

Geometria de Distâncias

Guilherme Philippi

8 de maio de 2020

Sumário

1	Geometria de Distâncias Euclidianas	1
1.1	Como tudo Começou	1
1.2	O Problema Fundamental	3
1.3	Os Diferentes Problemas em DG	4
1.3.1	Conformações Moleculares	4
1.3.2	Localização de Sensores	5
1.3.3	Dinâmicas em Cinemática Inversa	6
1.3.4	Escalonamento Multidimensional	6
1.4	A Busca de uma Solução	6
1.4.1	A Quantidade de Soluções do Problema	7
1.5	Combinatória do DGP	7
1.5.1	Realização de Grafos Completos	8
1.5.2	Trilateração	8
	Referências	10
	A Métricas	12
	B Lei dos Cos e Ângulos Entre dois Vetores no \mathbb{R}^3	12

1 Geometria de Distâncias Euclidianas

Apresenta-se nesta seção uma introdução a *Geometria de Distâncias Euclidianas*. O nome “Geometria de Distâncias” diz respeito ao conceito desta geometria basear-se em distâncias ao invés de pontos. A palavra “Euclidiana” é importante para caracterizar as arestas — elementos fundamentais associados as distâncias — como segmentos, sem restringir seus ângulos de incidência [1].

1.1 Como tudo Começou

Por volta de 300 AC, Euclides de Alexandria organizou o conhecimento de sua época acerca da Geometria em uma obra composta por treze volumes, onde construiu, a partir de um pequeno conjunto de axiomas fortemente baseado nos conceitos de pontos e linhas, a chamada *Geometria Euclidiana* [2]. Em contraponto a visão original de Euclides, os primeiros conceitos geométricos usando *apenas distâncias*

costumam estar associados aos trabalhos de Heron de Alexandria (10 a 80 DC) [1], com o desenvolvimento de um teorema que leva seu nome, como segue:

Teorema de Heron: Sejam s o *semiperímetro* de um triângulo (se p é o perímetro, $s = \frac{p}{2}$) e a , b e c os comprimentos dos três lados deste triângulo. Então, a área A do triângulo é

$$A = \sqrt{s(s-a)(s-b)(s-c)}. \quad (\text{Fórmula de Heron})$$

Pode-se dizer que esse foi o nascimento da *Geometria de Distâncias* (*Distance Geometry*, ou DG).

Algumas centenas de anos depois, em 1841, Arthur Cayley (1821 a 1895) generalizou a Fórmula de Heron através da construção de um determinante que calcula o conteúdo (volume n -dimensional) de um *simplex*¹ em qualquer dimensão [3]. Um século depois, em 1928, o matemático austríaco Karl Menger (1902 a 1985) re-organizou as ideias de Cayley e trabalhou em uma construção axiomática da geometria através de distâncias [4] — donde a alteração no nome do determinante de Cayley para como é conhecido hoje: “*Determinante de Cayley-Menger*”.

Definição: Sejam A_0, A_1, \dots, A_n $n+1$ pontos que definem os vértices de um n -simplex em um espaço euclidiano K -dimensional, onde $n \leq K$, e seja d_{ij} a distância entre os vértices A_i e A_j , onde $0 \leq i < j \leq n$. Então, o conteúdo v_n desse n -simplex é

$$v_n^2 = \frac{(-1)^{n+1}}{(n!)^2 2^n} \begin{vmatrix} 0 & d_{01}^2 & d_{02}^2 & \dots & d_{0n}^2 & 1 \\ d_{01}^2 & 0 & d_{12}^2 & \dots & d_{1n}^2 & 1 \\ d_{02}^2 & d_{12}^2 & 0 & \dots & d_{2n}^2 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{0n}^2 & d_{1n}^2 & d_{2n}^2 & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{vmatrix}. \quad (\text{Determinante de Cayley-Menger})$$

Mas foi só com Leonard Blumenthal (1901 a 1984) que, em 1953, o termo Geometria de Distâncias foi cunhado — com a publicação de seu livro “*Theory and Applications of Distance Geometry*” [5]. Blumenthal dedicou sua vida de trabalho para clarificar, organizar e traduzir as obras originais em alemão [1]. Ele acreditava que o problema mais importante nesta área era o “*Problema de Subconjunto*” (ou *Subset Problem*, originalmente), que consistia em encontrar condições necessárias e suficientes a fim de decidir quando uma matriz simétrica era, de fato, uma *matriz de distâncias*² [6]. Uma restrição desse problema à métrica euclidiana chama-se *Problema de Matrizes de Distâncias Euclidianas* (ou EDMP, do inglês *Euclidean Distance Matrix Problem*), como segue definida:

Problema de Matrizes de Distâncias Euclidianas: Determinar se, para uma dada matriz quadrada $D_{n \times n} = (d_{ij})$, existe um inteiro K e um conjunto $\{p_1, \dots, p_n\}$ de pontos em \mathbb{R}^K tal que $d_{ij} = \|p_i - p_j\|$ para todo $i, j \leq n$.

¹Um simplex é uma generalização do conceito de triângulo a outras dimensões, i.e.: O *0-simplex* é um ponto, *1-simplex* é um segmento de reta, *2-simplex* é um triângulo e o *3-simplex* é um tetraedro.

²Seja o par (\mathcal{X}, d) um *espaço métrico* (vide Apêndice A), onde $\mathcal{X} = \{x_1, \dots, x_n\}$. Uma *matriz de distância sobre \mathcal{X}* é uma matriz quadrada $D_{n \times n} = (d_{uv})$ onde, para todo $u, v \leq n$, temos $d_{uv} = d(x_u, x_v)$ [6].

Condições necessárias e suficientes para que uma matriz seja, de fato, uma matriz de distância euclidiana são dados em [7]. Para isso, apresenta-se um teorema onde se utiliza o Determinante de Cayley-Menger na criação de duas condições afirmando que, afim de $D_{n \times n}$ ser uma matriz de distâncias euclidianas, deve haver um K -simplex S de referência com conteúdo $v_K \neq 0$ em \mathbb{R}^K e que todos os $(K+1)$ -simplex e $(K+2)$ -simplex contendo S como uma das faces devem estar contidos em \mathbb{R}^K [6].

Blumenthal percebeu a importância em se respeitar as restrições métricas estabelecidas pelas matrizes de distâncias.

Quando temos como dado um conjunto de distâncias entre pares de pontos, a geometria das distâncias pode dar uma dica para encontrar um conjunto de coordenadas correto para pontos no espaço Euclidiano tridimensional, satisfazendo as restrições de distâncias dadas.

(Blumenthal, 1953, [5])

Pode-se dizer que resolver o Problema de Matrizes de Distâncias Euclidianas está intimamente relacionado com descobrir as coordenadas dos pontos que definem suas distâncias. Perceba que este é um problema inverso, onde o “problema direto” correspondente é calcular distâncias associadas a pares de pontos dados. Note que este estudo tem enorme aplicabilidade [6].

Adiante, em 1979, Yemini (atualmente professor emérito de Ciência da Computação na Universidade de Columbia) foi o primeiro a flexibilizar a definição do EDMP ao considerar um conjunto de distâncias esparso [8, 6] — i.e., que não se tem todas as distâncias dadas a priori. Com isso, introduziu-se o que se chamou de *Problema Posição - Localização*, onde deseja-se calcular a localização de todos os objetos imersos em um espaço geográfico [8].

Assim, foi possível re-formular o problema fundamental de Geometria de Distâncias, o qual pode ser caracterizado de forma mais moderna pela utilização da Teoria de Grafos [6].

1.2 O Problema Fundamental

Uma *realização* é uma função que mapeia um conjunto de vértices de um grafo G para um espaço euclidiano de alguma dimensão dada [1].

Problema de Geometria de Distâncias (DGP): Dados um grafo simples, ponderado e conectado $G = (V, E, d)$ e um inteiro $K > 0$, encontre uma realização $x : V \rightarrow \mathbb{R}^K$ tal que:

$$\forall \{u, v\} \in E, \quad \|x(u) - x(v)\| = d(u, v). \quad (1)$$

Desde que uma realização seja encontrada, também dá-se a ela o nome de *solução* do DGP. Por simplicidade — claramente um abuso de notação —, pode-se escrever x_u e d_{uv} no lugar de $x(u)$ e $d(u, v)$, respectivamente.

A principal diferença desta definição para o EDMP está acerca de que uma matriz de distância essencialmente representa um *grafo ponderado completo*. Em

contraponto, o DGP não empoe qualquer estrutura em G^3 , seguindo o conceito de matriz esparsa estabelecido por Yemini.

Por fim, na equação 1, utiliza-se a norma euclidiana $\| \cdot \|$ como métrica (ver Apêndice A), donde pode-se reescrever esta equação como

$$\forall \{u, v\} \in E, \quad \sqrt{\sum_{i=1}^K (x_{ui} - x_{vi})^2} = d_{uv}.$$

Como a definição de métrica garante a positividade das distâncias, pode-se esconder a raiz quadrada na equação acima, i.e.

$$\forall \{u, v\} \in E, \quad \sum_{i=1}^K (x_{ui} - x_{vi})^2 = d_{uv}^2. \quad (2)$$

1.3 Os Diferentes Problemas em DG

Em 2014, Leo Liberti *et al.* publicaram um ótimo compendio sobre a *Geometria de Distâncias Euclidianas e suas Aplicações* [6] e, em particular, desenvolveram um estudo taxonômico muito interessante sobre os problemas clássicos da área. No que se segue, devido a grande quantidade de siglas e variações dentro de DG, apresenta-se parte desse estudo, visando organizar os conceitos.

As principais aplicações em DG são no *calculo de estruturas moleculares* [9], na *localização de sensores em rede sem fio* (*Wireless Sensor Network Localization*, ou WSNL) [10], em *cinemática inversa* (*Inverse Kinematic*, ou IK) [11] e em *escalamento multidimensional* (*Multidimensional Scaling*, ou MDS) [12].

1.3.1 Conformações Moleculares

Existe uma relação muito forte com a forma geométrica das moléculas e suas funções em organismos vivos [13]. Projetar drogas para curar uma doença específica se trata basicamente de conhecer o que uma certa proteína pode fazer em um organismo [1]. Proteínas se ligam em outras moléculas através do equilíbrio de forças agindo entre elas⁴, por tanto, suas ligações dependem do seu formato.

Proteínas são constituídas por um grande conjunto de átomos e, alguns pares destes, trocam ligações químicas — sabe-se quais são esses átomos através de experimentos de cristalografia [15]. Então, se os átomos de uma molécula forem rotulados da forma $1, 3, 4, \dots, n$, então é possível inferir:

- O conjunto de ligações $\{u, v\}$, onde u, v são átomos em $\{1, \dots, n\}$;
- A distância entre u e v (para cara par ligado);

³A menos, é claro, no que diz respeito a seus vértices estarem conectados. Porém, caso G não seja conectado, então ele consiste de um conjunto de diferentes subgrafos conectados, donde, a fim de solucionar o DGP, pode-se realizar cada subgrafo separadamente [1].

⁴Ou seja, o equilíbrio da energia potencial das moléculas, proporcional, principalmente, as variações nos comprimentos das ligações covalentes, as variações nos ângulos entre duas ligações covalentes consecutivas, as rotações sobre as ligações covalentes e as interações de van der Waals e interações eletrostáticas entre átomos [14].

- O ângulo interno θ_v definido por duas ligações $\{u, v\}$ e $\{v, w\}$, com um átomo v em comum. (veja o Apêndice B)

Além desses dados, também é possível obter informações a partir de experimentos mais sofisticados, como a *Ressonância Magnética Nuclear* (RMN). Neste experimento é escolhida uma faixa de radiofrequência para bombardear uma amostra que está imersa em um campo magnético bastante intenso. Dependendo da radiofrequência utilizada (costuma-se usar a do hidrogênio), alguns núcleos atômicos irão absorver energia e outros não. Caso atinja-se uma frequência exata de ressonância dentro destes núcleos atômicos, é possível medir essa ressonância como um sinal de radiofrequência enviado dos núcleos atômicos — para calcular distâncias entre átomos próximos, com distâncias menores que 5\AA .

De posse dessas informações, deseja-se realizar (localizar) todos os átomos da molécula. Esse problema, com todas as informações moleculares disponíveis, denomina-se *Estrutura Proteica a partir de Dados Brutos* (*Protein Structure from Raw Data*, ou PSRD)

Em particular, como as coordenadas atômicas pertencem ao \mathbb{R}^3 , há uma particularização do DGP para o caso molecular, chamado *Problema de Geometria de Distâncias Moleculares* (*Molecular DGP*, ou MDGP). Trata-se do DGP com $K = 3$ fixo.

1.3.2 Localização de Sensores

O *Problema de Localização de Sensores em Rede sem Fio* (ou *WSNL Problem*) surge quando é necessário localizar um conjunto de objetos equipados com sensores eletrônicos capazes de medir distâncias entre si, geograficamente distribuídos, usando apenas medidas de distâncias entre pares destes objetos [10].

Por exemplo, *smartphones* com WIFI ativo podem criar uma rede conhecida por *Rede Ad-Hoc*, *i.e.*, eles conseguem criar uma rede para comunicar-se entre si, de forma *peer-to-peer*, sem a necessidade de uma torre central — cada aparelho funciona como uma pequena torre, de forma que a distância entre os aparelhos não pode ser excessiva. Dessa forma, os *smartphones* podem estimar a distância r de emparelhamento das suas conexões ao medir, por exemplo, qual a potência de transmissão do sinal, uma vez que sabe-se que a potência P de uma transmissão eletromagnética cai da forma

$$P = \frac{X}{r^n}, \quad (3)$$

onde X e n são constantes e dependem muito das condições do experimento, sendo obtidas experimentalmente [16].

Em essência, um problema do tipo WSNL segue a mesma definição do DGP, porém, com um subconjunto $A \subset V$ de vértices (chamados *âncoras*), onde os elementos de A tem uma posição em \mathbb{R}^k dada a priori — isso é feito pois, normalmente, interessa saber a posição relativa de um objeto a outro, como é o caso do Sistema de Posicionamento Global, onde temos os satélites como âncoras e desejamos saber a posição dos aparelhos GPS em relação aos satélites.

Por motivos práticos — semelhantes ao caso molecular — as variações de interesse desse problema tem o K fixo em $K = 2$ ou $K = 3$. É comum, também, que se defina um WSNL como *solucionável* somente se seu grafo possua uma única realização válida [1] — noção conhecida como *globalmente rígido*: Diz-se que um grafo

é *globalmente rígido* quando ele possui uma realização genérica x e, para todas as outras realizações x' , x é congruente a x' .

1.3.3 Dinâmicas em Cinemática Inversa

Muito utilizada em robótica e animação computadorizada, a cinemática inversa cerne sobre mecanismos e seus movimentos rígidos, onde restringe-se os movimentos de forma a preservar a geometria do sistema. Sem o auxílio computacional e matemático a manipulação de mecanismos com muitos graus de liberdade pode ser inviável: Imagine a manipulação manual de cem vértices em uma haste simulando o comportamento de um braço articulado em uma animação. Com o auxílio da DG, um animador pode apenas configurar a posição final de um pequeno grupo de vértices (como os da extremidade da aresta, por exemplo) e um algoritmo de cinemática inversa é capaz de verificar se aquela posição é ou não viável e, se viável, qual a realização de todo o conjunto de vértices em razão da posição configurada [11].

Visando tal restrição mecânica, define-se o *Problema de Cinemática Inversa* (*Inverse Kinematic Problem*, ou IKP) como uma variação do WSNL — logo, tem o objetivo de descobrir posições em relação a certos pontos previamente realizados — com uma restrição no grafo que define o problema: deve ser um caminho simples com seus vértices finais sempre sendo âncoras [6].

1.3.4 Escalonamento Multidimensional

O problema de *Escalonamento Multidimensional* (*Multidimensional Scaling*, ou MDS) é definido como [6]: Dado um conjunto X de vetores, encontre um conjunto Y de vetores com menor dimensão (com $|X| = |Y|$) tal que a distância entre cada i -ésimo e j -ésimo vetores de Y tenham, aproximadamente, a mesma distância que seus pares de vetores correspondentes em X .

Esse problema é muito aplicado na análise de dados em Big Data [1]. É um meio de facilitar a visualização do nível de similaridade entre casos individuais — que não necessariamente precisam ter uma conexão aparente — em um conjunto de dados. Pode-se usá-lo, por exemplo, para visualizar em uma escala bidimensional (\mathbb{R}^2) a evolução da locomoção de animais no espaço tridimensional utilizando dados de séries temporais (espaço em diferentes tempos, logo, dados em \mathbb{R}^4).

1.4 A Busca de uma Solução

A abordagem mais simples, pode-se pensar, para encontrar um conjunto de soluções que satisfaça a equação 2 é resolver o sistema de equações diretamente [17]. Infelizmente, para $K \geq 2$, há evidências de que uma solução de forma fechada onde todo componente de x é expresso por raízes, não é possível [1].

No entanto, pode-se re-formular o problema como um problema de otimização global [1], onde o objetivo é minimizar a soma dos erros⁵ entre as distâncias dadas a priori e as calculadas. Para isso, pode-se considerar uma única expressão que englobe todos os n erros, da forma

$$f(x_1, \dots, x_n) = \sum_{(i,j) \in E} (\|x_i - x_j\| - d_{ij})^2. \quad (4)$$

⁵Em otimização, vê-se a equação 1 de forma não exata: $\|x_u - x_v\| = d_{uv} + \varepsilon$, onde ε é chamado *erro*. Ou seja, para minimizar o erro, precisa-se minimizar a expressão $f(x_u, x_v) = \|x_u - x_v\| - d_{uv}$.

Fica claro que encontrar uma solução para o DGP é equivalente a encontrar realizações $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, se e somente se $f(x_1, \dots, x_n) = 0$ [1]. Pela definição de métrica (vide Apêndice A), 0 é o menor valor possível para f , donde diz-se que deseja-se *minimizar a função* $f: \mathbb{R}^n \rightarrow \mathbb{R}$ [17]. Ou seja,

$$\min_{x_i \in \mathbb{R}^n} f(x_1, \dots, x_n). \quad (5)$$

E, no caso da métrica euclidiana [1], temos

$$\min_{x_j \in \mathbb{R}^n} \sum_{(u,v) \in E} \left(\sum_{i=1}^K (x_{ui} - x_{vi})^2 - d_{uv}^2 \right)^2. \quad (6)$$

Por tanto, a equação 6 tem como objetivo a minimização de um polinômio de múltiplas variáveis de grau quatro.

Um dos desafios da Otimização Global é que muitos dos métodos existentes — em especial, os mais eficientes — não garantem que uma otimização *global* será encontrada [1]. Isso se dá pois, dependendo do comportamento da função, existem muitos ótimos locais e os métodos não conseguem diferenciá-los de um global.

Infelizmente, essa abordagem via otimização é custosa do ponto de vista computacional [1]. Saxe demonstrou em 1979 [18] que resolver um DGP para qualquer dimensão — i.e., para qualquer valor de K — tem a complexidade computacional da classe **NP**-Hard. Em outras palavras, isso significa que resolver um DGP pode exigir um custo computacional da proporcional a $2^{|V|}$.

1.4.1 A Quantidade de Soluções do Problema

Pode-se verificar que, dada uma solução de um DGP, obtêm-se facilmente uma quantidade infinita (não enumerável!) de outras realizações válidas, distintas, através de rotações e translações desta solução inicial [17] — infelizmente, essas costumam não serem soluções de interesse.

Contudo, se for desconsiderado essas soluções advindas de translações e rotações, a quantidade de soluções do DGP depende da estrutura geométrica do grafo que a define: podem não haver nenhuma realização; uma única realização; uma quantidade finita (não única) de realizações; ou, um número incontável de realizações [1]. Perceba que o conjunto solução de um DGP, curiosamente, somente não pode ser um número infinito enumerável — resultado obtido através da *Geometria Algébrica Real* [19].

Por tanto, supondo que o conjunto solução de um DGP seja não vazio, sabe-se que ele é não enumerável ou finito. Se for finito, além de aplicar os métodos de Otimização Global — já definidos como custosos coputacionalmente —, pode-se explorar outras abordagens, como a Otimização Combinatória [17].

1.5 Combinatória do DGP

Nesta seção, analisando o espaço de busca por uma solução, faz-se um estudo sobre as condições que garantem a finitude do conjunto solução do problema.

Em particular, para um DGP definido em um espaço de dimensão K , a classe de problemas mais simples de resolver são dos grafos $(K+1)$ -cliques, isto é, dos grafos completos com $K+1$ vértices [1].

1.5.1 Realização de Grafos Completos

Em um espaço unidimensional, i.e., uma reta, tem-se $K+1 = 2$. O grafo completo em dois vértices é $K_2 = (\{v_1, v_2\}, \{\{v_1, v_2\}\})$, ou seja, o grafo é dois vértices com uma aresta entre eles — uma linha (veja a Figura 1). No espaço bidimensional, o grafo completo é um triângulo e, no tridimensional, um tetraedro.

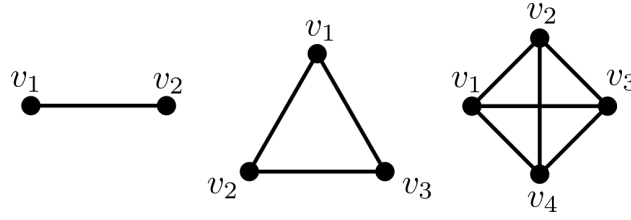


Figura 1: Em ordem: 2-clique; 3-clique; 4-clique.

No geral, se um $(K+1)$ -clique tem uma realização em \mathbb{R}^{K-1} , ela é única a menos de rotações e translações [1]. É trivial, por tanto, que um $(K+2)$ -clique tem, no máximo, uma realização no espaço \mathbb{R}^K — caracterizado como solução do DGP. Por tanto, dependendo da geometria do grafo que define um DGP — isto é, se ele possuir cliques suficientes —, pode-se utilizar estas estruturas como “blocos básicos de construção” para planejar uma realização interativa do grafo como um todo [1].

1.5.2 Trilateração

Considere um 3-clique com $V = \{1, 2, 3\}$, onde $d_{12} = d_{23} = 1$ e $d_{13} = 2$. Então, uma possível realização sobre a linha real \mathbb{R} que satisfaça todas as distâncias é $x_1 = 0$, $x_2 = 1$ e $x_3 = 2$. Uma forma de obter o valor de x_3 dado os valores de x_1 e x_2 e as distâncias d_{13} e d_{23} é a *trilateração*: sabendo que $d_{13} = \|x_3 - x_1\| = 2$ e $d_{23} = \|x_3 - x_2\| = 1$, tem-se

$$x_3^2 - 2x_1x_3 + x_1^2 = 4 \quad \text{e} \quad (7)$$

$$x_3^2 - 2x_2x_3 + x_2^2 = 1. \quad (8)$$

Subtraindo a equação 8 da 7, obtêm-se

$$2(x_1 - x_2)x_3 = x_1^2 - x_2^2 - 3 \Rightarrow 2x_3 = 4 \Rightarrow x_3 = 2.$$

E pode-se generalizar esse exemplo facilmente para $(K+1)$ -cliques em \mathbb{R}^{K-1} [1]: Precisa-se conhecer a posição de K vértices e as distâncias destes ao $(K+1)$ -ésimo vértice, assim, pode-se realizar o $(K+1)$ -ésimo vértice, em tempo linear, resolvendo um sistema de K equações como acima.

Isto é, sejam $x_1, \dots, x_k \in \mathbb{R}^{K-1}$ as posições para K vértices de um $(K+1)$ -clique e, para todo $j \leq K$, seja $d_{j,K+1}$ a distância associada com a aresta $\{j, K+1\}$. Seja $y \in \mathbb{R}^{K-1}$ a posição do $(K+1)$ -ésimo vértice; então, y deve respeitar as K equações quadráticas $\forall j \leq K$, $\|y - x_j\|^2 = d_{j,K+1}^2$, com $K-1$ incógnitas y_1, \dots, y_{k-1} :

$$\begin{cases} \|y\|^2 - 2x_1y + \|x_1\|^2 = d_{1,K+1}^2 \\ \vdots \\ \|y\|^2 - 2x_Ky + \|x_K\|^2 = d_{K,K+1}^2 \end{cases} \quad (9)$$

Para qualquer $h \leq K$, seja e_h a h -ésima equação no sistema de equações 9: pode-se tomar as diferenças e formar um novo sistema $\forall h < K$ ($e_h - e_K$) contendo $K - 1$ equações com $K - 1$ incógnitas:

$$\begin{cases} 2(x_1 - x_K) \cdot y = \|x_1\|^2 - \|x_K\|^2 - d_{1,K+1}^2 + d_{K,K+1}^2 \\ \vdots \\ 2(x_{K-1} - x_K) \cdot y = \|x_{K-1}\|^2 - \|x_K\|^2 - d_{K-1,K+1}^2 + d_{K,K+1}^2 \end{cases} \quad (10)$$

Note que o sistema de equações 10 é um *sistema linear* da forma

$$Ay = b, \quad (11)$$

onde $A = (2a_{ij})$ é uma matriz quadrada $(K - 1) \times (K - 1)$ com $a_{ij} = x_{ij} - x_{Kj}$ para todo $i, j < K$, e $b = (b_1, \dots, b_{K-1})^T$ com $b_i = \|x_i\|^2 - \|x_K\|^2 - d_{i,K+1}^2 + d_{K,K+1}^2$ para todo $i < K$.

Diferentes métodos para solução de sistemas lineares como esse são encontrados na bibliografia [20, 1] — no geral, a escolha do melhor depende de propriedades da matriz A , como, por exemplo, quão esparsa ela é.

Em especial, se A não é uma matriz singular, então ela possui uma inversa A^{-1} . Logo, podemos obter a posição do $(K + 1)$ -ésimo vértice da forma

$$Ay = b \Rightarrow A^{-1}Ay = A^{-1}b \Rightarrow x_{K+1} = y = A^{-1}b. \quad (12)$$

Porém, se A é singular, isso quer dizer que as linhas $a_i = x_i - x_K$ (para $i < K$) não são todas linearmente independentes [20]. Essa situação mostra algumas propriedades geométricas interessantes [1]. Por exemplo, se $K = 2$, significa que $x_1 - x_2 = 0 \Rightarrow x_1 = x_2$, ou seja, que o segmento entre x_1 e x_2 é um simples ponto. Se estamos imersos no $\mathbb{R}^{K-1} = \mathbb{R}$ (i.e., a reta real), geometricamente, a situação é que x_3 está posicionado ou a direita ou a esquerda de $x_1 = x_2$, mas não se pode escolher.

Também, se $K = 3$, a singularidade de A implica que o triângulo definido por x_1 , x_2 e x_3 é um apenas um segmento no plano (caso o rank de A é 1) ou um simples ponto (caso o rank for 0). Nesse primeiro caso, x_4 pode estar posicionado em ambos os lados da linha que contém o segmento e, no segundo caso, x_4 pode estar em qualquer um dos pontos formados pela circunferência com centro $x_1 = x_2 = x_3$ e raio $d_{14} = d_{24} = d_{34}$. Essa característica geométrica vale para valores maiores de K [1]: a singularidade de A implica que há sempre múltiplas soluções para x_{K+1} .

Deve-se mencionar que, a partir da equação 9, podemos chegar no sistema linear 11, mas a volta não é verdadeira [1]. Em particular, se o sistema 9 tem uma solução, então o sistema 11 tem a mesma solução. Porém, mesmo que o sistema 9 não tenha solução, o sistema 11 sempre terá uma solução única — desde que A não seja singular. Por tanto, para verificar a factibilidade de uma solução x_{K+1} advinda do sistema linear 11, deve ser verificado se as distâncias aos K vértices foram respeitadas — ou seja, se $\|x_{K+1} - x_i\| = d_{i,(K+1)}$, para todo $i \leq K$.

Note que um K -simplex é uma realização de um $(K + 1)$ -clique em \mathbb{R}^K . No geral,

Referências

- [1] Leo Liberti and Carlile Lavor. *Euclidean Distance Geometry*. Springer, 2017.
- [2] Irineu Bicudo et al. *Os elementos*. Unesp, 2009.
- [3] Arthur Cayley. A theorem in the geometry of position. *Cambridge Mathematical Journal*, 2:267–271, 1841.
- [4] Karl Menger. Untersuchungen über allgemeine metrik. *Mathematische Annalen*, 100(1):75–163, 1928.
- [5] Leonard M Blumenthal. Theory and applications of distance geometry. 1953.
- [6] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. Euclidean distance geometry and applications. *Society for Industrial and Applied Mathematics*, 56(1):3–69, February 2014.
- [7] Manfred J Sippl and Harold A Scheraga. Cayley-menger coordinates. *Proceedings of the National Academy of Sciences*, 83(8):2283–2287, 1986.
- [8] Yechiam Yemini. Some theoretical aspects of position-location problems. In *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*, pages 1–8. IEEE, 1979.
- [9] Gordon M Crippen, Timothy F Havel, et al. *Distance geometry and molecular conformation*, volume 74. Research Studies Press Taunton, 1988.
- [10] Yechiam Yemini. The positioning problem-a draft of an intermediate summary. Technical report, UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST, 1978.
- [11] Deepak Tolani, Ambarish Goswami, and Norman I Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical models*, 62(5):353–388, 2000.
- [12] Jan de Leeuw and Willem Heiser. 13 theory of multidimensional scaling. *Handbook of statistics*, 2:285–316, 1982.
- [13] David L Nelson and Michael M Cox. *Lehninger principles of biochemistry*. W.H.Freeman and Company, 2013.
- [14] CC Lavor. *Uma abordagem determinística para minimização global da energia potencial de moléculas*. PhD thesis, PhD thesis, COPPE/UFRJ, Rio de Janeiro, 2001.
- [15] GN Ramachandran, AS Kolaskar, C Ramakrishnan, and V Sasisekharan. The mean geometry of the peptide unit from crystal structure data. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 359(2):298–302, 1974.
- [16] Andreas Savvides, Chih-Chieh Han, and Mani B Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 166–179. ACM, 2001.

- [17] C. Lavor, N. Maculan, M. Souza, and R. Alves. *Álgebra e Geometria no Cálculo de Estrutura Molecular*. IMPA, Rio de Janeiro, RJ, 31^o colóquio brasileiro de matemática edition, 2017.
- [18] James B Saxe. Embeddability of weighted graphs in k-space is strongly np-hard. In *Proc. of 17th Allerton Conference in Communications, Control and Computing, Monticello, IL*, pages 480–489, 1979.
- [19] Riccardo Benedetti and Jean-Jacques Risler. In real algebraic and semi-algebraic sets (1990). *Berlin, Hermann, Paris*.
- [20] Elon Lages Lima. *Álgebra Linear*. SBM, Rio de Janeiro : IMPA, 1a edition, 2014.
- [21] Alfredo Steinbruch and Paulo Winterle. *Geometria Analítica*. Makron Books, São Paulo, SP, 2a edition, 1987.

A Métricas

Como esse texto utiliza fortemente o conceito de distância, é necessário e bem vindo que se gaste algum espaço para uma construção formal dessa ideia. A noção de distância está relacionada com o conceito de *métrica*, como segue.

Seja \mathcal{X} um espaço vetorial K -dimensional sobre \mathbb{R} . *Métrica* é uma função de dois argumentos que mapeia pares ordenados de elementos em \mathcal{X} para um número real não negativo. Precisamente, para todo x, y e $z \in \mathcal{X}$, uma função $d(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ é uma métrica se satisfaz os seguintes axiomas:

1. $d(x, y) = 0$ se, e somente se, $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, z) \leq d(x, y) + d(y, z)$;
4. $d(x, y) \geq 0$

Nesse trabalho, quando não é especificado qual métrica se está usando, fica implícita a utilização da *Métrica Euclidiana*, definida em função da *Norma Euclidiana*:

$$\forall x, y \in \mathcal{X}, d(x, y) = \|x - y\|_2 = \sqrt{\langle x - y, x - y \rangle} = \sqrt{\sum_{i=1}^K (x_i - y_i)^2}. \quad (\text{Norma Euclidiana})$$

O par (\mathcal{X}, d) é chamado *espaço métrico*. A noção de métrica não depende de espaços vetoriais, donde pode ser facilmente generalizada fazendo \mathcal{X} um conjunto qualquer.

B Lei dos Cos e Ângulos Entre dois Vetores no \mathbb{R}^3

A lei dos cossenos é uma propriedade trigonométrica válida para qualquer triângulo, permitindo encontrar o valor de um dos seus lados conhecendo apenas os outros lados e um ângulo. Porém, aqui utilizaremos a ideia reversa, onde, nesse caso, saberemos os lados e queremos descobrir os ângulos.

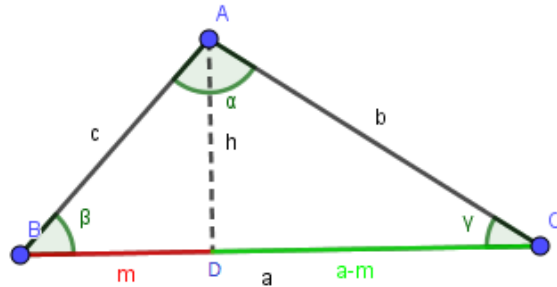


Figura 2: Triângulo para ilustrar a lei dos cossenos.

- **Demonstração Leis dos Cossenos:**

Dado um triângulo qualquer, traça-se uma altura relativa ao lado a . Aplicando o *Teorema de Pitágoras* no $\triangle ABD$:

$$c^2 = m^2 + h^2 \rightarrow h^2 = c^2 - m^2 \quad (13)$$

Aplicando novamente *Pitágoras*, porém, em $\triangle ADC$, obtemos:

$$b^2 = h^2 + (a - m)^2 \quad (14)$$

Substituindo na equação 14 o valor de h^2 obtido em 13:

$$b^2 = c^2 - m^2 + a^2 - 2am + m^2$$

$$b^2 = c^2 + a^2 - 2am$$

Analisando a Figura 2, pode-se perceber que $\frac{m}{c} = \cos \beta$, então:

$$b^2 = c^2 + a^2 - 2ac \cos \beta$$

Analogamente, obtém-se:

$$c^2 = a^2 + b^2 - 2ab \cos \gamma$$

$$a^2 = b^2 + c^2 - 2bc \cos \alpha$$

Note também que se o argumento dos cossenos for $\frac{\pi}{2}$ recaímos no Teorema de Pitágoras. ■

- **Ângulos Entre 2 Vetores:**

Sejam dois vetores \vec{u} e $\vec{v} \in \mathbb{R}^2$, representados na Figura 3

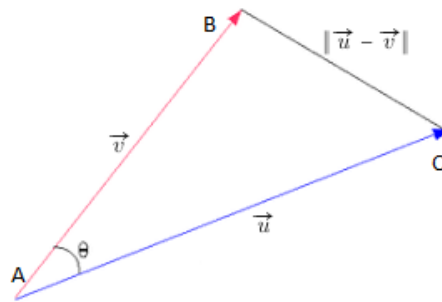


Figura 3: Diferença entre vetores u e v

Para encontrarmos o ângulo θ utilizaremos a lei dos cossenos aplicada a $\triangle ABC$:

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos \theta \quad (15)$$

Utilizando a definição do produto escalar [21]

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\vec{u} \cdot \vec{v} \quad (16)$$

Comparando a equação 15 com a 16, obtemos trivialmente

$$\|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos\theta = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\vec{u}\cdot\vec{v}$$

$$\vec{u}\cdot\vec{v} = \|\vec{u}\|\|\vec{v}\|\cos\theta$$

Logo,

$$\cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}$$

■