

Geometria de Distâncias

Guilherme Philippi

5 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	Sobre os Diferentes Problemas em GD	4
1.3.1	Conformações Moleculares	4
1.3.2	Localização de Sensores	5
1.3.3	Movimentos em Cinemática Inversa	5
1.3.4	Escalonamento Multidimensional	6
	Referências	7
A	Métricas	9
B	Lei dos Cos e Ângulos Entre dois Vetores no \mathbb{R}^3	9

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 (GD).

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 $i, j \leq n$.

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

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].

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.

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].

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 Sobre os Diferentes Problemas em GD

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 GD, apresenta-se parte desse estudo, visando organizar os conceitos.

As principais aplicações em GD 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 *escalonamento 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);
- 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

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].

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 $K = 3$.

1.3.3 Movimentos 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 GD, 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).

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] 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*:

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

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.

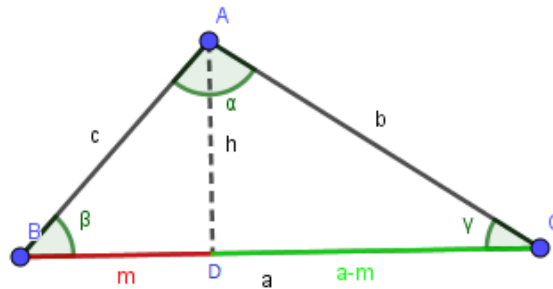


Figura 1: Triângulo para ilustrar a lei dos cossenos.

- **Demonstração Leis dos Cossenos:**

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

$$c^2 = m^2 + h^2 \rightarrow h^2 = c^2 - m^2 \quad (4)$$

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

$$b^2 = h^2 + (a - m)^2 \quad (5)$$

Substituindo na equação 5 o valor de h^2 obtido em 4:

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

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

Analisando a Figura 1, 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 2

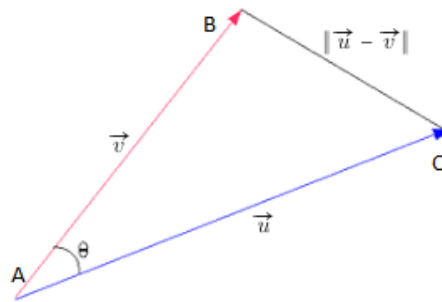


Figura 2: Diferença entre vetores u e v

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

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos \theta \quad (6)$$

Utilizando a definição do produto escalar [17]

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\vec{u} \cdot \vec{v} \quad (7)$$

Comparando a equação 6 com a 7, 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,

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

■