Chapter 2
The Distance Geometry Problem (DGP)

2.1 Definition of the DGP

The fundamental problem of DG, as we have previously stated, is to determine all the coordinates of a set points, in a given geometric space, for which some of the distances are known. Depending on the application, these points can represent stars, reachable points for a robot arm, atoms, or people. Each one of these objects can be represented by a vertex of a graph, and if the distance between them is known, we have an edge connecting the correspondent vertices. Formally, we have the following definition of the Distance Geometry Problem (DGP) [57].

Definition 2.1 (DGP) Given a integer $K > 0$ and a simple connected graph $G = (V, E)$ with weights on the edges given by $d : E \to (0, \infty)$, find a function $x : V \to \mathbb{R}^K$ such that

$$\forall \{u, v\} \in E, \ \|x(u) - x(v)\| = d(u, v). \tag{2.1}$$

Remark 2.1 The norm of (2.1) is general and will depend on the application. This monograph uses the Euclidean norm.

A solution to the DGP associates each vertex of G to a point in $\mathbb{R}^K$ satisfying Eq. (2.1). That is, we wish to position the vertices $u, v \in V$ such that, for $\{u, v\} \in E$, we have situated them in $\mathbb{R}^K$ so that the calculated distance $\|x(u) - x(v)\|$ is the given value $d(u, v)$. The function x is called a *realization* of G. A realization of a graph is a "representation" of its vertices in some Euclidean space $\mathbb{R}^K$. Note that the dimension K and the graph G are inputs/data of the problem. Some DGP variants have the dimension K as part of the problem [57]. This monograph, however, assumes that the dimension K is given a priori.

A realization that satisfies all Eq. (2.1) is a *valid realization*. In order to simplify the notation, we will use x_u, x_v instead of $x(u), x(v)$, and d_{uv} instead of $d(u, v)$.

The focus here is on the cases $K = 2$ and $K = 3$. However, all results can be extended to $\mathbb{R}^K$ [57].

Exercise 2.1 Can there exist more than one solution of a DGP? Can the solution set be empty?

Exercise 2.2 When we "draw" a graph on paper, are we solving a DGP?

We obtain the following system of equations when we use the Euclidean norm in the definition of the DGP with $K = 2$:

$$\sqrt{(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2} = d_{uv} \ \forall \{u, v\} \in E, \tag{2.2}$$

where $x_u^T = (x_{u1}, x_{u2})$ and $x_v^T = (x_{v1}, x_{v2})$. Thus, we have a system with $2|V|$ variables and $|E|$ equations. From (2.2), by squaring both sides, we immediately derive:

$$(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 = d_{uv}^2 \ \forall \{u, v\} \in E. \tag{2.3}$$

Trying to solve the system (2.1) or its associated quadratic system (2.3) as a closed formula is, in general, impossible [7]. Solving the problem numerically also presents difficulties [57].

Exercise 2.3 Consider a DGP with $K = 2, V = \{u, v, r\}, E = \{\{u, v\}, \{v, r\}\}$, and $d_{uv} = d_{vr} = 1$. Solve the problem graphically.

Exercise 2.4 Considering the previous exercise, what would be the solution if we add $\{u, r\}$ to E with $d_{ur} = 1$?

Before we consider solution methods to solve the DGP, we will discuss two important aspects of the problem: (i) the cardinality of the solution set and (ii) the complexity of the problem.

2.2   Number of Solutions of the DGP

What is the importance of knowing the number of solutions of a DGP? Is it possible to have this information before we solve the problem? Besides its theoretical importance, the cardinality of the DGP solution set may help us to solve the problem, since we know how many solutions are being sought. These questions will be discussed more fully in Chap. 5.

We saw from the first two exercises that the number of solutions of a DGP can be infinite. However, does this always happen? We know that, given three points x_u, x_v, x_w in $\mathbb{R}^2$, the triangle inequality

$$d_{uw} \leq d_{uv} + d_{vw}$$

must be satisfied, where d_{uv}, d_{vw}, d_{uw} are the distances between the given points. This means that if we add the edge $\{u, r\}$ to E (in the last exercise), the problem does not have a solution if the triangle inequality is not satisfied.

It is also clear that once we have a solution, there will be an unaccountably infinite number of other solutions by simply rotating and/or translating the given solution. However, excluding rotations, translations, and reflections, the exercises suggest that the set of solutions of a DGP can be of three kinds:

- empty,
- finite,
- unaccountably infinite.

There is one more case to consider. Is it possible to have a DGP with a countably infinite number of solutions? It turns out that this case is impossible, but the proof is not simple enough to be presented here [8].

Exercise 2.5 Consider a DGP with $K = 2, V = \{u, v, r, s\}, E = \{\{u, v\}, \{u, r\}, \{v, r\}, \{v, s\}\}$ and $d_{uv} = d_{ur} = d_{vr} = d_{vs} = 1$. Excluding rotations, translations, and reflections, how many solutions exist?

Exercise 2.6 Considering the previous exercise, how many solutions will exist if we add $\{u, s\}$ to E, with $d_{us} = \sqrt{2}$?

The two exercises above illustrate the fact that the addition of a single edge can engender a change from "uncountably many" to "finitely many" solutions. This is an evidence (not a proof) that the DGP cannot have countably infinitely many solutions.

2.3   Complexity of the DGP

The focus of this section is to give an intuition of the computational difficulty we face when solving a DGP (formally, this is investigated in computational complexity theory). Let us first consider the DGP whose associated graph is complete. To this end, consider a DGP with $K = 1$, $V = \{u, v, r\}$, $E = \{\{u, v\}, \{u, r\}, \{v, r\}\}$, $d_{uv} = d_{vr} = 1$, and $d_{ur} = 2$. If we fix $x_u = 0$ and $x_v = 1$, we have

$$\|x_r - x_u\| = 2$$
$$\|x_r - x_v\| = 1.$$

Squaring both terms of equalities, we have

$$x_r^2 - 2x_r x_u + x_u^2 = 4$$
$$x_r^2 - 2x_r x_v + x_v^2 = 1.$$

Subtracting one equation from the other, we obtain

$$-2x_r x_u + 2x_r x_v + x_u^2 - x_v^2 = 3 \Rightarrow 2x_r(x_v - x_u) = x_v^2 - x_u^2 + 3.$$

Using the fixed values for x_u and x_v, we get

$$2x_r = 4 \Rightarrow x_r = 2.$$

We could have solved this problem by simply drawing the graph. However, the interesting part of this procedure is that it can be generalized for $\mathbb{R}^K$. Let us see what happens in $\mathbb{R}^2$.

Consider a DGP with $K = 2$,

$$V = \{u, v, r, s\} \text{ and } E = \{\{u, v\}, \{u, r\}, \{u, s\}, \{v, r\}, \{v, s\}, \{r, s\}\}.$$

Assume that u, v, r are fixed, that is, we can find $x_u, x_v, x_r \in \mathbb{R}^2$ such that $\|x_u - x_v\| = d_{uv}$, $\|x_u - x_r\| = d_{ur}$, $\|x_v - x_r\| = d_{vr}$. Given these three points in $\mathbb{R}^2$, we can construct a quadratic system to obtain the coordinates of x_s in the following way:

$$\|x_s - x_u\| = d_{us}$$
$$\|x_s - x_v\| = d_{vs}$$
$$\|x_s - x_r\| = d_{rs}.$$

Squaring both terms of equalities,

$$\|x_s\|^2 - 2x_s \cdot x_u + \|x_u\|^2 = d_{us}^2$$
$$\|x_s\|^2 - 2x_s \cdot x_v + \|x_v\|^2 = d_{vs}^2$$
$$\|x_s\|^2 - 2x_s \cdot x_r + \|x_r\|^2 = d_{rs}^2,$$

and subtracting the first equation from the other two, we obtain

$$2(x_v - x_u) \cdot x_s = \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2$$
$$2(x_r - x_u) \cdot x_s = \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2.$$

Thus, we have a linear system

$$Ax = b,$$

where

$$A = 2 \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix},$$

$$b = \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2 \end{bmatrix},$$

and

$$x = \begin{bmatrix} x_{s1} \\ x_{s2} \end{bmatrix}.$$

If the matrix A is invertible, then we have only one solution x^* given by

$$x^* = A^{-1}b. \tag{2.4}$$

Remark 2.2 Note that supposing that the matrix A is invertible, the solution x^* is obtained for any values of d_{us}, d_{vs}, d_{rs}. What is the connection of this fact with the solution of the original quadratic system?

If A has no inverse, what do we do? We can, for example, fix three different vertices of $\{u, v, r, s\}$ in order to obtain an invertible matrix and hence the position of the fourth vertex. Therefore, if the graph G of a DGP is complete and we assume the existence of a solution, we can obtain a realization of G by solving a sequence of linear systems, considering that all the associated matrices are invertible.

Exercise 2.7 In Exercise 2.5, suggest a method for fixing the positions of u, v, r. Does it generalize to $\mathbb{R}^3$?

Exercise 2.8 In Exercise 2.5, what conditions can we impose on x_u, x_v, x_r to guarantee that the associated linear system has a solution?

Exercise 2.9 In Exercise 2.5, if the associated linear system has a solution, can we guarantee that we have a solution to the original quadratic system?

Exercise 2.10 Still considering Exercise 2.5, if none of possible choices for the three first vertices produce an invertible matrix, does this mean that the DGP has no solution?

If we require that all the matrices associated to a DGP with complete graph possess inverses, the problem has a unique solution (modulo rotations, translations, and reflections) that can be obtained at a computational cost proportional to $|V|$ [20]. However, in the majority of applications, we do not have complete graphs. So, the strategy discussed above may yield just a partial realization of the graph.

Before continuing our exploration of the DGP, we state a theoretical result pertaining to the computational complexity of the DGP [72].

Theorem 2.1 *The DGP is NP-Hard for all $K \in \mathbb{N}$, and in particular it is NP-Complete for $K = 1$.*

This means that any algorithm capable of solving the DGP is very likely to run (in the worst case) in a number of steps which is exponential in the size of the memory used to store the instance data, that is, G and d.

Exercise 2.11 To solve an instance of the DGP in linear time, does the associated graph necessarily need to be complete?

Exercise 2.12 Can we have just a unique solution (modulo rotations, translations and reflections) to the DGP even though the graph is not complete?

2.4   DGP Instances

It is important to be able to generate instances/examples of DGPs in order to test algorithms and to analyze the relationship between the graph of a DGP and its set of solutions. This section presents a method for the reader to generate example problems of a DGP for $K = 3$.

The procedure given here stems from the calculation of 3D molecular structures. To simplify the process without making the generated instance easy to solve, we consider a sequence of covalently connected atoms, denoted by $1, \ldots, n$. That is, each atom is connected to only two others, except the first and the last of the sequence.

We will use a Cartesian coordinate system $x_1, \ldots, x_n \in \mathbb{R}^3$ to define a spatial structure of our molecule in terms of an *internal coordinate system* [77], given by the *lengths of the covalent bonds* $d_{1,2}, \ldots, d_{n-1,n}$, by the *planar angles* $\theta_{1,3}, \ldots, \theta_{n-2,n}$ (formed by three consecutive atoms), and by the *torsion angles* $\omega_{1,4}, \ldots, \omega_{n-3,n}$ (formed by four consecutive atoms). Each torsion angle $\omega_{i-3,i}$ is, in fact, the angle between the normals of the planes defined by the atoms $i-3, i-2, i-1$ and $i-2, i-1, i$, respectively (see Fig. 2.1).
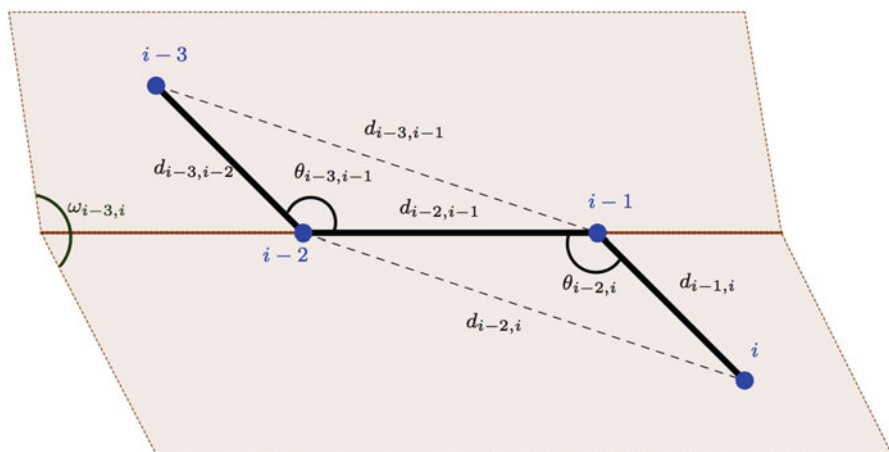

Fig. 2.1 Internal coordinates

We now fix the lengths of the covalent bonds (for example, $d_{i-1,i} = 1.526$) and the values of the planar angles (for example, $\theta_{i-2,i} = 1.91$ radians). In this way, except for rotations, translations, and reflections, a structure for our molecule will be determined by the torsion angles $\omega_{1,4}, \ldots, \omega_{n-3,n}$, each of which can vary in the interval $[0, 2\pi]$.

A way to proceed is to choose randomly values for $\omega_{i-3,i} \in [0, 2\pi]$, as well as pairs of points i, j whose Euclidean distances d_{ij} are smaller than a given value. To simulate DGP instances associated with the calculation of molecular structures using distance information provided by NMR experiments, we can choose pairs of points i, j for which $d_{ij} \leq 5$ [40] (see Chap. 6).

However, how do we determine the pairs of points i, j without knowing the distances d_{ij}? For this, we need to obtain the Cartesian coordinates from the internal coordinates, as follows:

$$
\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ 1 \end{bmatrix} = B_1 B_2 \cdots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \forall i = 1, \ldots, n,
$$

where

$$
B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -d_{1,2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},
$$

$$
B_3 = \begin{bmatrix} -\cos\theta_{1,3} & -\sin\theta_{1,3} & 0 & -d_{2,3}\cos\theta_{1,3} \\ \sin\theta_{1,3} & -\cos\theta_{1,3} & 0 & d_{2,3}\sin\theta_{1,3} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},
$$

and

$$
B_i = \begin{bmatrix} -\cos\theta_{i-2,i} & -\sin\theta_{i-2,i} & 0 & -d_{i-1,i}\cos\theta_{i-2,i} \\ \sin\theta_{i-2,i}\cos\omega_{i-3,i} & -\cos\theta_{i-2,i}\cos\omega_{i-3,i} & -\sin\omega_{i-3,i} & d_{i-1,i}\sin\theta_{i-2,i}\cos\omega_{i-3,i} \\ \sin\theta_{i-2,i}\sin\omega_{i-3,i} & -\cos\theta_{i-2,i}\sin\omega_{i-3,i} & \cos\omega_{i-3,i} & d_{i-1,i}\sin\theta_{i-2,i}\sin\omega_{i-3,i} \\ 0 & 0 & 0 & 1 \end{bmatrix},
$$

for $i = 4, \ldots, n$.

Using above matrices and fixing the values $d_{1,2}, d_{2,3}, \theta_{1,3}$, the coordinates of the first three atoms are given by:

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$x_2 = \begin{bmatrix} -d_{1,2} \\ 0 \\ 0 \end{bmatrix},$$

$$x_3 = \begin{bmatrix} -d_{1,2} + d_{2,3}\cos\theta_{1,3} \\ d_{2,3}\sin\theta_{1,3} \\ 0 \end{bmatrix}.$$

Note that we are using matrices in $\mathbb{R}^{4\times4}$ to generate points in $\mathbb{R}^3$. These are related to *homogeneous coordinates*, which are very useful in computer graphics [29].

Exercise 2.13 Since 3D molecular structures can be described by Cartesian coordinates or internal coordinates, what is the difference in using one or the other coordinate system to generate a DGP instance?

Exercise 2.14 Is it possible to use the matrices above to generate DGP instances for $K = 2$?