinality of the symmetry group is the same as $|X|$.

For all leaf nodes $\alpha \in \mathcal{V}$ with $\mu(\alpha) = \boxplus$ let $\chi(\alpha) = (\lambda(\beta) \mid \beta \in \mathsf{p}(\alpha))$; since embeddings in X are also in correspondence with leaf $\boxplus$-nodes of $\mathcal{T}$ by Alg. 1, Step 8, χ defines a relation on $X \times \{0,1\}^n$.

Lemma 5.1. *With probability 1, the relation χ is a function.*

Let $\Xi = \{\chi(y) \mid y \in X\}$. For $y \in X$ let y^i be its subsequence $(x_1, \ldots, x_i)$. We extend χ to be defined on all such subsequences by simply setting $\chi^i = (\chi(y)_1, \ldots, \chi(y)_i)$; $\chi(y)$ is valid if y is a valid embedding.

Let $N = \{1, \ldots, n\}$ and g be the $n \times n$ binary matrix such that $g_{ij} = 1$ if $i \leq j$ and 0 otherwise (the upper triangular $n \times n$ all-1 matrix); let g_i be its i-th row vector and $\Gamma = \{g_i \mid i \in N\}$. Consider the elementwise modulo-2 addition in the set $\mathbb{F}_2^n$ (denoted $\oplus$): this endows $\mathbb{F}_2^n$ with an additive group structure with identity $e = (0, \ldots, 0)$ where each element is idempotent. Thus, $\mathcal{G} = (\mathbb{F}_2^n, \oplus) \cong C_2^n$. This group naturally acts on itself (and subsets thereof) using the same $\oplus$ operation. It is not difficult to prove that Γ is a set of group generators for $\mathcal{G}$ and a linearly independent set of the vector space $\mathcal{V}$ given by $\mathcal{G}$ with scalar multiplication over $\mathbb{F}_2$. For all $S \subseteq N$, let

$$g_S = \bigoplus_{i \in S} g_i,$$

and define a mapping $\phi : \mathcal{P}(N) \to \mathcal{G}$ given by $\phi(S) = g_S$.

Lemma 5.2. *ϕ is injective.*

The following result shows essentially that groups of partial reflections have power of two cardinality.

Lemma 5.3. *For all $H \subseteq \Gamma$, $|\langle H \rangle| = 2^{|H|}$.*

Let I be the set of levels of $\mathcal{T}$ for which from all nodes with two valid children there is a path going to a feasible leaf through both children. Let $L = \{g_i \in \Gamma \mid i \in I\}$ and $\Lambda = \langle L \rangle$ be the subgroup of $\mathcal{G}$ of "allowed partial reflections" generated by L. In the following (the main result of this section) we relate partial reflections to χ representations of valid embeddings. We show that any valid embedding, in its χ representation, generates the whole set of valid embeddings by means of the action of the group of allowed partial reflections.

Theorem 5.4. *If $\Xi \neq \emptyset$, for all $\xi \in \Xi$ we have $\xi \oplus \Lambda = \Xi$ with probability 1.*

The main result of the paper is now simply a corollary of Thm. 5.4.

Corollary 5.5. *If a DMDGP instance is YES, $|X|$ is a power of two with probability 1.*

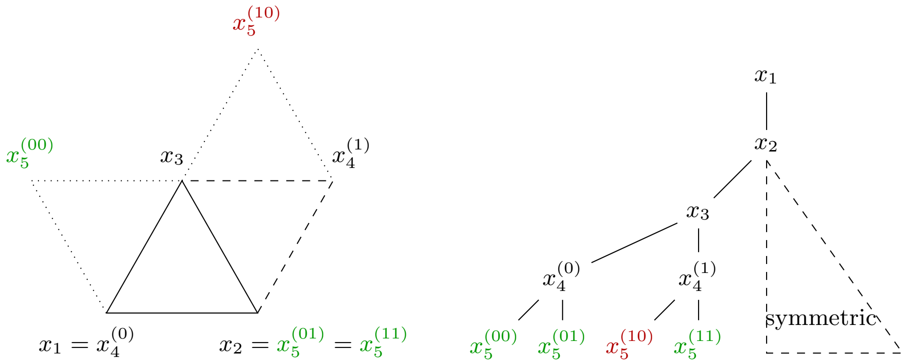
6 Counterexamples

6.1 Disproving the "power of two" conjecture

We first discuss a class of counterexamples to the conjecture that *all* DMDGP instances have a number of solutions which is a power of two (also see Lemma 5.1 in [22]). All these counterexamples are hand-crafted and have the property that two distinct embeddings x, x' have at least a level i where $x_i = x'_i$, which is an event which happens with probability 0. For any $K \geq 1$, let $n = K + 3$, $V = \{1, \ldots, n\}$, $E = \{\{i, j\} \mid 0 < i - j \leq K\} \cup \{\{1, n\}\}$ and $d_{ij} = 1$ for all $\{i, j\} \in E$. The first $n - 2 = K + 1$ points can be embedded in the vertices of a regular simplex in dimension K; then either $x_{n-1} = x_1$ or x_{n-1} is the symmetric position from x_1 with respect to the hyperplane through $\{x_2, \ldots, x_{n-2}\}$. In the first case, the two positions for x_n are valid, in the second only $x_n = x_2$ is possible (see Fig. 2 for the 2-dimensional case), yielding a YES instance where $|X| = 6$.

6.2 Necessity of immediate predecessors

Lastly, Fig. 3 shows an example where the $(ii) \Rightarrow (i)$ implication of Lemma 4.3 fails for instances in DDGP $\smallsetminus$ DMDGP. This shows that any generalization of our result to the DDGP is nontrivial. Let $V = \{1, \ldots, 6\}$ (the graph drawing is the same as the embedding in $\mathbb{R}^2$). The nodes $5', 6'$ linked with dashed lines show alternative node placements. Let $U_5 = \{3, 4\}$ and $U_6 = \{1, 2\}$. The line through the points $3, 4$ does not provide a valid reflection mapping 6 to $6'$. This happens because U_6 does not consist of the two *immediate* predecessors of 6.



(a) Positions of the points on the plane. (b) BP tree.

Fig. 2. The counterexample in the case $K = 2$. Embeddings $x_5^{(00)}$, $x_5^{(01)}$, and $x_5^{(11)}$ are valid, while $x_5^{(10)}$ is not.

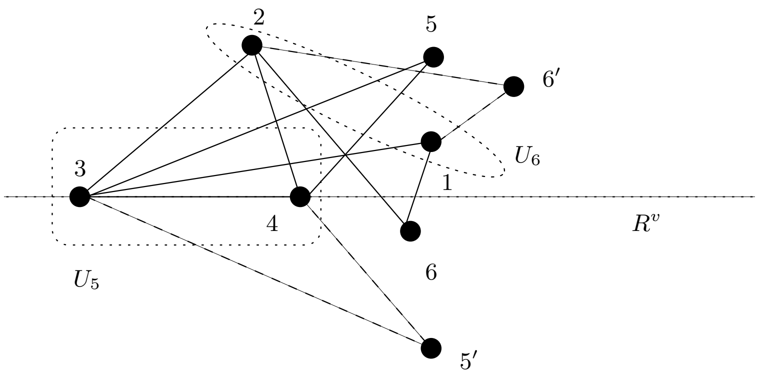


Fig. 3. A counterexample to Lemma 4.3 applied to DDGP ∖ DMDGP.

7 Conclusion

In this paper we showed that YES instances of the DDGP have a number of solutions which is a power of two with probability 1. This settles a question which arose from an empirical observation in [22]. One of the partial results (Thm. 5.4) leading to the proof of this fact will also have practical implications, since all solutions can be expressed in function of one solution by means of a set of flip operations on binary sequences; we are going to test this idea computationally in future work.

References

1. Lavor, C., Liberti, L., Maculan, N.: Computational experience with the molecular distance geometry problem. In Pintér, J., ed.: Global Optimization: Scientific and Engineering Case Studies. Springer, Berlin (2006) 213–225
2. Liberti, L., Lavor, C., Maculan, N., Marinelli, F.: Double variable neighbourhood search with smoothing for the molecular distance geometry problem. Journal of Global Optimization **43** (2009) 207–218
3. Saxe, J.: Embeddability of weighted graphs in k-space is strongly NP-hard. Proceedings of 17th Allerton Conference in Communications, Control and Computing (1979) 480–489
4. Huang, H.X., Liang, Z.A., Pardalos, P.: Some properties for the Euclidean distance matrix and positive semidefinite matrix completion problems. Journal of Global Optimization **25** (2003) 3–21
5. Hendrickson, B.: The molecule problem: exploiting structure in global optimization. SIAM Journal on Optimization **5** (1995) 835–857
6. Eren, T., Goldenberg, D., Whiteley, W., Yang, Y., Morse, A., Anderson, B., Belhumeur, P.: Rigidity, computation, and randomization in network localization. IEEE Infocom Proceedings (2004) 2673–2684
7. Krislock, N., Wolkowicz, H.: Explicit sensor network localization using semidefinite representations and facial reductions. SIAM Journal on Optimization **20** (2010) 2679–2708
8. Gunther, H.: NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry. Wiley, New York (1995)
9. Schlick, T.: Molecular modelling and simulation: an interdisciplinary guide. Springer, New York (2002)
10. Santana, R., Larrañaga, P., Lozano, J.: Combining variable neighbourhood search and estimation of distribution algorithms in the protein side chain placement problem. Journal of Heuristics **14** (2008) 519–547
11. Lavor, C., Mucherino, A., Liberti, L., Maculan, N.: Discrete approaches for solving molecular distance geometry problems using nmr data. International Journal of Computational Biosciences **2010** (2010) 88–94
12. Lavor, C., Liberti, L., Maculan, N., Mucherino, A.: The discretizable molecular distance geometry problem. Computational Optimization and Applications (to appear)
13. Liberti, L., Lavor, C., Maculan, N.: A branch-and-prune algorithm for the molecular distance geometry problem. International Transactions in Operational Research **15** (2008) 1–17
14. Mucherino, A., Lavor, C., Liberti, L.: The discretizable distance geometry problem. Optimization Letters (in revision)
15. Lavor, C., Lee, J., John, A.L.S., Liberti, L., Mucherino, A., Sviridenko, M.: Discretization orders for distance geometry problems. Optimization Letters (to appear)
16. Lavor, C., Mucherino, A., Liberti, L., Maculan, N.: On the computation of protein backbones by using artificial backbones of hydrogens. Journal of Global Optimization (to appear)
17. Liberti, L., Lavor, C., Mucherino, A., Maculan, N.: Molecular distance geometry methods: from continuous to discrete. International Transactions in Operational Research **18** (2010) 33–51

18. Lavor, C., Liberti, L., Maculan, N., Mucherino, A.: Recent advances on the discretizable molecular distance geometry problem. European Journal of Operational Research (submitted (invited survey))
19. Blumenthal, L.: Theory and Applications of Distance Geometry. Oxford University Press, Oxford (1953)
20. Connelly, R.: Generic global rigidity. Discrete Computational Geometry **33** (2005) 549–563
21. Brady, T., Watt, C.: On products of Euclidean reflections. American Mathematical Monthly **113** (2006) 826–829
22. Lavor, C., Liberti, L., Maculan, N.: The discretizable molecular distance geometry problem. Technical Report q-bio/0608012, arXiv (2006)
23. Dong, Q., Wu, Z.: A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data. Journal of Global Optimization **26** (2003) 321–333
24. Coope, I.: Reliable computation of the points of intersection of n spheres in $\mathbb{R}^n$. Australian and New Zealand Industrial and Applied Mathematics Journal **42** (2000) C461–C477

A Appendix: Proofs

Lemma A.1 (3.1). *If $|N(v) \cap \gamma(v)| = K$ then there are at most two distinct extensions of x that are valid for G^v. If one valid extension exists, then with probability 1 there are exactly two distinct valid extensions.*

Proof. Since $|N(v) \cap \gamma(v)| = K$, $U_v = N(v) \cap \gamma(v)$ and v is at the intersection of exactly K spheres in $\mathbb{R}^K$ (each centered at x_u with radius d_{uv}, where $u \in U_v$). The position $z \in \mathbb{R}^K$ of v must then satisfy:

$$\forall u \in U_v \quad \|z - x_u\| = d_{uv} \Rightarrow \|z\|^2 - 2x_u \cdot z + \|x_u\|^2 = d_{uv}^2. \tag{6}$$

As in [23], we choose an arbitrary $w \in U_v$, say $w = \max_< U_v$, and subtract from the Eq. (6) indexed by w the other equations of (6), obtaining the system:

$$\left.\begin{array}{r} \forall u \in U_v \smallsetminus \{w\} \quad 2(x_u - x_w) \cdot z = (\|x_u\|^2 - d_{uv}^2) - (\|x_w\|^2 - d_{wv}^2) \\ \|z\|^2 - 2x_w \cdot z + \|x_w\|^2 = d_{wv}^2. \end{array}\right\} \tag{7}$$

The system (7) consists of a set of $K-1$ linear equations and a single quadratic equation in the K-vector z. We write the linear equations as the system $Az = b$, where the (u,j)-th component of A is $2(x_{uj} - x_{wj})$, the u-th component of b is $\|x_u\|^2 - \|x_w\|^2 - d_{uv}^2 + d_{wv}^2$, A is $(K-1) \times K$ and $b \in \mathbb{R}^{K-1}$. By strict simplex inequality, A has full rank (for otherwise $\sum_{u \neq w} \lambda_u (x_u - x_w) = 0$ implies that x_w is in the span of $\{x_u \mid u \in U_v\}$, and hence that $\Delta_{K-1}(U_v) = 0$); so without loss of generality assume that the square matrix B formed by the first $K-1$ columns of A is invertible. Let z_B be the vector consisting of the first $K-1$ components of z; then the linear part (first $K-1$ equations) of (7) yields $z_B = B^{-1}(b - Nz_K)$, where $N = 2(x_{uK} - x_{wK} \mid u \in U_v \smallsetminus \{w\}) \in \mathbb{R}^{K-1}$. After replacement of z_B in (7) with $z_B(z_K)$, we obtain the following quadratic equation in z_K:

$$(\|\bar{N}\|^2 + 1)z_K^2 - 2((\bar{b} + x_{wB})\bar{N} + x_{wK})z_k + (\|x_{wB} - \bar{b}\|^2 + x_{wK}^2 - d_{wv}^2) = 0, \tag{8}$$

where $\bar{b} = B^{-1}b$ and $\bar{N} = B^{-1}N$. If the discriminant of (8) is negative then no extension of $\bar{x}$ to v is possible and the result follows. If the discriminant is nonnegative, (8) has solutions z'_K, z''_K yielding points $z' = (z_B(z'_K), z'_K)$ and $z'' = (z_B(z''_K), z''_K) \in \mathbb{R}^K$, which are distinct with probability 1 because the discriminant is zero with probability 0. The extended embeddings, distinct with probability 1, are given by (x, z') and (x, z''). □

Lemma A.2 (3.2). *If $|N(v) \cap \gamma(v)| > K$ then, with probability 1, there is at most one extension of x.*

Proof. Consider a subset $S \subseteq N(v) \cap \gamma(v)$ such that $|S| = K+1$ and $S \supseteq U_v$. Either there is at least one point x_v such that (x, x_v) is an embedding of $G[S \cup \{v\}]$ that is valid w.r.t. the system:

$$\forall u \in S \quad \sum_{k \leq K} (x_{vk}^2 - 2x_{uk}x_{vk} + x_{uk}^2) = d_{uv}^2, \tag{9}$$

or the system has no solution. In the latter case, the result follows, so we assume now that there is a point x_v satisfying (9). Since the points x_u are known for all $u \in S$, (9) is a quadratic system with K variables and $K+1$ equations. As in the proof of Lemma 3.1, we derive an equivalent linear system from (9). Since d satisfies the strict simplex inequalities on U_v with probability 1 and $S \supseteq U_v$, by [24] $\{x_u \mid u \in S\}$ are not co-planar and the system has exactly one solution. $\square$

Lemma A.3 (3.3). *With the notation of Lemma 3.1, if $\bar{x}$ is a valid embedding for $G[U_v]$, then z'' is a reflection of z' with respect to the hyperplane through the K points of $\bar{x}$.*

Proof. Any sphere in $\mathbb{R}^K$ is symmetric with respect to any hyperplane through its center; so the intersection of up to K spheres in $\mathbb{R}^K$ is symmetric with respect to the hyperplane containing all the centers. $\square$

Lemma A.4 (3.4). *At termination of Alg. 1, X contains all valid embeddings of G extending $\bar{x}$.*

Proof. Z exists with probability 1 by Lemma 3.1. Every embedding in X is valid because of Steps 17 and 19-20. No other valid extension of $\bar{x}$ exists because of Lemmata 3.1-3.2. $\square$

Proposition A.5 (3.5). *With probability 1, there is no level $i \leq n$ having two distinct feasible nodes $\beta, \theta \in \mathcal{V}_i$ such that $|\{\alpha \in N^+(\beta) \mid \mu(\alpha) = \boxplus\}| = 1$ and $|\{\alpha \in N^+(\theta) \mid \mu(\alpha) = \boxplus\}| = 2$.*

Proof. We show that for all $i \leq n$ the event of having two distinct nodes $\beta, \theta \in \mathcal{V}_i$, with $\rho^{-1}(i) = v$, such that β has one feasible subnode and θ has two has probability 0. Consider $T_v = N(v) \cap \gamma(v)$: if $|T_v| = K$ then by Lemma 3.1 β should have exactly two feasible subnodes with probability 1; since it only has one, the event $|T_v| = K$ occurs with probability 0. Since $|T_v| \geq K$ by (4), the event $|T_v| > K$ occurs with probability 1. Thus by Lemma 3.2 there is at most one valid embedding extending the partial embedding at v, which means that the two feasible subnodes of θ represent the same embedding, an event that occurs with probability 0. $\square$

Lemma A.6 (4.3). *Let $\alpha \in \mathcal{V}_{i-1}$ for some $i > 1$ such that $N^+(\alpha) = \{\eta', \beta'\}$, $u = \rho^{-1}(i)$; $v > u$ with $\rho(v) = \ell$, and consider two feasible nodes $\eta, \beta \in \mathcal{V}_\ell$ such that $\eta' = \mathsf{p}(\eta) \cap \mathcal{V}_i$ and $\beta' = \mathsf{p}(\beta) \cap \mathcal{V}_i$. Then, with probability 1, the following statements are equivalent:*

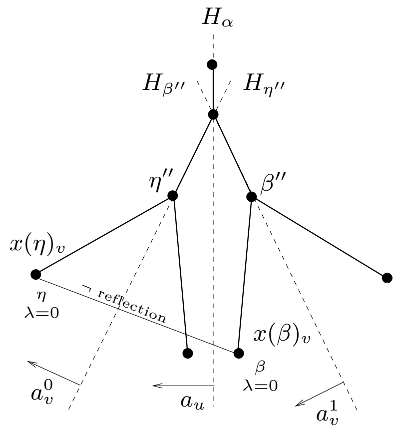
(i) $\forall\, i \leq j \leq \ell$, $x(\beta'')_w = R^u x(\eta'')_w$, *where* $\eta'' = \mathsf{p}(\eta) \cap \mathcal{V}_j$, $\beta'' = \mathsf{p}(\beta) \cap \mathcal{V}_j$, *and* $w = \rho^{-1}(j)$;
(ii) $\forall\, i \leq j \leq \ell$, $\lambda(\eta'') = 1 - \lambda(\beta'')$, *with* $\eta'' = \mathsf{p}(\eta) \cap \mathcal{V}_j$ *and* $\beta'' = \mathsf{p}(\beta) \cap \mathcal{V}_j$.

Proof. Let ${a_v^0}^\top x = a_{v0}^0$, ${a_v^1}^\top x = a_{v0}^1$ be the equations of the hyperplanes H_η, H_β defined respectively by $x(\eta)[U_v]$ and $x(\beta)[U_v]$, with the normals oriented as explained on page 5. We prove by induction on $\ell - i$ that the following assumption is equivalent to (i) and (ii):

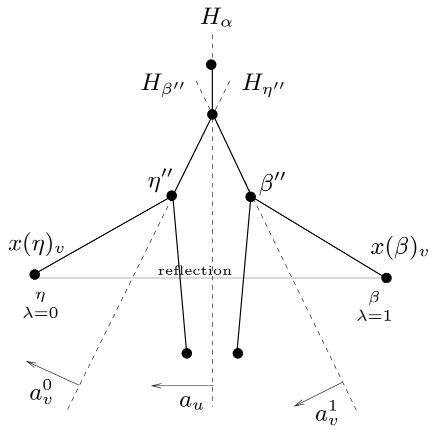
(*iii*) for all $i \le j \le \ell$, $x(\beta'')_w = R^u x(\eta'')_w$ and $a_u \cdot a_w^0 = a_u \cdot a_w^1$, where $\eta'' = \mathsf{p}(\eta) \cap \mathcal{V}_j$, $\beta'' = \mathsf{p}(\beta) \cap \mathcal{V}_j$, $w = \rho^{-1}(j)$, and a_w^0 and a_w^1 are the normal vectors of the hyperplanes $H_{\eta''}$ and $H_{\beta''}$ oriented as usual.

If $\ell = i$, then (*i*), (*ii*), and (*iii*) hold simultaneously. Indeed, $\eta = \eta'$ and $\beta = \beta'$, hence $x(\beta)_v = R^u x(\eta)_v$ (Lemma 3.3) and $\lambda(\eta) = 1 - \lambda(\beta)$ (Alg. 1, Steps 25 and 27). In addition, we have $H_\eta = R^u H_\beta$, therefore $|a_u \cdot a_v^0| = |a_u \cdot a_v^1|$. Because the orientation of a_v^0, a_v^1 is such that $a_u \cdot a_v^0, a_u \cdot a_v^1 \ge 0$, the result holds. Assume that the equivalence stated above holds for level $\ell - 1$, we show that it is still the case at level ℓ. In the sequel, denote $t = \rho^{-1}(\ell - 1)$.

(*i*) $\Leftrightarrow$ (*ii*). Suppose for all $i \le j < \ell$, $x(\beta'')_w = R^u x(\eta'')_w$ and $\lambda(\eta'') = 1 - \lambda(\beta'')$ (by the induction hypothesis, both statements are equivalent). Hence, $H_{\eta''} = R^u H_{\beta''}$ holds for all j, because the K points generating the hyperplanes either belong to H_α, or are reflections of each other. This is true in particular if we choose $\eta'', \beta'' \in \mathcal{V}_{\ell-1}$. In addition, if we use the induction hypothesis (*i*) $\Rightarrow$ (*iii*)), we have $a_u \cdot a_t^0 = a_u \cdot a_t^1$, so a_t^0, a_t^1 are directed similarly w.r.t a_u, and $\lambda(\eta) = 1 - \lambda(\beta)$ if and only if $x(\beta)_v = R^u x(\eta)_v$ (see Fig. 4).



(a) ${a_v^0}^\top x(\eta)_v > a_{v0}^0$ and ${a_v^1}^\top x(\beta)_v > a_{v0}^1$



(b) ${a_v^0}^\top x(\eta)_v > a_{v0}^0$ and ${a_v^1}^\top x(\beta)_v < a_{v0}^1$

Fig. 4. Proof of Lemma 4.3: Case (4a) shows the contradiction deriving from $\lambda(\eta) = \lambda(\beta) = 0$ (or $x(\beta)_v \neq R^u x(\eta)_v$), and case (4b) the situation that actually occurs.

(*ii*) $\Rightarrow$ (*iii*). Suppose for all $i \le j \le \ell$, $\lambda(\eta'') = 1 - \lambda(\beta'')$. By the previous result, we also know that $i \le j \le \ell$, $x(\beta'')_w = R^u x(\eta'')_w$. It remains to prove that $a_u \cdot a_v^0 = a_u \cdot a_v^1$, i.e. that the angles θ_v^0 and θ_v^1 formed by these vectors have the same cosine. Notice once again that $H_\eta = R^u H_\beta$. By induction, we know that the angles θ_t^0, θ_t^1 formed by a_u and respectively a_t^0, a_t^1, have same cosine. With probability 1, the hyperplanes H_η, H_β are not parallel, hence their normal vectors cannot be identical, therefore, $\theta_t^0 = -\theta_t^1$ (see the illustration on Fig. 5).

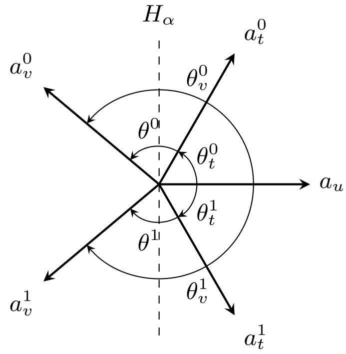


Fig. 5. Proof of Lemma 4.3: illustration of the fact that $a_u \cdot a_v^0 = a_u \cdot a_v^1$.

Denote θ^0, θ^1 the angles formed respectively by a_t^0 and a_v^0, and by a_t^1 and a_v^1. We also have, $H_{\eta''} = R^u H_{\beta''}$, where $\eta'', \beta'' \in \mathcal{V}_{\ell-1}$, hence the normal vectors of these 4 hyperplanes are also symmetric, which implies $\theta^0 = -\theta^1$ or $\theta^0 = \pi - \theta^1$. By the definition of a_v^0 and a_v^1 (page 5), since the scalar products are positive, $-\pi/2 \leq \theta^0, \theta^1 \leq \pi/2$, thus $\theta^0 = -\theta^1$. Therefore, $\theta_v^0 = \theta_t^0 + \theta^0 = -\theta_t^1 - \theta^1 = -\theta_v^1$, which concludes this part of the proof. $(iii) \Rightarrow (i)$. Obvious. $\square$

Proposition A.7 (4.4). *Consider a subtree $\mathcal{T}'$ of $\mathcal{T}$ consisting of $K+2$ consecutive levels $i-K-1, \ldots, i$ (where $i \geq 2K+1$), rooted at a single node η and such that all nodes at all levels are marked $\boxplus$. Let $p = 2^{K+1}$ and consider the set $Y' = \{y_j \mid j \leq p\}$ of partial embeddings of G at the leaf nodes $\{\alpha_j \mid j \leq p\}$ of $\mathcal{T}'$. Let $u = \rho^{-1}(i-K-1)$ and $v = \rho^{-1}(i)$. Then with probability 1 there are two distinct positive reals r, r' such that $\|y_j(\alpha_j)_u - y_j(\alpha_j)_v\| \in \{r, r'\}$ for all $j \leq p$.*

Proof. Fig. 1 shows a graphical proof sketch for $K = 2$. With a slight abuse of notation, for a vertex $w \in V$ in this proof we denote by R^w the set of all reflections at level w. We order the α_j nodes so that the action of R^v on $(\alpha_1, \ldots, \alpha_p)$ is the permutation $\prod_{j \bmod 2=1}(j, j+1)$. Let $t = \rho^{-1}(i-1)$. Since all nodes are feasible, $\|y_j(\alpha_j)_v - y_j(\alpha_j)_t\| = d_{tv}$ and $\|y_j(\alpha_j)_u - y_j(\alpha_j)_t\| = d_{ut}$ for all $j \leq p$ (we remark that $\{t, v\}$ and $\{u, t\}$ must be in E by the definition of the DMDGP). With probability 1, the segments through $y_j(\alpha_j)_u$ and $y_j(\alpha_j)_t$ (where $j \leq p$) do not respectively lie within the hyperplanes defining the reflections R^v; and the same holds for the segments through $y_j(\alpha_j)_t$ and $y_j(\alpha_j)_v$. Thus, there is a set Q of positive reals $r_1, \ldots, r_p$ s.t. for all $j \leq p$ with $j \bmod 2 = 1$ we have $\|y_j(\alpha_j)_u - y_i(\alpha_j)_v\| = r_j$ and $\|y_{j+1}(\alpha_{j+1})_u - y_{j+1}(\alpha_{j+1})_v\| = r_{j+1}$, which shows $|Q| \leq p = 2^{K+1}$. By Lemma 4.3 the action of R^t on $(\alpha_1, \ldots, \alpha_p)$ is the permutation $\prod_{j \bmod 4=1}(j, j+3)(j+1, j+2)$: this implies that $r_j = r_{j+3}$ and $r_{j+1} = r_{j+2}$ for all $j \bmod 4 = 1$, which shows $|Q| \leq p/2 = 2^K$. Inductively, for a vertex w s.t. $i - K \leq \rho(w) \leq i-1$ the action of R^w is $\prod_{j \bmod 2^{i-\rho(w)+1}}(j, j + 2^{i-\rho(w)+1} - 1)(j+1, j + 2^{i-\rho(w)+1} - 2) \cdots (j + 2^{i-\rho(w)} - 1, j + 2^{i-\rho(w)})$, which implies that $|Q| \le 2^{K+1-i+\rho(w)}$. Therefore $\rho(w) = i - K$ proves that $|Q| \le 2$. The case $|Q| = 1$ can only occur if $y_j(\alpha_j)_u$, $y_j(\alpha_j)_t$ and $y_j(\alpha_j)_v$ are collinear for all $j \le p$, an event that occurs with probability 0. □

Corollary A.8 (4.6). *Consider a subtree $\mathcal{T}'$ of $\mathcal{T}$ consisting of $K+q+1$ consecutive levels $i-K-q,\ldots,i$ (where $i \ge 2K+q$ and $q \ge 1$), rooted at a single node η and such that all nodes at all levels are marked $\boxplus$. Let $p = 2^{K+q}$ and consider the set $Y' = \{y_j \mid j \le p\}$ of partial embeddings of G at the leaf nodes $\{\alpha_j \mid j \le p\}$ of $\mathcal{T}'$. Let $u = \rho^{-1}(i-K-q)$ and $v = \rho^{-1}(i)$. Then with probability 1 there is a set $H^{uv} = \{r_j \mid j \le 2^q\}$ of 2^q distinct positive reals such that $\|y_i(\alpha_i)_u - y_i(\alpha_i)_v\| \in H^{uv}$ for all $i \le p$.*

Proof. The proof of Prop. 4.4 can be generalized to span an arbitrary number of levels by induction on q. Two distances $r_{j_1}, r_{j_2} \in H^{uv}$ can only be equal by collinearity of some subsets of points, an event occurring with probability 0. □

Corollary A.9 (4.7). *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. If $\{u,w\} \in E$ with $u < v < w$ and $\rho(w) - \rho(u) > K$ then $d_{uw} \in H^{uw}$ with probability 1.*

Proof. Since $\Upsilon(y,v)$ holds, then the DMDGP instance is YES and there must exist at least two feasible nodes at level $\rho(w)$ in $\mathcal{T}$. If $d_{uw} \notin H^{uw}$ the probability that a completely determined sphere contains two arbitrary points in $\mathbb{R}^K$ is zero. Since the instance is a YES one, however, the BP algorithm does not prune all feasible nodes due to d_{uw}. By Cor. 4.6 the only remaining possibility (which therefore occurs with probability 1) is that $d_{uw} \in H^{uw}$. □

Corollary A.10 (4.8). *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. If $u \in V$ with $u > v$ then $R^v y_u$ belongs to a valid extension of $y[U_v]$.*

Proof. If there is no edge $\{w,u\} \in E$ with $\rho(u) - \rho(w) > K$ the result follows by Cor. 4.1. Otherwise, by Cor. 4.7, $d_{wu} \in H^{wu}$. As in the proof of Prop. 4.4, all pairs of points that are feasible w.r.t. d_{wu} are reflections of each other w.r.t. R^v. □

Theorem A.11 (4.9). *Let $y \in X$ and $v \in V \smallsetminus V_0$ such that $\Upsilon(y,v)$ holds. Then $\tilde{R}^v y \in X$ with probability 1.*

Proof. We have to show that $\tilde{R}^v y$ is a valid embedding for $G = (V,E)$. Partition E into three subsets E_1, E_2, E_3, where $E_1 = \{\{t,u\} \in E \mid t,u < v\}$, $E_2 = \{\{t,u\} \in E \mid t,u \ge v\}$ and $E_3 = \{\{t,u\} \in E \mid t < v \wedge u \ge v\}$. For E_1, by definition $\|(\tilde{R}^v y)_t - (\tilde{R}^v y)_u\| = \|Iy_t - Iy_u\| = \|y_t - y_u\| = d_{tu}$ as claimed. For E_2, $\|(\tilde{R}^v y)_t - (\tilde{R}^v y)_u\| = \|R^v y_t - R^v y_u\| = \|y_t - y_u\| = d_{tu}$ because R^v is an isometry. For E_3, we aim to show that $\|Iy_t - R^v y_u\| = d_{tu}$. Since $y \in X$, by Lemma 3.4 there is a feasible leaf node α with $x(\alpha) = y$. Because $\Upsilon(y,v)$, $\exists \eta \in \mathcal{V}_{\rho(v)-1}$ such that $x(\eta) = y[\gamma(v)]$ and $N^+(\eta) = \{\beta, \beta'\}$ with $\mu(\beta) = \mu(\beta') = \boxplus$; we can assume without loss of generality that $\mathsf{p}(\alpha) \cap \mathcal{V}_{\rho(v)} = \{\beta\}$; furthermore, again by $\Upsilon(y,v)$, there is at least one feasible leaf node α' such that $\mathsf{p}(\alpha') \cap \mathcal{V}_{\rho(v)} =$

$\{\beta'\}$. Let $\{\omega\} = \mathsf{p}(\alpha) \cap \mathcal{V}_{\rho(u)}$ and $\{\omega'\} = \mathsf{p}(\alpha') \cap \mathcal{V}_{\rho(u)}$. Because ω' is feasible, $\|x(\omega')_t - x(\omega')_u\| = d_{tu}$; because η is an ancestor of both α and α' at level $\rho(v)-1$ and $t < v$, $\mathsf{p}(\alpha') \cap \mathcal{V}_{\rho(t)} = \mathsf{p}(\alpha) \cap \mathcal{V}_{\rho(t)}$, which implies that $x(\omega')_t = x(\omega)_t = y_t$. Thus, $\|y_t - y_u\| = d_{tu} = \|y_t - x(\omega')_u\|$. Furthermore, because $\beta' \in \mathsf{p}(\omega') \cap \mathcal{V}_{\rho(v)}$, $x(\omega')$ extends $x(\beta')$. By Alg. 1, Steps 25 and 27, $\lambda(\beta) = 1 - \lambda(\beta')$. Because α is feasible, at every level $\rho(u') \in V$ such that $v \le u' < u$ the node $\theta \in \mathsf{p}(\alpha) \cap \mathcal{V}_{\rho(u')}$ has $f \in \{1, 2\}$ feasible subnodes; by Prop. 3.5, the node $\theta' \in \mathsf{p}(\alpha') \cap \mathcal{V}_{\rho(u')}$ also has f feasible subnodes. If $f = 2$, by Cor. 4.8 it is possible to choose α' so that $\lambda(\theta') = 1 - \lambda(\theta)$ with probability 1; if $f = 1$ then by Alg. 1, Steps 31 and 33, all feasible nodes inherit the same λ value as their parents, so $\lambda(\theta') = 1 - \lambda(\theta)$. By Lemma 4.3, $x(\omega')_u = R^v y_u$ with probability 1. Hence $\|y_t - R^v y_u\| = d_{tu}$ as claimed. □

Lemma A.12 (5.1). *With probability 1, the relation χ is a function.*

Proof. For χ to fail to be well-defined, there must exist an embedding x which is in relation with two distinct binary sequences χ', χ'', which corresponds to the discriminant of the quadratic equation in the proof of Lemma 3.1 taking value zero at some rank $> K$, which happens with probability 0. □

Lemma A.13 (5.2). *ϕ is injective.*

Proof. We show that for all $S, T \subseteq N$, if $g_S = g_T$ then $S = T$.

$$\begin{array}{lll}
 & & g_S = g_T \\
 & \Rightarrow & \bigoplus_{i\in S} g_i = \bigoplus_{i\in T} g_i \\
 & \Rightarrow & \bigoplus_{i\in S} g_i \oplus \bigoplus_{i\in T} g_i^{-1} = e \\
\text{idempotency} & \Rightarrow & \bigoplus_{i\in S} g_i \oplus \bigoplus_{i\in T} g_i = e \\
g_i \oplus g_i = g_i^2 & \Rightarrow & \bigoplus_{i\in S\triangle T} g_i \oplus \bigoplus_{i\in S\cap T} g_i^2 = e \\
\text{idempotency} & \Rightarrow & \bigoplus_{i\in S\triangle T} g_i = e \\
\text{linear independence} & \Rightarrow & S\triangle T = \emptyset \\
 & \Rightarrow & S = T.
\end{array}$$

This concludes the proof. □

Lemma A.14 (5.3). *For all $H \subseteq \Gamma$, $|\langle H\rangle| = 2^{|H|}$.*

Proof. The restriction of function ϕ to $\mathcal{P}(H)$ is injective by Lemma 5.2. Furthermore, each element g of $\langle H\rangle$ can be written as $\bigoplus_{i\in S} g_i$ for some $S \subseteq H$ because H is a spanning set for the vector space H over $\mathbb{F}_2^n$, which is setwise equal to the group $\langle H\rangle$. Thus ϕ is surjective too. Hence ϕ is a bijection between $\mathcal{P}(H)$ and $\langle H\rangle$, which yields the result. □

Theorem A.15 (5.4). *If $\Xi \neq \emptyset$, for all $\xi \in \Xi$ we have $\xi \oplus \Lambda = \Xi$ with probability 1.*

Proof. ($\Rightarrow$) We show that $\xi \oplus \Lambda \subseteq \Xi$ with probability 1; because $\langle L \rangle = \Lambda$ it suffices to show that $\xi \oplus g_i \in \Xi$ for an arbitrary $g_i \in L$, i.e. that there exists a valid embedding $w \in X$ such that $\chi(w) = \xi \oplus g_i$. Let $y \in \chi^{-1}(\xi)$ and $v = \rho^{-1}(i)$ such that $\Upsilon(y, v)$, and define $w = \tilde{R}^v y$ (where $\tilde{R}^v$ is defined in Thm. 4.9 above); by Thm. 4.9, $w \in X$. Let α' be the leaf node of $\mathcal{T}$ such that $x(\alpha') = y$; by Lemma 3.4, there is a leaf node β' such that $x(\beta') = w$. We have to show that for all $\ell \geq i$ the node $\beta \in \mathsf{p}(\beta') \cap \mathcal{V}_\ell$ is such that $\lambda(\beta) = 1 - \lambda(\alpha)$, where α is the node in $\mathsf{p}(\alpha') \cap \mathcal{V}_\ell$. We proceed by induction on ℓ. For $\ell = i$ this holds by Lemma 3.3. For $\ell > i$, the induction hypothesis allows us to apply Lemma 4.3 and conclude that the event $\lambda(\alpha) = 1 - \lambda(\beta)$ occurs with probability 1.

($\Leftarrow$) Now we show that $\Xi \subseteq \xi \oplus \Lambda$ with probability 1, i.e. for any $\eta \in \Xi$ there is $g \in \Lambda$ with $\xi \oplus g = \eta$. We proceed by induction on n, which starts when $n = K+1$: if $K+1 \notin I$ then $|\Xi| = 1$, $L = \emptyset$ and the theorem holds; if $K+1 \in I$ then $|\Xi| = 2$, $L = \{g_{K+1}\}$ and the theorem holds. Now let $n > K+1$; for all $j \in \{K+1, \ldots, n-1\}$ define $\Xi^j = \{\xi^j \mid \xi \in \Xi\}$ and $L^j = \{g_\ell \in \Gamma \mid \ell \in I \wedge \ell \leq j\}$. By the induction hypothesis, for all $\xi' \in \Xi^j$ ($\xi' \oplus \langle L^j \rangle = \Xi^j$). Now, either $n \notin I$ or $n \in I$; by Prop. 3.5, with probability 1 if $n \notin I$ then nodes in $\mathcal{V}_{n-1}$ can only have zero or one feasible subnode (let B_1^n be the set of all such feasible subnodes), and if $n \in I$ then nodes in $\mathcal{V}_{n-1}$ can only have zero or two feasible subnodes β (let B_2^n be the set of all such feasible subnodes). In the former case we let $\Xi^n = \{\xi(x(\beta)) \mid \beta \in B_1^n\}$ and $L^n = L^{n-1}$; in the latter we let $\Xi^n = \{\xi(x(\beta)) \mid \beta \in B_2^n\}$ and $L^n = L^{n-1} \cup \{g_n\}$. In both cases it is easy to verify that the theorem holds for Ξ^n, L^n: in the former case it follows by the induction hypothesis, and in the latter case it follows because $g_n = (0, \ldots, 0, 1)$, namely, if $\eta \in \Xi$ and $n \in I$ then take $\xi = \eta \oplus g_n$ (the result follows by idempotency of g_n). □

Corollary A.16 (5.5). *If a DMDGP instance is feasible, $|X|$ is a power of two with probability 1.*

Proof. By Lemma 5.1 χ is a function with probability 1. Let $x, x' \in X$ be distinct; then by Alg. 1, Steps 25, 27, 31, and 33, the map $\chi : X \to \Xi$ is injective. By definition of Ξ it is also surjective, hence $|X| = |\Xi|$. By Thm. 5.4 $|\Xi| = |\chi \oplus \Lambda|$ for all $\chi \in \Xi$ with probability 1. It is easy to show that $|\chi \oplus \Lambda| = |\Lambda|$, so by Lemma 5.3 $|X|$ is a power of two with probability 1. □