

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICA**

Letícia Figueredo de Carvalho

MATRIZES DE DISTÂNCIAS EUCLIDIANAS

Florianópolis

2018

Letícia Figueredo de Carvalho

MATRIZES DE DISTÂNCIAS EUCLIDIANAS

Trabalho de Conclusão de Curso submetido ao Departamento de Matemática para a obtenção do Grau de Licenciada em Matemática.

Orientador: Prof. Dr. Douglas Soares Gonçalves

Florianópolis

2018

Catálogo na fonte elaborada pela biblioteca da
Universidade Federal de Santa Catarina

A ficha catalográfica é confeccionada pela Biblioteca Central.

Tamanho: 7cm x 12 cm

Fonte: Times New Roman 9,5

Maiores informações em:

<http://www.bu.ufsc.br/design/Catalogacao.html>

Letícia Figueredo de Carvalho

MATRIZES DE DISTÂNCIAS EUCLIDIANAS

Este Trabalho de Conclusão de Curso foi julgado aprovado para a obtenção do Título de “Licenciada em Matemática”, e aprovado em sua forma final pelo Departamento de Matemática

Florianópolis, 14 de Novembro 2018

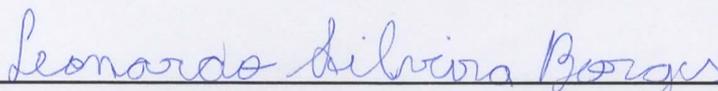


Prof. Dra. Sonia Elena Palomino Castro
Coordenador do Curso

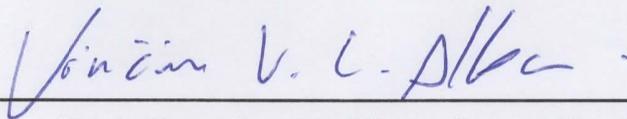
Banca Examinadora:



Prof. Dr. Douglas Soares Gonçalves
Orientador



Prof. Dr. Leonardo Silveira Borges



Prof. Dr. Vinícius Viana Luiz Albani

AGRADECIMENTOS

Agradeço primeiramente aos meus pais Marnes Filgueira de Carvalho Júnior e Maria Terezinha Figueredo de Carvalho e a minha irmã Luísa Figueredo de Carvalho por todo o apoio dado a mim todos esses anos. Agradeço também à minha amiga Luísa Maciel da Costa por estar na minha vida desde muito nova e por sempre acreditar e confiar em mim. Agradeço também ao meu namorado Bruno Teixeira da Motta com toda a ajuda na construção dos códigos dos algoritmos e por todo o apoio, carinho, confiança e orgulho que tem de mim.

Agradeço também aos amigos que o curso de matemática me proporcionou e, principalmente agradeço à Bruna da Silva Donadel, Victória Foyes Gittens, André Borges Carlos, Ana Carolina Altomani por terem vivido intensamente comigo esses 4 anos de graduação.

Agradeço aos membros da banca por aceitarem o convite de avaliar o meu trabalho e agradeço ao professor Douglas Soares Gonçalves por me orientar e ter me dado a oportunidade de trabalhar com ele. Agradeço ainda à professora Silvia Janesch por ter me aberto as portas para ingressar no curso de matemática na UFSC e, ainda, agradeço aos professores Rubens Starke e Luciano Bedin por terem me apresentado a matemática e enchido meu coração de paixão por tal carreira.

RESUMO

O problema de geometria de distâncias (PGD) consiste em determinar a posição de um conjunto de objetos em um dado espaço euclidiano, com base nas distâncias conhecidas entre alguns pares desses objetos. Este problema, que encontra aplicações em bioquímica, robótica, estatística e telecomunicações, está intimamente relacionado ao conceito de matrizes de distâncias euclidianas. Este trabalho apresenta um algoritmo de reconstrução sequencial que, explorando submatrizes de distâncias euclidianas de uma matriz de distâncias possivelmente incompleta, é capaz de resolver uma classe particular de PGDs que satisfazem certas hipóteses. Também estuda-se um algoritmo de reordenação do conjunto de objetos capaz de encontrar, quando existente, uma ordem que satisfaça as hipóteses necessárias ao algoritmo de reconstrução. Experimentos numéricos mostram que os algoritmos propostos são eficientes e apresentam boa precisão na reconstrução de conjuntos com milhares de pontos com base em uma lista incompleta de distâncias.

Palavras-chave: Geometria de Distâncias. Matrizes de Distâncias Euclidianas. Algoritmo de Reconstrução Geométrica Sequencial. Algoritmo de Ordenação.

ABSTRACT

The distance geometry problem (DGP) consists in determining the position of a set of objects in a given Euclidean space, based on the known distances between pairs of objects. This problem, which finds applications in biochemistry, robotics, statistics and telecommunications, is closely related to the concept of euclidean distance matrix. This work presents a sequential reconstruction algorithm which, by exploiting submatrices of an (incomplete) euclidean distance matrix, is able to solve a particular class of DGPs when some suitable assumptions hold. Moreover, we study an ordering algorithm which is able to identify an order for the objects such that the assumptions required by the reconstruction algorithm are satisfied (when one exists). Numerical experiments show that the proposed algorithms are efficient and present good precision in the reconstruction of sets with thousands of points based on an incomplete distance list.

Keywords: Distance Geometry. Euclidean Distance Matrix. Sequential Reconstruction Algorithm. Ordering Algorithm.

SUMÁRIO

1 INTRODUÇÃO	13
2 OBJETIVOS	15
3 METODOLOGIA	17
4 CONCEITOS PRELIMINARES	19
4.1 NOÇÕES SOBRE GRAFOS	19
4.2 ALGUNS RESULTADOS DE ÁLGEBRA LINEAR	20
5 PROBLEMA DE GEOMETRIA DE DISTÂNCIAS ..	23
5.1 DEFINIÇÃO	23
5.2 SOLUÇÕES DO PGD	23
6 MATRIZES DE DISTÂNCIAS EUCLIDIANAS	29
6.1 MATRIZ DE GRAM	32
6.2 O TEOREMA DE SCHOENBERG	33
7 SOBREPOSIÇÃO DE CONJUNTOS DE PONTOS ..	39
7.1 CENTRALIZAÇÃO DE PONTOS AO REDOR DA ORIGEM	39
7.2 PROBLEMA DE PROCRUSTES ORTOGONAL	42
8 ALGORITMO DE RECONSTRUÇÃO GEOMÉTRICA SEQUENCIAL	43
8.1 HIPÓTESES	43
8.2 DESCRIÇÃO DO ALGORITMO	44
8.3 EXPERIMENTOS NUMÉRICOS	45
9 ORDENAÇÃO DO CONJUNTO DE VÉRTICES	49
9.1 HIPÓTESES	49
9.2 DESCRIÇÃO DO ALGORITMO DE ORDENAÇÃO	49
9.3 EXPERIMENTOS NUMÉRICOS	51
10 CONSIDERAÇÕES FINAIS	55
REFERÊNCIAS	57

1 INTRODUÇÃO

O problema de geometria de distâncias (PGD) consiste em determinar a posição de um conjunto de pontos em um dado espaço \mathbb{R}^k , com base nas distâncias conhecidas entre alguns pares de pontos (LAVOR; LIBERTI, 2014).

Além da teoria matemática associada à Geometria de Distâncias (GD), o interesse por esse tópico de pesquisa explica-se pela riqueza e variedade de suas aplicações. Pode-se citar, como exemplo, aplicações em astronomia, bioquímica, estatística, nanotecnologia, robótica e telecomunicações. (LAVOR; LIBERTI, 2014)

Em astronomia, o problema está relacionado à determinação da posição de estrelas, utilizando informação de distâncias entre elas.

Em bioquímica, o problema aparece na determinação de estruturas tridimensionais de moléculas de proteínas, utilizando dados de ressonância magnética nuclear. Em estatística, o problema está relacionado à visualização de dados e redução dimensional. Essa aplicação também está ligada a um tema atual de pesquisa, denominado Big Data.

Em nanotecnologia, o problema possui semelhanças com o problema em bioquímica, mas só que em escala “nano”. Em robótica, o problema aparece no posicionamento dos braços de um robô para realizar uma tarefa específica.

Apesar desta vasta lista de aplicações para o PGD, o presente trabalho vai focar na aplicação referente à telecomunicações, ou seja, o interesse é em encontrar a localização dos sensores de rede, em que alguns pontos são previamente fixados em situações em que nem todas as distâncias são conhecidas.

No que se refere à parte teórica do estudo, quando *todas* as distâncias são conhecidas de forma *exata*, o problema é de fácil resolução. O problema maior surge quando nem todas as distâncias são conhecidas. Nesse caso, surge a pergunta: é possível encontrar a posição exata de todos esses pontos, satisfazendo as distâncias conhecidas, em um espaço Euclidiano de dimensão apropriada?

Este trabalho tem como objetivo apresentar o Problema de Geometria de Distâncias na teoria e, também, construir algoritmos de resolução para uma classe de problemas práticos modelados pelo PGD. Este trabalho está organizado da seguinte forma. No Capítulo 2 apresentam-se os objetivos do trabalho. O Capítulo 3 trata da metodologia utilizada. O Capítulo 4 traz algumas definições básicas de teoria de grafos

e alguns conceitos de álgebra linear necessários para os estudos realizados. A definição do PGD e uma discussão a respeito da cardinalidade do seu conjunto solução é apresentada no Capítulo 5. As definições de matriz de distâncias, matriz de distâncias euclidianas e matriz de Gram, bem como as relações entre elas, estão no Capítulo 6. O Capítulo 7 trata da superimposição de conjuntos de pontos, falando sobre o problema de centralização de pontos ao redor da origem e do Problema de Procrustes Ortogonal. Nos Capítulos 8 e 9 serão vistos os algoritmos de reconstrução geométrica sequencial e ordenação de conjunto de vértices, respectivamente. Ambos os capítulos trazem as hipóteses dos algoritmos, suas descrições e também os experimentos numéricos realizados para comprovar suas aplicabilidades. O Capítulo 10 traz as considerações finais.

2 OBJETIVOS

O presente trabalho tem como objetivo apresentar ao leitor o Problema de Geometria de Distâncias e a teoria referente a matrizes de distâncias euclidianas que permitirá determinar se o problema tem solução e, ainda, desenvolver algoritmos para resolução de alguns problemas práticos.

Sendo assim, este trabalho conterà os seguintes tópicos para a realização do objetivo:

- Revisão dos conceitos necessários de Álgebra Linear
- Revisão da literatura básica sobre Geometria de Distâncias
- Desenvolvimento de um Algoritmo de Reconstrução Sequencial e do Algoritmo de Ordenação de Conjunto de Pontos
- Implementação e testes computacionais dos algoritmos em problemas artificiais com um conjunto incompleto de distâncias

3 METODOLOGIA

O trabalho utilizará a metodologia descritiva, com abordagem qualitativa. Para os conceitos de álgebra linear mais relevantes foram utilizadas como bibliografias: (LIMA, 2009), (STRANG, 2011), (MEYER, 2000) e (GOLUB; LOAN, 1996). Na parte do estudo do PGD e de matrizes de distâncias euclidianas o estudo foi baseado principalmente nas bibliografias (LIBERTI et al., 2012) (MUCHERINO et al., 2012) (LAVOR; LIBERTI, 2014) (DOKMANIC et al., 2015) (GONÇALVES, to appear, 2017). O assunto deste Trabalho de Conclusão de Curso foi estudado por um período de um ano e meio (sendo o primeiro ano parte da iniciação científica).

A rotina de estudo aconteceu da seguinte maneira. O primeiro semestre (2017.2) do trabalho foi voltado a estudos teóricos; foram revisados os assuntos de álgebra linear necessários e um pouco da teoria do problema de geometria de distâncias e matrizes de distâncias euclidianas. Ainda, todas as demonstrações foram desenvolvidas também nesse período.

O segundo semestre de estudos foi voltado ao algoritmo de reconstrução geométrica sequencial. O foco foi utilizar toda a argumentação teórica já consolidada anteriormente para a criação do algoritmo e realização dos experimentos numéricos para comprovação de aplicabilidade do algoritmo em alguns problemas práticos.

No terceiro semestre focou-se em desenvolver o algoritmo de ordenação de conjuntos de pontos e na conclusão da escrita do TCC. Esse outro algoritmo foi criado para auxiliar na resolução de problemas que possuem hipóteses mais fracas que as hipóteses do algoritmo inicial.

Sendo assim, a partir de novos problemas gerados e modificando um pouco o algoritmo criado no semestre anterior, foram realizados experimentos numéricos e a aplicabilidade do algoritmo geométrico em conjunto com o algoritmo de ordenação (pré-processamento) foi comprovada. Durante todo o estudo foram realizadas reuniões semanais com o orientador apresentando os resultados obtidos e/ou estudos realizados de uma semana para a outra.

4 CONCEITOS PRELIMINARES

Na parte teórica do trabalho foram revisados alguns conceitos de álgebra linear que servem de base para os estudos realizados baseados nas referências (LIMA, 2009) e (STRANG, 2011). Os estudos começaram a partir do conceito de grafo, uma vez que os problemas modelados pelo PGD são representados através de grafos. As definições abaixo estão baseadas em (LAVOR; LIBERTI, 2014).

4.1 NOÇÕES SOBRE GRAFOS

Definição 1. *Um grafo é formado por um conjunto $V \neq \emptyset$ e um conjunto E formado por pares não ordenados de elementos de V . Nesse caso usa-se a notação $G = (V, E)$.*

Os elementos de V são chamados de vértices e os elementos de E são chamados de arestas. Denota-se cada aresta por $\{v, w\}; v, w \in V$.

Note que nada foi dito a respeito de v ser diferente de w , logo, as arestas podem ser do tipo $\{v, v\}$, nesse caso, essa aresta é dita um *laço* ou um *loop*.

Definição 2. *Um grafo é dito simples quando os vértices de uma aresta são elementos distintos de V .*

Definição 3. *Um grafo é dito conectado quando não é possível separar o conjunto de seus vértices V em dois subconjuntos A e B com $A \cap B = \emptyset$ e $V = A \cup B$, de tal maneira que não exista nenhuma aresta $\{a, b\} \in E$ em que $a \in A$ e $b \in B$.*

Definição 4. *Um grafo é dito não orientado quando dados $a, b \in V$, $\{a, b\} = \{b, a\}$.*

Definição 5. *Um grafo é dito completo se seu conjunto de arestas possui todos os pares não-ordenados possíveis.*

Obs: dado V , a quantidade de arestas possíveis em um grafo completo é dada por:

$$\binom{|V|}{2} = \frac{|V|(|V| - 1)}{2}, \text{ em que } |V| \text{ é a cardinalidade de } V.$$

Definição 6. Um grafo é dito ponderado quando é associado um número real a cada aresta desse grafo, ou seja, tem-se uma aplicação $d: E \rightarrow \mathbb{R}$. Nesse caso usa-se a notação $G = (V, E, d)$.

Definição 7. Um clique em um grafo $G = (V, E)$ é um outro grafo $G' = (V', E')$ tal que $V' \subset V$ e $E' \subset E$ e G' é completo.

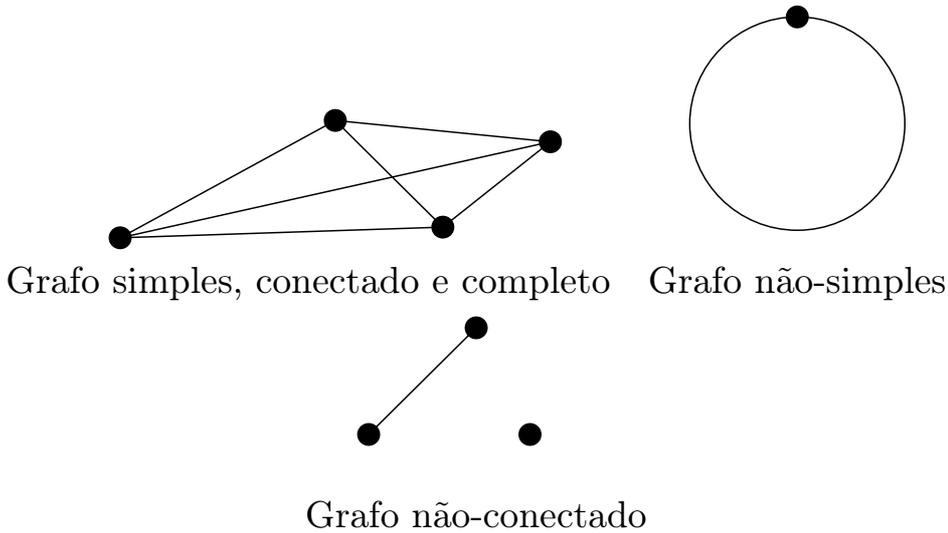


Figura 1: Exemplos de Grafos

Agora que já foi definido o conceito e alguns tipos de grafos, seguem mais algumas definições e relações que serão úteis no decorrer do texto.

4.2 ALGUNS RESULTADOS DE ÁLGEBRA LINEAR

Primeiro recorda-se da relação entre Norma Euclidiana $\|\cdot\|_2$ e o produto interno $\langle \cdot, \cdot \rangle$ entre vetores de \mathbb{R}^k . Denota-se tais vetores como vetores colunas e x^T representa o transposto do vetor x .

Dados x_i e x_j em \mathbb{R}^k , tem-se: $\|x_i\|_2 = \sqrt{\langle x_i, x_i \rangle}$, ou seja, $\|x_i\|_2^2 = \langle x_i, x_i \rangle = x_i^T x_i$, e, além disso,

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 - 2\langle x_i, x_j \rangle + \|x_j\|_2^2. \quad (4.1)$$

Definição 8. A norma de Frobenius de uma matriz A é: $\|A\|_F^2 = \text{tr}(A^T A) = \sum_i \sum_j a_{ij}^2$, em que $\text{tr}(B)$ é o traço de B .

5 PROBLEMA DE GEOMETRIA DE DISTÂNCIAS

A seguir, define-se o problema fundamental de Geometria de Distâncias, segundo (LAVOR; LIBERTI, 2014).

5.1 DEFINIÇÃO

Definição 11. *Dado um inteiro $k > 0$ e $G = (V, E, d)$, G simples, conectado, não orientado e ponderado com pesos positivos, resolver o PGD significa encontrar uma aplicação x que associa cada vértice do grafo a uma posição no espaço \mathbb{R}^k . Ou seja, o interesse é encontrar $x : V \rightarrow \mathbb{R}^k$ tal que*

$$\forall \{u, v\} \in E, \|x(u) - x(v)\|_2 = d(u, v). \quad (5.1)$$

Resolver o problema é, então, associar cada vértice do grafo a um único ponto em \mathbb{R}^k satisfazendo (5.1), ou seja, ao posicionar os vértices $u, v \in V$ em \mathbb{R}^k , quer-se que a distância calculada $\|x(u) - x(v)\|_2$ coincida com o valor dado $d(u, v)$ (peso da aresta), para todo $\{u, v\} \in E$. Note que da definição tem-se que o peso das arestas é a distância correspondente entre dois vértices do grafo.

A partir de agora o trabalho tratará sempre com a norma Euclidiana ($\|\cdot\| = \|\cdot\|_2$), a não ser quando mencionado o contrário.

5.2 SOLUÇÕES DO PGD

Uma solução x para o PGD é chamada *realização*. Com uma dada solução do PGD, pode-se obter uma quantidade infinita (não enumerável) de realizações válidas distintas através de rotações, translações e reflexões totais. Eliminando, então, as soluções obtidas dessa forma, tem-se que o conjunto solução de um PGD pode ser vazio, finito ou infinito não-enumerável (LAVOR; LIBERTI, 2014).

Para ilustrar as possíveis soluções do PGD, considere os exemplos a seguir.

Exemplo 1. Considere o conjunto de vértices r, u, v tais que $d_{rv} = 1$ e $d_{uv} = 1$ e $d_{ru} = 10$. Existe solução em alguma dimensão k para esse conjunto de pontos?

Pense primeiro em \mathbb{R} . Note nesse caso que para fixar-se dois pontos em \mathbb{R} tais que a distância entre eles seja 1 basta pegarmos quaisquer dois pontos com essa distância (veja figura abaixo).

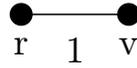


Figura 2: pontos r e v fixados

Agora, tente fixar o ponto u de forma que as distâncias de u até os pontos r e v sejam satisfeitas. Pode-se tentar fixar u da seguinte maneira:



Figura 3: $d_{ru} = 10$ e $d_{vu} = 9$

Note que nessa representação, a distância entre r e u é satisfeita mas a distância entre v e u não é satisfeita e, portanto, não é possível fixar u nessa posição. Tente fixar então u à esquerda do ponto r , nesse caso tem-se:



Figura 4: $d_{ru} = 10$ e $d_{vu} = 11$

Note então que nesse caso também é possível satisfazer a distância entre r e u , mas a distância entre v e u não é satisfeita. Dessa forma, concluí-se que não é possível fixar em \mathbb{R} um terceiro ponto u satisfazendo as distâncias dadas.

A pergunta natural que surge agora é: será que existe solução para esse problema em alguma dimensão $k > 1$?

Tentando encontrar a solução em \mathbb{R}^2 , pode-se também fixar dois dos três pontos tais que as distâncias entre eles seja 1. Fixe, sem perda de generalidade, o ponto r na origem do plano cartesiano, ou seja,

$r = (0, 0)$. Agora, pode-se fixar o próximo ponto v em qualquer lugar na circunferência de raio 1 pois em qualquer ponto dessa circunferência tem-se $\|r - v\| = 1$. Fixe então $v = (1, 0)$.

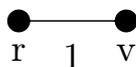


Figura 5: pontos r e v fixados

Já foi possível fixar dois pontos que satisfaçam as distâncias dadas, agora o problema se resume a fixar o terceiro ponto u tal que $\|r - u\| = 10$ e $\|v - u\| = 1$. Note que encontrar a posição para o terceiro ponto de forma a satisfazer todas as distâncias significa “fechar” um triângulo em \mathbb{R}^2 com lados medindo 1, 1 e 10. Mas isso não ocorre pois essas medidas violam a desigualdade triangular (veja figura 6).

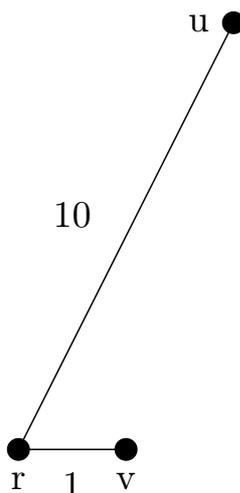


Figura 6: pontos r, u e v fixados

Conclui-se dessa forma que não há solução para o problema em \mathbb{R}^2 e, conseqüentemente, não há solução para o problema em nenhuma dimensão $k > 3$ pois em dimensões maiores apenas acrescentariam-se componentes nulas nas entradas dos vetores.

Sendo assim, o conjunto solução é vazio para todo $k > 0$ e o problema não possui solução.

Note que nesse exemplo se tivesse sido definido $d_{ru} = \sqrt{2}$, ter-se-ia encontrado a posição do vértice u e, conseqüentemente, encontraria-se uma solução única para o problema a menos de translações, rotações e reflexões totais e, nesse caso, tem-se um conjunto solução finito (veja

figura abaixo).

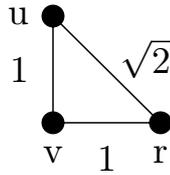


Figura 7: pontos r, u, v fixados

Exemplo 2. Considere novamente o conjunto de vértices r, u, v só que agora tem-se que $d_{rv} = 1$ e $d_{ru} = 4$ e a distância d_{uv} é desconhecida. Nesse caso, pergunta-se: existe solução em alguma dimensão k para esse conjunto de pontos?

Pense primeiro de novo na dimensão 1. Nesse caso, posicione de novo os pontos r e v conforme o exemplo anterior.

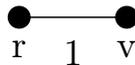


Figura 8: pontos r e v fixados

Agora precisa-se posicionar o ponto u de forma a satisfazer $d_{ru} = 4$. Como a dimensão de interesse é 1, pode-se posicionar o ponto u à esquerda ou à direita de r , conforme figura abaixo.

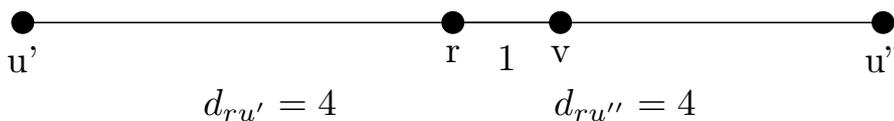


Figura 9: pontos r, u, v fixados

Considere u' e u'' como sendo a posição de u à esquerda de r e a posição de u à direita de r , respectivamente. Note que se o leitor considerar u' , chega-se em $d_{uv} = 5$ e se o leitor considerar u'' chega-se em $d_{uv} = 3$.

Portanto, dado o conjunto de vértices r, u, v , o problema em \mathbb{R} possui duas soluções (pois a distância entre u e v não foi previamente fixada).

Agora, pense no caso em \mathbb{R}^2 . Fixe novamente os pontos r e v .

Agora note que como $d_{ru} = 4$, pode-se posicionar u em qualquer ponto da circunferência de centro em r e raio 4, conforme figura abaixo.

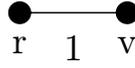


Figura 10: pontos r e v fixados

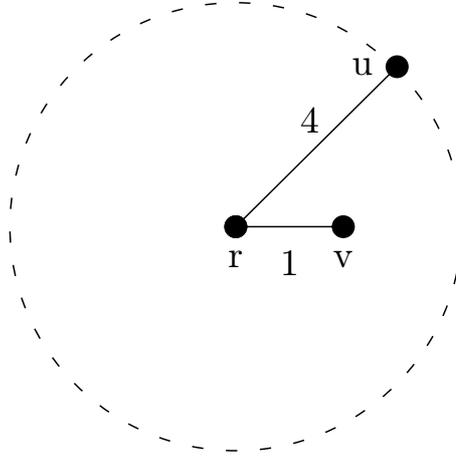


Figura 11: pontos r, u e v fixados

Sendo assim, note que pelo fato de d_{uv} ser desconhecida, o ponto u pode percorrer toda a circunferência e nesse caso tem-se infinitas soluções para o problema e, portanto, o conjunto solução é infinito não-enumerável. Note também que para dimensões maiores continuaria-se tendo infinitas soluções pois só acrescentaria-se 0 nas coordenadas dos pontos.

Exemplo 3. *Seja agora o seguinte conjunto de vértice em \mathbb{R}^2 : u, v, r tais que $d_{uv} = d_{ur} = d_{vr} = 1$. Pode-se fixar os pontos por exemplo, da forma conforme figura abaixo:*

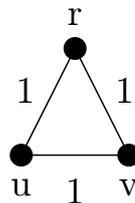


Figura 12: Pontos u, v, r fixados

Nesse caso, note que a desigualdade triangular é satisfeita estritamente e tem-se solução única, conforme a figura. Mas note também que não existe solução em \mathbb{R} pois não é possível posicionar esses 3 pontos em \mathbb{R} de forma que todas as distâncias sejam satisfeitas.

Exemplo 4. *Pense agora no seguinte conjunto de pontos em \mathbb{R}^2 dados*

por: u, v, r, s tais que $d_{uv} = d_{ur} = d_{vs} = d_{us} = d_{vr} = d_{rs} = 1$. Será que é possível encontrar a posição dos pontos em \mathbb{R}^2 tais que todas essas distâncias sejam satisfeitas?

Pode-se fixar os pontos u, v, r da seguinte forma, conforme figuras abaixo:

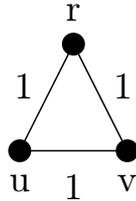


Figura 13: Pontos r, u, v fixados

Perceba agora então que não há como posicionar em \mathbb{R}^2 o ponto s de forma que as distâncias sejam satisfeitas. Por exemplo, não é possível colocá-lo na seguinte posição:

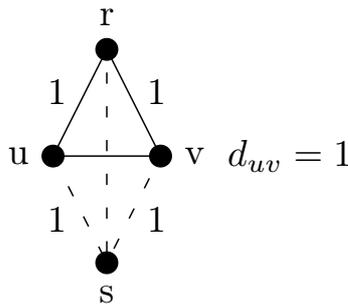


Figura 14: Tentativa de fixar s

Note que nesse caso, $d_{rs} = 1$ não é satisfeita. Por um raciocínio análogo, qualquer tentativa de fixar s não satisfaz alguma das distâncias dadas no enunciado do exemplo. Sendo assim, não há realização em \mathbb{R}^2 para esse conjunto de pontos.

No entanto, se \mathbb{R}^3 for considerado, é possível encontrar a posição de todos os pontos satisfazendo todas as distâncias dadas. Basta para isso posicionar os pontos na estrutura de um tetraedro regular de aresta 1. Conclui-se que nesse exemplo não encontra-se solução para \mathbb{R} ou \mathbb{R}^2 , mas considerando \mathbb{R}^3 ou dimensões $k > 3$ encontra-se solução única.

Considerando os exemplos acima, é possível perceber que tanto a quantidade de distâncias conhecidas, o valor dessas distâncias e a dimensão escolhida para a realização influenciam na solução do problema. Os exemplos mostraram problemas sem solução, com solução única e com infinitas soluções.

6 MATRIZES DE DISTÂNCIAS EUCLIDIANAS

Dando continuidade ao estudo do PGD, segue agora a definição de matrizes de distâncias Euclidianas (em inglês, Euclidean Distance Matrix - EDM). A teoria envolvendo estas matrizes será fundamental para o desenvolvimento de algoritmos para classes especiais do PGD. As definições e resultados deste capítulo estão baseadas em (DOKMANIC et al., 2015).

Definição 12. *Uma matriz $M \in \mathbb{R}^{n \times n}$ simétrica com entradas não negativas e diagonal nula é chamada de **matriz de distâncias**.*

Exemplo 5. $M_1 = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \end{bmatrix}$, $M_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$ e $M_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

são matrizes de distâncias.

Definição 13. *Dada uma matriz $D \in \mathbb{R}^{n \times n}$ de distâncias, D é uma **matriz de distâncias Euclidianas** (ou EDM) se para algum inteiro $k > 0$, existem vetores $x_1, x_2, \dots, x_n \in \mathbb{R}^k$ tais que*

$$D_{ij} = \|x_i - x_j\|^2, \quad \forall i, j \in \{1, \dots, n\} \quad (6.1)$$

*Ainda, o menor k para o qual (6.1) é satisfeito é chamado **dimensão de realização**.*

Segue então das definições acima que, toda matriz de distâncias euclidianas é uma matriz de distâncias, mas a recíproca não é verdadeira.

Exemplo 6. *Considere a matriz de distâncias*

$$M = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \end{bmatrix}$$

Será que existem $x_1, \dots, x_n \in \mathbb{R}^k$, para algum $k > 0$ tal que $\forall i, j : \|x_i - x_j\|^2 = M_{ij}$?

Pense primeiro em \mathbb{R}^2 . Como a matriz M é quadrada de dimensão 3, o objetivo é tentar encontrar 3 vetores x_1, x_2 e $x_3 \in \mathbb{R}^2$ tais que:

$$\|x_1 - x_2\|^2 = 1 \quad (6.2)$$

$$\|x_1 - x_3\|^2 = 2 \quad (6.3)$$

$$\|x_2 - x_3\|^2 = 3 \quad (6.4)$$

Pode-se fixar sem perda de generalidade os pontos x_1 e x_2 da seguinte forma: $x_1 = (0, 0)^T$ e $x_2 = (1, 0)^T$. Note que fixados dessa forma, a equação (6.2) é satisfeita. Precisa-se agora encontrar x_3 que satisfaz as equações (6.3) e (6.4).

Ou seja,

$$\|x_3\|^2 - 2x_3^T x_1 + \|x_1\|^2 = 2, \quad (6.5)$$

$$\|x_3\|^2 - 2x_3^T x_2 + \|x_2\|^2 = 3. \quad (6.6)$$

Subtraindo (6.5) de (6.6) tem-se:

$$-2x_3^T (x_2 - x_1) + \|x_2\|^2 - \|x_1\|^2 = 1,$$

implicando em

$$(x_2 - x_1)^T x_3 = -\frac{1}{2}(1 + \|x_1\|^2 - \|x_2\|^2).$$

Mas, note que $\|x_1\|^2 = 0$ e $\|x_2\|^2 = 1$ e assim, segue que

$$(x_2 - x_1)^T x_3 = -\frac{1}{2}(1 - 1) = 0.$$

Agora, sabendo que $(x_2 - x_1)^T = [1 \ 0]$ e sendo $x_3 = \begin{bmatrix} a \\ b \end{bmatrix}$,

tem-se:

$$[1 \ 0] \begin{bmatrix} a \\ b \end{bmatrix} = 0 \Rightarrow a = 0 \text{ e } b \text{ é livre.}$$

Assim, concluí-se que $x_3 = \begin{bmatrix} 0 \\ b \end{bmatrix}$, $b \in \mathbb{R}$.

Até aqui, x_1, x_2 e x_3 acima são as soluções do sistema linear montado anteriormente. A pergunta que surge agora é: quais dessas soluções do sistema linear são soluções do PGD? Para isso, substitui-se x_3 encontrado em alguma das equações (6.2) ou (6.3) e encontra-se o valor exato de b que é a solução do PGD.

Sendo assim, substituindo x_3 em (6.3) tem-se que:

$$\|x_3 - x_1\|^2 = b^2 = 2 \Rightarrow b = \pm\sqrt{2}$$

e assim o conjunto dos pontos

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ e } x_3 = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

é uma solução do PGD para \mathbb{R}^2 .

Note que se fosse feita a escolha de $b = -\sqrt{2}$, os pontos $x_1 = (0, 0)^T$, $x_2 = (1, 0)^T$ e $x_3 = (0, -\sqrt{2})^T$ também seriam solução para o PGD em \mathbb{R}^2 .

Dessa forma, concluí-se que a matriz dada no exemplo é uma matriz de distâncias Euclidianas, com realização em \mathbb{R}^2 . Note que a matriz M continua admitindo realizações para dimensões maiores que 2, basta acrescentar coordenadas nulas em cada um dos pontos x_1, x_2 e x_3 . No entanto, nada garante que essa matriz irá admitir realizações em dimensões menores. Para comprovar esse fato, suponha então que o problema tenha solução para \mathbb{R} . Fixe as posições de $x_1 = 0$ e $x_2 = 1$ da mesma forma que anteriormente. Tem-se então a seguinte ilustração:

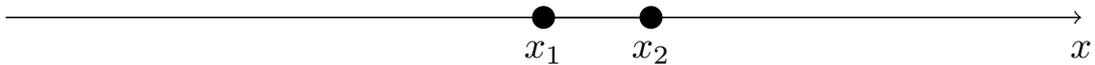


Figura 15: Pontos x_1 e x_2 posicionados

Precisa-se agora fixar o ponto x_3 de forma que as equações (6.3) e (6.4) sejam satisfeitas. Imagine inicialmente que x_3 satisfaça (6.3). Sendo assim, as seguintes posições para x_3 são:

$$|x_3| = \sqrt{2} \Rightarrow x_3 = \pm\sqrt{2}$$

Supondo agora que x_3 satisfaça (6.4), as seguintes posições para x_3 são, também:

$$|1 - x_3| = 3 \Rightarrow x_3 = -2 \text{ ou } x_3 = 4.$$

Perceba então que x_3 não satisfaz as equações (6.3) e (6.4) simultaneamente e, por isso, o PGD não possui solução em \mathbb{R} . Portanto, note que a matriz dada no exemplo é uma matriz de distâncias e também é EDM pois o PGD admite solução para $k \geq 2$.

6.1 MATRIZ DE GRAM

Definição 14. A matriz de Gram é uma matriz de produtos internos $G = X^T X$, com $X \in \mathbb{R}^{k \times n}$, ou seja, X é uma matriz formada por n vetores coluna $x_i \in \mathbb{R}^k$, ou seja,

$$G = \begin{bmatrix} \langle x_1, x_1 \rangle & \dots & \langle x_1, x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \dots & \langle x_n, x_n \rangle \end{bmatrix}_{n \times n}.$$

O interesse agora é saber qual a relação entre D , uma matriz de distâncias euclidianas com a matriz de produtos internos. Busca-se então escrever D em função de G e vice-versa.

Nas proposições e teoremas a seguir, a notação $\text{diag}(G)$ é usada para denotar um vetor coluna, com os elementos da diagonal da matriz G e e denota um vetor coluna de uns, de dimensão apropriada, i.e., $e^T = (1, 1, \dots, 1, 1)$.

Proposição 2. Sejam D uma matriz de distâncias Euclidianas e G a matriz de produtos internos associada. Então

$$D = \text{diag}(G)e^T + e\text{diag}(G)^T - 2G.$$

Demonstração. Se D é tal que $D_{ij} = \|x_i - x_j\|^2$, então

$$\begin{aligned} D &= \begin{bmatrix} \|x_1\|^2 - 2\langle x_1, x_1 \rangle + \|x_1\|^2 & \dots & \|x_1\|^2 - 2\langle x_1, x_n \rangle + \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 - 2\langle x_n, x_1 \rangle + \|x_1\|^2 & \dots & \|x_n\|^2 - 2\langle x_n, x_n \rangle + \|x_n\|^2 \end{bmatrix} \\ &= \begin{bmatrix} \|x_1\|^2 + \|x_1\|^2 & \dots & \|x_1\|^2 + \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 + \|x_1\|^2 & \dots & \|x_n\|^2 + \|x_n\|^2 \end{bmatrix} \\ &\quad - 2 \begin{bmatrix} \langle x_1, x_1 \rangle & \dots & \langle x_1, x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \dots & \langle x_n, x_n \rangle \end{bmatrix} \\ &= \begin{bmatrix} \|x_1\|^2 & \dots & \|x_1\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 & \dots & \|x_n\|^2 \end{bmatrix} + \begin{bmatrix} \|x_1\|^2 & \dots & \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_1\|^2 & \dots & \|x_n\|^2 \end{bmatrix} - 2G, \end{aligned}$$

e conclui-se que

$$D = \text{diag}(G)e^T + e\text{diag}(G)^T - 2G.$$

□

6.2 O TEOREMA DE SCHOENBERG

Agora, busca-se escrever G em função de D . Para isso, prova-se os seguintes lemas.

Lema 1. *Sejam $G = X^T X$, em que $Xe = 0$ e $J = (I - \frac{1}{n}ee^T)$. Então $G = JGJ$.*

Demonstração.

$$\begin{aligned} JGJ &= \left(I - \frac{1}{n}ee^T\right) G \left(I - \frac{1}{n}ee^T\right) \\ &= \left(I - \frac{1}{n}ee^T\right) X^T X \left(I - \frac{1}{n}ee^T\right) \\ &= \left(X^T X - \frac{1}{n}ee^T X^T X\right) \left(I - \frac{1}{n}ee^T\right) \\ &= (X^T X) \left(I - \frac{1}{n}ee^T\right) \\ &= X^T X - \frac{1}{n}X^T(Xe)e^T \\ &= X^T X. \end{aligned}$$

Portanto, $JGJ = G$.

□

Lema 2. *Dada*

$$A = \begin{bmatrix} \|x_1\|^2 & \dots & \|x_1\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 & \dots & \|x_n\|^2 \end{bmatrix},$$

tem-se que $J(A + A^T)J = 0$.

Demonstração.

$$\begin{aligned}
J(A + A^T)J &= \left(I - \frac{1}{n}ee^T\right)(A + A^T)\left(I - \frac{1}{n}ee^T\right) \\
&= \left(A + A^T - \frac{1}{n}ee^T A - \frac{1}{n}ee^T A^T\right)\left(I - \frac{1}{n}ee^T\right) \\
&= A + A^T - \frac{1}{n}ee^T A - \frac{1}{n}ee^T A^T - \frac{1}{n}Aee^T \\
&\quad - \frac{1}{n}A^T ee^T + \frac{1}{n^2}ee^T Aee^T + \frac{1}{n^2}ee^T A^T ee^T.
\end{aligned}$$

Agora, note que:

$$ee^T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ & \ddots & \\ 1 & \dots & 1 \end{bmatrix}_{n \times n}$$

e defina

$$M = \begin{bmatrix} \|x_1\|^2 + \dots + \|x_n\|^2 & \dots & \|x_1\|^2 + \dots + \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_1\|^2 + \dots + \|x_n\|^2 & \dots & \|x_1\|^2 + \dots + \|x_n\|^2 \end{bmatrix}.$$

Com isso tem-se que:

$$\begin{aligned}
J(A + A^T)J &= A + A^T - \frac{1}{n}ee^T A - \frac{1}{n}ee^T A^T - \frac{1}{n}Aee^T - \frac{1}{n}A^T ee^T \\
&\quad + \frac{1}{n^2}ee^T Aee^T + \frac{1}{n^2}ee^T A^T ee^T \\
&= A + A^T - \frac{1}{n}M - \frac{1}{n}nA^T - \frac{1}{n}nA - \frac{1}{n}M \\
&\quad + \frac{1}{n^2}nM + \frac{1}{n^2}nM \\
&= A + A^T - \frac{1}{n}M - A^T - A - \frac{1}{n}M + \frac{1}{n}M + \frac{1}{n}M = 0.
\end{aligned}$$

□

Proposição 3. *Seja D EDM, G matriz de produtos internos e seja J tal que $J = (I - \frac{1}{n}ee^T)$. Então $G = -\frac{1}{2}JDJ$.*

Demonstração.

$$\begin{aligned}
G &= \begin{bmatrix} \langle x_1, x_1 \rangle & \dots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \dots & \langle x_2, x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \dots & \langle x_n, x_n \rangle \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & \dots & x_1^T x_n \\ \vdots & \ddots & \vdots \\ x_n^T x_1 & \dots & x_n^T x_n \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} \|x_1\|^2 + \|x_1\|^2 & \dots & \|x_1\|^2 - \|x_1 - x_n\|^2 + \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 - \|x_n - x_1\|^2 + \|x_1\|^2 & \dots & \|x_n\|^2 + \|x_n\|^2 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} \|x_1\|^2 & \dots & \|x_1\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n\|^2 & \dots & \|x_n\|^2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \|x_1\|^2 & \dots & \|x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_1\|^2 & \dots & \|x_n\|^2 \end{bmatrix} \\
&\quad - \frac{1}{2} \begin{bmatrix} 0 & \dots & \|x_1 - x_n\|^2 \\ \vdots & \ddots & \vdots \\ \|x_n - x_1\|^2 & \dots & 0 \end{bmatrix} = \frac{1}{2} (A + A^T - D).
\end{aligned}$$

Assim, $G = \frac{1}{2} (A + A^T - D)$.

Do *Lema 1* e do *Lema 2* segue que:

$$\begin{aligned}
G &= JGJ = \frac{1}{2} J (A + A^T - D) J \Rightarrow \\
G &= \frac{1}{2} J (A + A^T) J - \frac{1}{2} JDJ \Rightarrow G = -\frac{1}{2} JDJ.
\end{aligned}$$

□

Apresenta-se agora a relação entre matrizes de distâncias euclidianas e matrizes positivas semidefinidas (DOKMANIC et al., 2015). Este resultado é conhecido como Teorema de Schoenberg (SCHOENBERG, 1935).

Teorema 3. *Dada $D \in \mathbb{R}^{n \times n}$, D matriz de distâncias, D é EDM se, e somente se, $-\frac{1}{2}JDJ$ é positiva semidefinida. Além disso, a dimensão de realização de D corresponde ao posto de $-\frac{1}{2}JDJ$.*

Demonstração. Suponha inicialmente que D é EDM e o objetivo é mos-

trar que $-\frac{1}{2}JDJ$ é positiva semidefinida. Como D é EDM, existe uma matriz $X \in \mathbb{R}^{k \times n}$ com $Xe = 0$ tal que $D_{ij} = \|x_i - x_j\|^2, \forall \{i, j\}$, de onde segue das Proposições e Lemas acima a relação $X^T X = G = -\frac{1}{2}JDJ$. Assim, $v^T (-\frac{1}{2}JDJ) v = v^T G v = v^T (X^T X) v = (Xv)^T (Xv) = \|Xv\|_2^2 \geq 0$.

Por outro lado, seja D uma matriz de distâncias e $\tilde{G} = -\frac{1}{2}JDJ$ positiva semidefinida. Como D é matriz de distâncias, tem-se que D é matriz simétrica e, assim, $-\frac{1}{2}JDJ$ também é simétrica. Pelo teorema espectral (veja Capítulo 4) existe Q matriz ortogonal tal que $\tilde{G} = Q\Lambda Q^T$, em que Λ é matriz diagonal com entradas que são os autovalores de \tilde{G} . Dessa forma, segue que:

$$\tilde{G} = Q\Lambda Q^T = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix}$$

Pode-se considerar a matriz \tilde{G} em blocos da seguinte forma:

$$\begin{aligned} \tilde{G} &= \begin{bmatrix} Q_k & \vdots & \tilde{Q}_k \end{bmatrix} \begin{bmatrix} \Lambda_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{bmatrix} Q_k^T \\ \vdots \\ \tilde{Q}_k^T \end{bmatrix} \\ &= Q_k \Lambda_k Q_k^T = \left(Q_k \sqrt{\Lambda_k} \right) \left(\sqrt{\Lambda_k} Q_k^T \right), \end{aligned}$$

em que Λ_k é matriz diagonal com os autovalores positivos de \tilde{G} na diagonal, Q_k é matriz ortogonal com entradas sendo os k autovetores associados aos k autovalores de Λ_k e \tilde{Q}_k é matriz ortogonal com os demais autovetores associados aos autovalores nulos.

Assim, as colunas de $X = \sqrt{\Lambda_k} Q_k^T$ formam uma realização para D em \mathbb{R}^k e portanto D é EDM com matriz de Gram associada $\tilde{G} = -\frac{1}{2}JDJ$ e concluí-se que \tilde{G} é uma matriz de Gram.

Por fim, note pelo feito acima que o número de linhas não nulas de \tilde{G} é determinado pelo bloco Λ_k , sendo assim, o número de linhas desse bloco determina o posto de \tilde{G} e o mesmo coincide com a menor dimensão de realização de D (dimensão k). \square

A partir de agora, para saber se uma certa matriz D é EDM, basta verificar se $-\frac{1}{2}JDJ$ é positiva semidefinida.

A expressão $(Q_k\sqrt{\Lambda_k})(\sqrt{\Lambda_k}Q_k^T)$ também nos permite obter as coordenadas $X \in \mathbb{R}^{k \times n}$ em dimensão k . No entanto, as coordenadas obtidas por $X = \sqrt{\Lambda_k}Q_k^T$ estarão, em geral, em um sistema de coordenadas (referencial) diferente do desejado. Nesse caso, transformações rígidas são necessárias para transformar/levar tais vetores coordenadas para o referencial ideal.

7 SOBREPOSIÇÃO DE CONJUNTOS DE PONTOS

No algoritmo que busca-se desenvolver nesse trabalho será preciso sobrepor dois conjuntos de pontos. Para realizar a sobreposição, primeiro é necessário centralizá-los em torno da origem, no caso em que os conjuntos não estejam centralizados e, em seguida, determinar uma rotação que faça com que os dois conjuntos de pontos centralizados se sobreponham da melhor forma possível.

7.1 CENTRALIZAÇÃO DE PONTOS AO REDOR DA ORIGEM

Para realizar a centralização ao redor da origem utiliza-se o conceito de centro de massa.

Considere o seguinte conjunto de pontos $x_1, \dots, x_n \in \mathbb{R}^k$. Quer-se centralizar esse conjunto de pontos em torno da origem. Para essa centralização ocorrer basta que todos os pontos do conjunto sofram uma translação conveniente.

Começa-se encontrando o centro de massa desse conjunto de pontos. Como $x_i \in \mathbb{R}^k$, denotam-se as coordenadas de um vetor x_i por $x_i = (x_i^1, x_i^2, \dots, x_i^k)$.

As coordenadas do vetor centro de massa $c = (x_{1c}, x_{2c}, \dots, x_{kc})$ são dadas por:

$$\begin{aligned} x_{1c} &= \frac{x_1^1 + x_2^1 + \dots + x_n^1}{n} \\ x_{2c} &= \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} \\ &\vdots \\ x_{kc} &= \frac{x_1^k + x_2^k + \dots + x_n^k}{n} \end{aligned}$$

Dessa forma, para transladar todo o conjunto de pontos basta fazer com que o vetor centro de massa translade até a origem para coincidir com o vetor nulo. Então para transladar os pontos até a origem,

basta fazer $x_1 - c, x_2 - c, \dots, x_n - c$ e tem-se como resposta o conjunto x_1^*, \dots, x_n^* inicialmente desejado.

Observação: Note que pode-se centralizar um conjunto de pontos em torno de qualquer outro ponto de escolha, basta subtrair um vetor conveniente para tal.

Exemplo 7. Considere o conjunto de pontos $x_1 = (3, 2)$, $x_2 = (4, 5)$, $x_3 = (5, 5)$. Centralize tal conjunto em torno da origem.

Considere a seguinte ilustração desse conjunto de pontos:

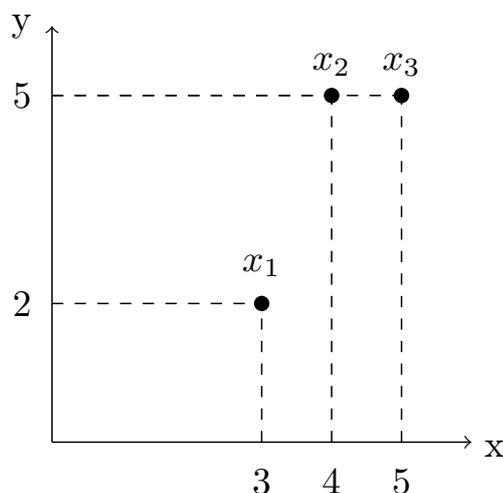


Figura 16: Posição dos pontos x_1 , x_2 e x_3

Seguindo o raciocínio anterior o primeiro passo é encontrar o vetor c . Sendo assim:

$$x_{1c} = \frac{3 + 4 + 5}{3} \Rightarrow x_{1c} = 4$$

$$x_{2c} = \frac{2 + 5 + 5}{3} \Rightarrow x_{2c} = 4$$

Agora sabe-se que o vetor centro de massa é o vetor $c = (4, 4)$. Note a posição do vetor c na imagem abaixo:

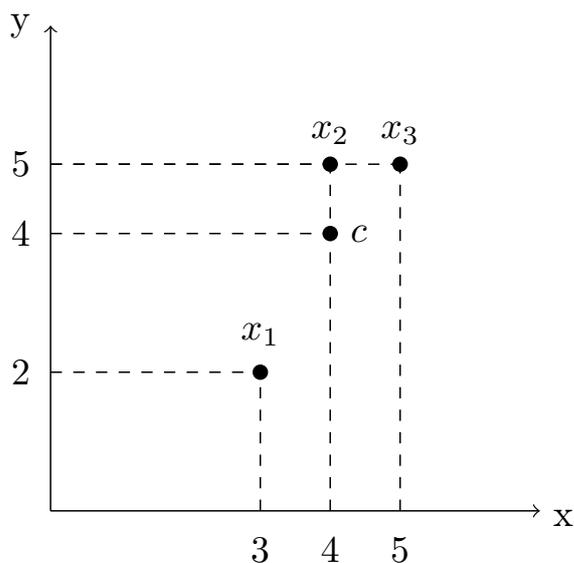


Figura 17: Coordenadas do vetor centro de massa

Com isso, centraliza-se os pontos ao redor da origem da seguinte forma:

$$x_1^* = x_1 - c = (3, 2) - (4, 4) \Rightarrow x_1^* = (-1, -2)$$

$$x_2^* = x_2 - c = (4, 5) - (4, 4) \Rightarrow x_2^* = (0, 1)$$

$$x_3^* = x_3 - c = (5, 5) - (4, 4) \Rightarrow x_3^* = (1, 1)$$

Repare na posição dos pontos x_1^* , x_2^* e x_3^* na imagem abaixo:

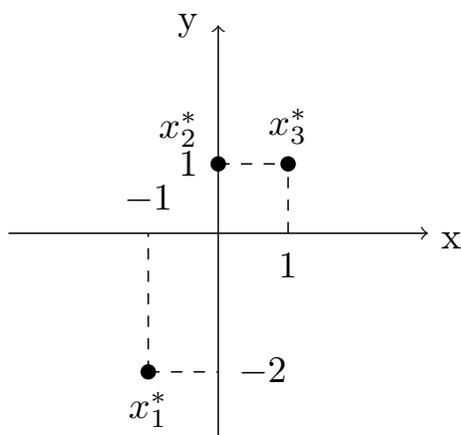


Figura 18: x_1 , x_2 e x_3 centralizados ao redor da origem

E, portanto, x_1^* , x_2^* , x_3^* estão centralizados ao redor da origem.

7.2 PROBLEMA DE PROCRUSTES ORTOGONAL

Agora, além de centralizar um conjunto de pontos ao redor da origem, precisa-se ainda tratar de um problema de aproximação de matrizes bem conhecido em álgebra linear: o *Problema de Procrustes Ortogonal* (SCHÖNEMANN, 1966). Tal problema pode ser formulado como o seguinte problema de otimização:

$$\begin{aligned} \min \quad & \|QX_c - Y_c\|_F^2 \\ \text{s.a} \quad & Q^T Q = Q Q^T = I, \end{aligned} \tag{7.1}$$

em que X_c e Y_c são matrizes cujas colunas são coordenadas de pontos já centralizados. O objetivo é encontrar novas coordenadas para os pontos do conjunto X_c para que tais coordenadas se encaixem da melhor forma possível (com o menor erro possível) no conjunto Y_c . Para ocorrer tal encaixe é necessário encontrar uma matriz ortogonal Q que transformará os pontos de X_c para coincidirem com Y_c .

O problema (7.1) equivale a encontrar a matriz ortogonal Q mais próxima da matriz $M = Y_c X_c^T$ (para depois aplicá-la à matriz X_c para que ocorra a rotação e o encaixe dos pontos). Para encontrar tal matriz Q utiliza-se a SVD (veja o Capítulo 4). Dada a SVD de M , $M = U \Sigma V^T$, segue que a matriz Q procurada é dada por $Q = UV^T$ (GOLUB; LOAN, 1996).

8 ALGORITMO DE RECONSTRUÇÃO GEOMÉTRICA SEQUENCIAL

O algoritmo desenvolvido pertence à classe de algoritmos de reconstrução geométrica sequencial (LIBERTI et al., 2012), (SIT; WU; YUAN, 2009), (DONG; WU, 2002), (DONG; WU, 2003) e utiliza-se de (sub)matrizes de distâncias euclidianas, do Teorema de Schoenberg e do problema de Procrustes ortogonal para encontrar a posição de um conjunto de n pontos em certa dimensão k de interesse, satisfazendo uma lista de distâncias incompleta, baseado em (GONÇALVES, to appear, 2017). Para tanto, algumas hipóteses devem ser satisfeitas.

8.1 HIPÓTESES

Dado um PGD representado através de um grafo $G(V, E, d)$, existe uma ordenação dos vértices (v_1, \dots, v_n) tal que:

1. todas as distâncias entre os $k+2$ primeiros vértices são conhecidas;
2. para os vértices $k+3, \dots, n$, as distâncias entre tais vértices e seus $k+1$ predecessores imediatos são conhecidas.

Por exemplo, considerando-se um PGD com 8 vértices e dimensão de realização igual a 2, espera-se que a matriz D apresente-se da seguinte forma:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & d_{14} & * & * & * & * \\ d_{21} & 0 & d_{23} & d_{24} & d_{25} & * & * & * \\ d_{31} & d_{32} & 0 & d_{34} & d_{35} & d_{36} & * & * \\ d_{41} & d_{42} & d_{43} & 0 & d_{45} & d_{46} & d_{47} & * \\ * & d_{52} & d_{53} & d_{54} & 0 & d_{56} & d_{57} & d_{58} \\ * & * & d_{63} & d_{64} & d_{65} & 0 & d_{67} & d_{68} \\ * & * & * & d_{74} & d_{75} & d_{76} & 0 & d_{78} \\ * & * & * & * & d_{85} & d_{86} & d_{87} & 0 \end{bmatrix}, \quad (8.1)$$

com as entradas em $*$ sendo as distâncias desconhecidas (e que não influenciarão no algoritmo).

8.2 DESCRIÇÃO DO ALGORITMO

Entrada: D EDM possivelmente incompleta e $k > 0$ dimensão de interesse

Passo 1: Posicionar os $k + 2$ primeiros vértices utilizando D_0 submatriz inicial de dimensão $(k + 2) \times (k + 2)$;

Passo 2: Para os vértices $i = k + 3, \dots, n$

1. Utilizar a submatriz D_i de dimensão $(k + 2) \times (k + 2)$ correspondente aos vértices $\{v_{i-k-1}, \dots, v_{i-1}, v_i\}$ para:
 2. calcular as coordenadas de X_i ;
 3. centralizar e transladar X_i até a origem do sistema e obter um novo conjunto \hat{X}_i de pontos;
 4. rotacionar \hat{X}_i para encaixar com o conjunto \tilde{X}_{i-1} (matriz que contém os $k + 1$ vértices anteriormente posicionados);
 5. rotacionar e transladar x_i para o referencial dos primeiros $k + 2$ vértices.

Saída: Matriz X_t cujas colunas são as coordenadas dos n pontos na dimensão k

Para o posicionamento dos $k + 2$ primeiros pontos trabalha-se com a primeira submatriz D_0 que contém as $k + 2$ primeiras linhas e colunas de D . Para tal posicionamento utiliza-se o Teorema de Schoenberg (vide capítulo 6, seção 6.2).

Para os próximos pontos, calculam-se as coordenadas de X_i , dos vértices $\{v_{i-k-1}, \dots, v_{i-1}, v_i\}$ pelo Teorema de Schoenberg, usando a submatriz D_i . As coordenadas obtidas estão em um sistema de coordenadas arbitrário, é necessário então aplicar uma rotação e translação convenientes para levar as coordenadas do novo ponto x_i para o referencial dos primeiros $k + 2$ vértices. Na parte da centralização e translação do conjunto X_i utiliza-se o centro de massa, que foi visto no capítulo 7, na seção 7.1. Por fim, na parte de rotacionar \hat{X}_i para se sobrepor ao conjunto \tilde{X}_{i-1} utiliza-se o Problema de Procrustes Ortogonal visto no capítulo 7, seção 7.2.

8.3 EXPERIMENTOS NUMÉRICOS

Os experimentos computacionais foram realizados no ambiente Matlab/Octave. As matrizes de distâncias dos experimentos realizados foram geradas da seguinte maneira: primeiro foi construído um conjunto de n pontos aleatórios em dimensão k no hipercubo unitário e esse conjunto de pontos foi armazenado como entradas de uma matriz $X \in \mathbb{R}^{k \times n}$. Após isso, as entradas D_{ij} de D tais que $|i - j| \leq k + 1$ recebem como valor os quadrados das distâncias entre os respectivos pontos i e j . Gera-se então um conjunto de pontos original X e uma EDM incompleta inicial D .

Para avaliar a precisão do algoritmo, duas medidas de erro foram consideradas, as chamadas LDE (em inglês, Largest Distance Error) e RMSD (em inglês, Root Mean Square Deviation). O erro LDE analisa a maior violação dentre todas as distâncias disponíveis e o erro RMSD é uma média dos erros de posicionamento ponto-a-ponto entre dois conjuntos de pontos (GONÇALVES, to appear, 2017).

O LDE é definido da seguinte maneira:

$$LDE(X, D, E) = \max_{\{i,j\} \in E} \{ | \|x_i - x_j\| - d_{ij} | \}.$$

O RMSD é dado pela expressão:

$$RMSD(X_c, Y_c) = \frac{1}{\sqrt{n}} \min_{Q^T Q = I} \|Q X_c - Y_c\|_F.$$

O cálculo do erro RMSD é feito através do problema de Procrustes Ortogonal, sendo X_c o conjunto dos pontos gerados pelo algoritmo de reconstrução geométrica sequencial X_t já centralizado e Y_c o conjunto de pontos original X após centralização.

Para cada dimensão k e número de vértices n , foram geradas 10 instâncias e os erros foram calculados para cada uma delas.

As Tabelas 1 e 2 apresentam os valores mínimo, máximo e a média dos erros LDE e RMSD para cada par (k, n) .

k	n	Mínimo LDE	Máximo LDE	Média LDE
2	10	3.3307e-016	1.6653e-015	9.9364e-016
	100	2.9976e-015	1.8874e-014	8.3544e-015
	500	8.4377e-015	2.8422e-014	1.7042e-014
	1000	1.4100e-014	7.9936e-014	2.8921e-014
3	10	7.7716e-016	2.9976e-015	1.8929e-015
	100	7.5495e-015	2.4758e-014	1.5832e-014
	500	3.0642e-014	1.4833e-013	6.0907e-014
	1000	5.7288e-014	3.2085e-013	1.3194e-013

Tabela 1: Erro LDE

k	n	Mínimo RMSD	Máximo RMSD	Média RMSD
2	10	1.9672e-016	4.4359e-016	2.7900e-016
	100	5.7453e-016	2.5030e-015	1.3045e-015
	500	1.2124e-015	4.5308e-015	2.5658e-015
	1000	1.9597e-015	7.5044e-015	3.8232e-015
3	10	3.1881e-016	2.0143e-015	7.7723e-016
	100	1.6918e-015	5.5827e-015	3.0931e-015
	500	5.5852e-015	1.8449e-014	8.6203e-015
	1000	8.8947e-015	4.7863e-014	1.6945e-014

Tabela 2: Erro RMSD

É possível perceber pela ordem de grandeza dos valores dos erros a precisão do algoritmo proposto. Observe que a maior violação entre as distâncias foi da ordem de 10^{-13} se considerado o erro LDE e da ordem de 10^{-14} se considerado o erro RMSD. Ressalta-se ainda que cada uma das instâncias foi resolvida em no máximo 30 segundos, em um laptop com processador de 2,20 GHz e 8 GB de memória RAM.

As Figuras 19 e 20 ilustram uma instância com 100 pontos e uma instância com 1000 pontos em \mathbb{R}^2 respectivamente, mostrando o conjunto de pontos original X e os pontos gerados pelo algoritmo X_t plotados pelo software Matlab. Os círculos representam os pontos originais e as cruces representam os pontos reconstruídos.

Analisando as tabelas e as figuras, concluí-se que o algoritmo construído é bastante preciso e é possível utilizá-lo para encontrar a solução de PGDs no caso em que nem todas as distâncias são conhecidas.

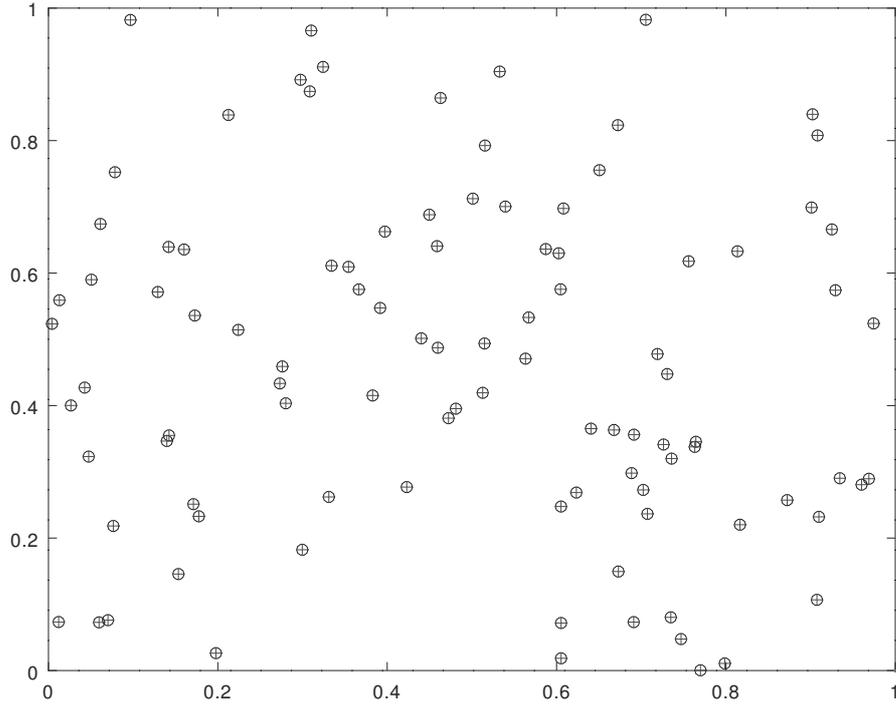


Figura 19: Posicionamento de 100 pontos em \mathbb{R}^2

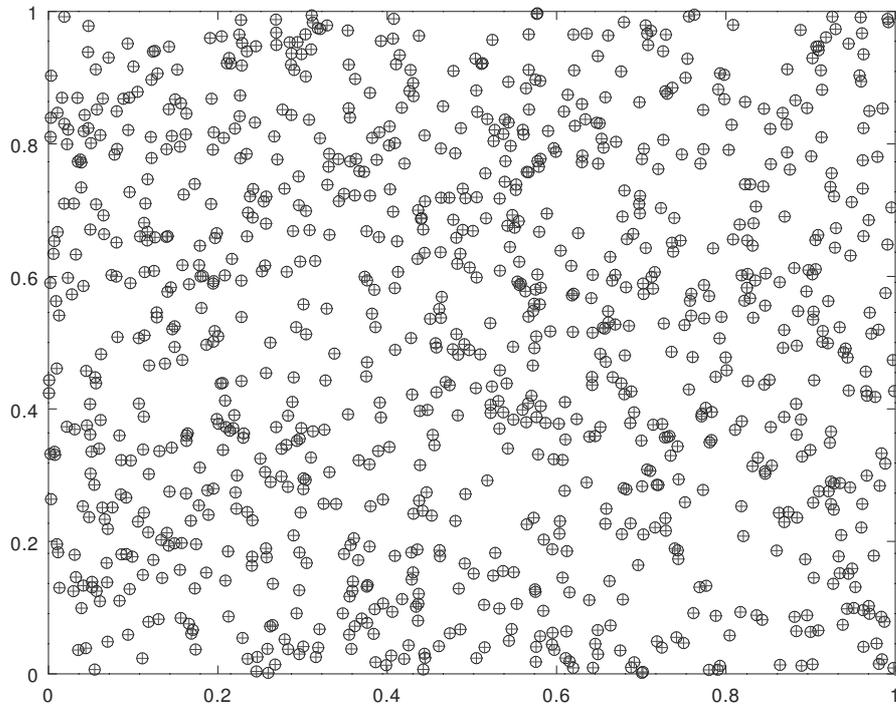


Figura 20: Posicionamento de 1000 pontos em \mathbb{R}^2

9 ORDENAÇÃO DO CONJUNTO DE VÉRTICES

Até o capítulo anterior o interesse era descobrir as coordenadas de conjuntos de vértices que satisfaziam as hipóteses do algoritmo de reconstrução geométrica sequencial. No entanto, tais hipóteses podem ser relativamente fortes na prática, isto é, a ordem natural dada aos vértices pode não satisfazer tais hipóteses.

Então surge um novo problema: será que dado um conjunto de vértices, é possível encontrar uma ordenação de tais vértices de forma que as hipóteses sejam satisfeitas?

Em verdade serão trabalhadas hipóteses um pouco mais fracas que as anteriores.

9.1 HIPÓTESES

Dado um PGD representado através de um grafo $G(V, E, d)$, existe uma ordenação(enumeração) dos vértices (v_1, \dots, v_n) tal que:

1. os primeiros $k + 1$ vértices formam um clique;
2. para cada vértice v_i , com $i > k + 1$, existem pelo menos $k + 1$ predecessores adjacentes a v_i .

Na hipótese acima, v_j é um predecessor adjacente a v_i se $\{v_j, v_i\} \in E$ e $j < i$ na ordem definida por (v_1, \dots, v_n) .

Considerando tais hipóteses, caso exista uma ordenação válida, é possível aplicar o algoritmo sequencial com poucos ajustes: a saber, as submatrizes D_i são construídas a partir das distâncias entre v_i a seus predecessores adjacentes (não necessariamente imediatos) e eventuais distâncias faltantes em tais submatrizes podem ser completadas (calculadas) com as posições já conhecidas dos predecessores.

9.2 DESCRIÇÃO DO ALGORITMO DE ORDENAÇÃO

Entrada: D EDM incompleta, $k > 0$ dimensão de interesse e n número de vértices do problema

Passo 1: Para cada $(k + 1)$ -clique C induzido pela matriz D , aplicar o Passo 2.

Passo 2: Usando os vértices em C como iniciais, tentar completar uma ordenação que satisfaça as hipóteses acima (MUCHERINO et al., 2012).

Detalha-se abaixo o algoritmo (sub-rotina) invocado no Passo 2 que tenta completar uma ordenação válida a partir de um dado clique inicial C . Neste algoritmo, $\delta(v)$ denota os vértices adjacentes a v .

Entrada: (V, E) e um clique inicial C

Passo 2a: Faça $B = C$ e $C_b = V \setminus B$.

Passo 2b: Para todo $v \in C_b$, faça $\alpha(v) = |\delta(v) \cap B|$.

Passo 2c: Se $C_b = \emptyset$, pare.

Passo 2d: Seja $v = \operatorname{argmax}\{\alpha(u) : u \in C_b\}$.

Passo 2e: Se $\alpha(v) < k + 1$, pare(falha).

Passo 2f: Para cada $u \in \delta(v) \setminus B$, atualize $\alpha(u) = \alpha(u) + 1$.

Passo 2g: Atualize $B = B \cup \{v\}$, $C_b = C_b \setminus \{v\}$ e volte ao Passo 2c.

Saída: Se não parou por falha, B representa a ordenação procurada.

Para realizar o passo 1, para cada uma das possíveis combinações de $k + 1$ dentre n vértices, verifica-se se todas as distâncias na combinação são conhecidas. Se sim, passa-se ao Passo 2; senão a próxima combinação é checada. Se nenhuma das combinações for um clique, o algoritmo termina com “falha”.

Para realizar o Passo 2 segue-se a seguinte resolução, baseada em (MUCHERINO et al., 2012): dado um $k + 1$ clique C , os vértices em C serão as entradas de um conjunto (lista) chamado de B . Cria-se um conjunto $C_b = V \setminus B$. Para cada $v \in C_b$, calcula-se o número de vértices adjacentes a v que já estão em B , e tais quantidades são armazenadas como entradas de um vetor α . Para completar a ordenação (caso possível), enquanto $C_b \neq \emptyset$, adiciona-se ao conjunto B o vértice v em C_b que possui o maior número de vértices adjacentes em B (analisando os valores correspondentes no vetor α – empates são quebrados arbitrariamente). Assim, $B = B \cup \{v\}$ e $C_b = C_b \setminus \{v\}$ (em verdade v ocupa a nova posição na ordem definida em B). Após o vértice em C_b ter

seja incorporado a lista B , os valores das entradas de α são atualizados (soma-se 1 a cada valor $\alpha(u)$ caso u seja adjacente a v , o vértice recém incluído em B). O processo se repete para os próximos vértices em C_b , até que $C_b = \emptyset$. Caso exista ordenação válida a partir de C , esta é revelada ao final do processo na lista B . Se durante o processo, todos os valores de $\alpha(v)$, para $v \notin B$, forem menores que $k + 1$, o algoritmo para pois não há ordenação a partir do clique inicial C , e então outro clique é testado.

A corretude do algoritmo de ordenação descrito acima foi provada em (MUCHERINO et al., 2012).

9.3 EXPERIMENTOS NUMÉRICOS

As matrizes de distâncias dos experimentos realizados nesse caso foram geradas da seguinte maneira: foi gerado um conjunto de n pontos aleatórios em dimensão k no hipercubo unitário e computadas as distâncias entre todos os pares de pontos. A distância D_{ij} é então incluída na matriz D somente se $\|X_i - X_j\| \leq r$, onde r é uma tolerância definida a priori.

Com este procedimento, agora D não mais possui a estrutura de (8.1) e, em geral, faz-se necessária a re-ordenação dos vértices a fim de cumprir as hipóteses deste capítulo.

Sendo assim, após a matriz D ser gerada, aplica-se então o algoritmo da ordenação para saber se é possível ou não ordenar os vértices. Nos casos em que foi possível encontrar a ordenação, sendo O o vetor que representa a ordenação dos vértices, a matriz D será reorganizada segundo essa nova ordenação através do comando (Matlab): $D_t = D(O, O)$. Com isso, através das alterações feitas no código do algoritmo de reconstrução geométrica sequencial, é possível aplicá-lo em D_t e descobre-se assim as coordenadas dos vértices.

Novamente, o erro $RMSD$ foi utilizado para verificar a precisão desse novo algoritmo.

Os experimentos desse algoritmo foram feitos com a dimensão $k = 2$ fixa e os valores de n e r foram variados. A Tabela 3 mostra o tempo para a ordenação (coluna Tempo 1), o tempo do algoritmo de reconstrução (coluna Tempo 2), o erro calculado e a porcentagem (%) de distâncias conhecidas para cada par (n, r) .

n	r	Tempo 1	Tempo 2	Erro RMSD	%
10	0.1	-	-	-	0.15
	0.5	8.31	5.86	2.48e-016	0.82
	0.7	4.06	3.49	3.22e-016	0.91
100	0.01	-	-	-	0.02
	0.1	6.64	6.05	1.59e-015	0.22
	0.5	5.10	3.89	8.80e-016	0.73
	0.7	2.85	2.85	2.86e-015	0.87
500	0.01	-	-	-	0.02
	0.1	14.37	6.15	1.98e-014	0.22
	0.5	8.61	4.47	6.60e-015	0.76
	0.7	8.23	3.36	3.49e-015	0.88
1000	0.01	28.38	6.37	5.48e-014	0.02
	0.1	23.21	3.38	7.12e-014	0.24
	0.5	36.42	3.71	4.58e-015	0.77
	0.7	24.75	3.02	3.27e-014	0.87

Tabela 3: Resultados para as novas instâncias

Na tabela os “-” indicam que não foi possível encontrar a ordenação e a porcentagem de distâncias conhecidas é calculada através da expressão $|E|/(n(n-1)/2)$. Analisando a coluna do erro, é possível perceber a precisão do algoritmo de reconstrução em um tempo satisfatório para cada par (n, r) . Portanto, o algoritmo de reconstrução geométrica sequencial, em conjunto com o algoritmo da reordenação, pode ser utilizado para resolver problemas práticos. As precisões podem ser observadas nas Figuras 21, 22 e 23.

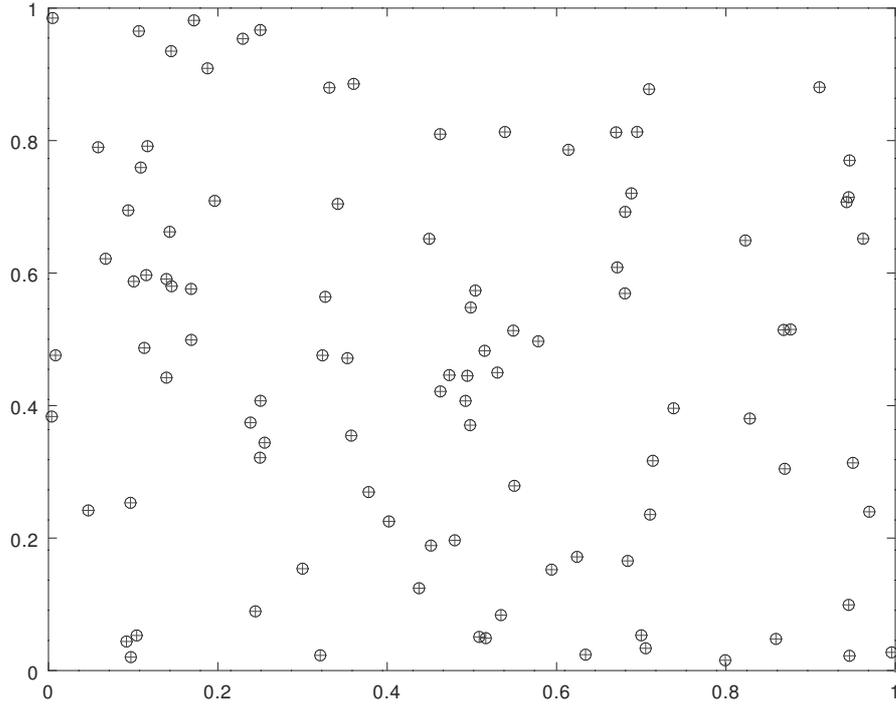


Figura 21: Posicionamento de 100 pontos com $r = 0.1$

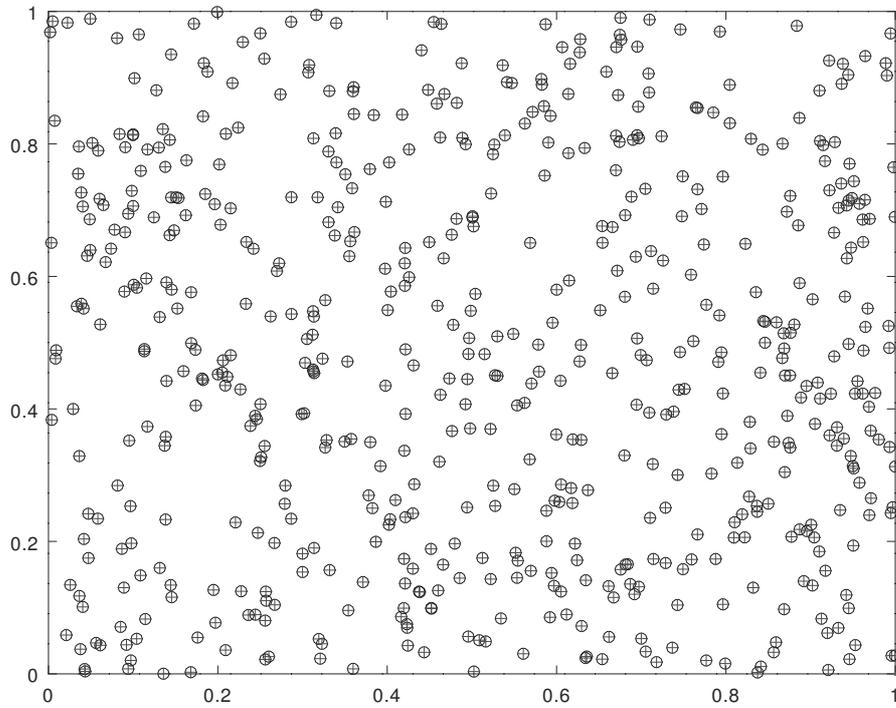


Figura 22: Posicionamento de 500 pontos com $r = 0.7$

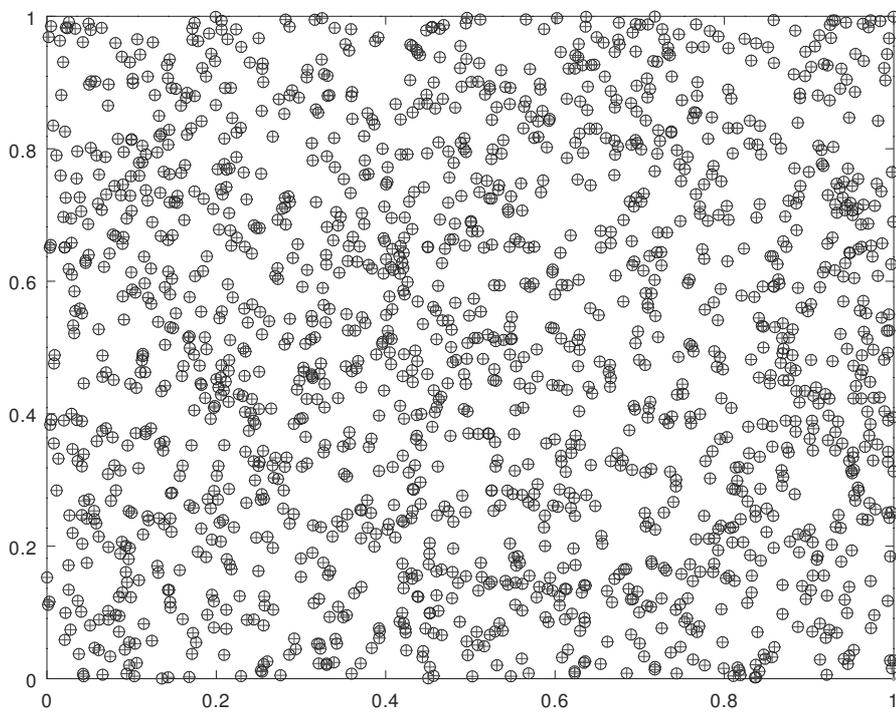


Figura 23: Posicionamento de 1000 pontos com $r = 0.01$

10 CONSIDERAÇÕES FINAIS

Nesse Trabalho de Conclusão de Curso foi estudado o Problema de Geometria de Distâncias e sua resolução através de Matrizes de Distâncias Euclidianas (EDM). Foi proposto um algoritmo de reconstrução geométrica sequencial que explora submatrizes EDM de uma matriz de distâncias incompleta. Tal algoritmo foi capaz de resolver diversos problemas gerados de forma aleatória em dimensões 2 e 3 e com milhares de pontos. Ainda, com experimentos computacionais foi possível confirmar a precisão do algoritmo nas resoluções dos problemas. É possível concluir então que, satisfeitas as hipóteses (do Capítulo 8), o algoritmo resolve o problema gerado com precisão satisfatória, cumprindo assim seu objetivo e podendo ser usado de forma eficiente na prática no caso em que nem todas as distâncias são conhecidas.

Como em problemas práticos a ordem natural dada aos vértices (objetos) pode não satisfazer as hipóteses, foi estudado o problema de obter uma ordenação de um conjunto de vértices dada uma matriz de distâncias incompleta (e ainda relaxando as hipóteses do Capítulo 8). Como verificado nos experimentos computacionais, em geral foi possível encontrar uma ordenação válida desde que a porcentagem de distâncias conhecidas na EDM seja aproximadamente superior a 1% do total $n(n-1)/2$. Nestes casos, através do algoritmo de reconstrução modificado, foi possível encontrar as coordenadas dos pontos com boa precisão, mesmo para conjuntos com milhares de pontos.

REFERÊNCIAS

DOKMANIC, I. et al. Euclidean distance matrices: essential theory, algorithms and applications. *IEEE Signal Processing Magazine*, v. 32, p. 12–30, 2015.

DONG, Q.; WU, Z. A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Global Optimization*, v. 22, p. 365–375, 2002.

DONG, Q.; WU, Z. A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data. *Global Optimization*, v. 26(3), p. 321 – 333, 2003.

GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. Baltimore, USA: The Johns Hopkins University Press, 1996. 723 p.

GONÇALVES, D. S. A least-squares approach for discretizable distance geometry problems with inexact distances. *Optimization Letters*, to appear, 2017.

LAVOR, C.; LIBERTI, L. Um convite à geometria de distâncias. *Sociedade Brasileira de Matemática Aplicada e Computacional*, v. 71, p. 69, 2014.

LIBERTI, L. et al. The discretizable molecular distance geometry problem. *Computational Optimization and Applications*, v. 52, p. 115 – 146, 2012.

LIMA, E. L. *Álgebra Linear*. Rio de Janeiro, Brasil: IMPA, 2009. 357 p.

MEYER, C. D. *Matrix Analysis and Applied Linear Algebra*. Philadelphia, USA: SIAM, 2000. 718 p.

MUCHERINO, A. et al. Discretization orders for distance geometry problems. *Optimization Letters*, v. 6, p. 783 – 796, 2012.

SCHOENBERG, I. Remarks to maurice fréchet’s article “sur la définition axiomatique d’une classe d’espaces distanciés vectoriellement applicable sur l’espace de hilbert”. *Annals of Mathematics*, v. 36, p. 724–732, 1935.

SCHÖNEMANN, P. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, v. 31, p. 1–10, 1966.

SIT, A.; WU, Z.; YUAN, Y. A geometric buildup algorithm for the solution of the distance geometry problem using least-squares approximation. *Bulletin of Mathematical Biology*, v. 71, p. 1914–1933, 2009.

STRANG, G. *Álgebra Linear e suas aplicações*. USA: Cengage Learning, 2011. 444 p.