



Estimação de distâncias em redes de sensores utilizando técnicas de completamento de matrizes de posto reduzido

Nilson Moreira

Instituto de Matemática, Estatística e Computação Científica - UNICAMP
nilsonmat27@gmail.com

Carlile Campos Lavor

Instituto de Matemática, Estatística e Computação Científica - UNICAMP
clavor@ime.unicamp.br

Cristiano Torezzan

Faculdade de Ciências Aplicadas - UNICAMP
cristiano.torezzan@fca.unicamp.br

RESUMO

Um problema clássico no contexto de redes de sensores consiste em determinar a localização dos sensores na rede com base no conhecimento das distâncias entre eles. No entanto, na prática, é comum que apenas um subconjunto do total de distâncias seja conhecido. Nesses casos é necessário obter algum tipo de aproximação ou estimativa dos dados faltantes de modo a completar a matriz de distâncias para, então, determinar a localização dos sensores. Neste trabalho propomos a aplicação de técnicas de completamento de matrizes de posto reduzido no contexto de localização em redes de sensores. Além da descrição do método, apresentamos, também, resultados computacionais que mostram que é possível recuperar matrizes de distâncias com alta percentagem de dados faltantes com erro absoluto arbitrariamente pequeno.

PALAVRAS CHAVE. Rede de sensores, completamento, matriz de distância, posto reduzido.

Tópicos (SS4 - Sessão Especial 4: Otimização, Grafos e Geometria de Distâncias, PM - Programação Matemática, OC - Otimização Combinatória)

ABSTRACT

A classic problem in sensor networks is to determine the location of the sensors based on the distances between them. However, often in practice, only a subset of the pairwise distances are known. In these cases it is useful to get some approximation or estimation for the missing data, in order to complete the distance matrix, and then to determine the location of the sensors. In this paper, we approach this problem using a low rank matrix completion technique. Besides the method, we also present computational results showing that it is possible to recover Euclidean distance matrices with high percentage of missing data with arbitrarily small absolute error.

KEYWORDS. Sensor network, matrix completion. Euclidean distances, low rank.

Paper topics (MP - Mathematical programming, CO - Combinatorial optimization)



1. Introdução

Distâncias Euclidianas aparecem em muitas aplicações do mundo real. Por exemplo, no estudo de conformação molecular, alguns pares de átomos estão conectados e a distância entre eles pode ser estimada usando técnicas apropriadas, tais como a ressonância nuclear magnética [Havel e Wüthrich, 1985]. Outro exemplo é o problema da localização em rede de sensores, em que alguns pares de sensores se comunicam diretamente por meio de sinais, permitindo que suas distâncias sejam estimadas [P. Biswas e Ye., 2006]. No entanto, devido a várias razões, em muitas situações práticas, não é possível mensurar todas as distâncias e precisamos encontrar maneiras de estimar os valores em falta. Esse problema pode ser compreendido como um problema da classe de *Completamento de Matrizes* [Candès e Recht, 2009].

De maneira geral, problemas práticos envolvem o estudo de vários pontos em espaços bi-dimensionais, ou tri-dimensionais. Se considerarmos uma matriz D , cujas entradas representam o quadrado da distância entre os pontos, pode-se mostrar (Lema 2.4) que o posto de D excede no máximo em duas unidades a dimensão do espaço que contém os pontos. Neste sentido, quando o número de pontos é bem maior que a dimensão do espaço podemos considerar que a matriz D tem posto baixo (ou reduzido). Por exemplo, se 1000 pontos forem distribuídos uniformemente em uma região do plano, a matriz D terá mil linhas e mil colunas e, no máximo, posto igual a 4.

O completamento de matrizes de posto reduzido tem sido estudado em diversos contextos e, sob certas condições, existem métodos capazes de recuperar os dados em falta com alta precisão [Candès e Recht, 2009]. Neste trabalho investigamos a aplicação dessas técnicas para estimar dados faltantes de uma matrizes de distâncias Euclidianas associadas à rede de sensores. Mais especificamente propomos uma adaptação de um método proposto em [Mazumder et al., 2010] de forma a melhor modelar o completamento de matrizes de distâncias e mostramos que, em alguns casos, é possível recuperar uma matriz de distância com alta percentagem de dados faltantes, com erro absoluto inferior a 10^{-4} . Também investigamos a influência do número de âncoras (pontos de localização conhecida) na rede de sensores com a capacidade de estimação de distâncias faltantes.

O restante do artigo está organizado como se segue. Na seção 2 apresentamos as notações e definições fundamentais para o desenvolvimento do artigo. Na seção 3 descrevemos o problema e apresentamos um algoritmo para completamento de matrizes de posto reduzido. Na seção 4 apresentamos alguns resultados numéricos para o problema de localização em redes de sensores e, finalmente, na seção 5 as conclusões e perspectivas.

2. Notações e definições preliminares

Seja $X = [x_1 \ x_2 \ \dots \ x_n]$, $x_i \in \mathbb{R}^d$ uma matriz cujas colunas representam n pontos num espaço Euclidiano d -dimensional. O quadrado da distância (d_{ij}) entre dois pontos quaisquer em \mathbb{R}^d é definido por:

$$d_{ij} = \|x_i - x_j\|_2^2 \triangleq \langle x_i - x_j, x_i - x_j \rangle = x_i^T x_i - 2x_i^T x_j + x_j^T x_j = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j. \quad (1)$$

Definição 2.1 (Matriz de distâncias Euclidiana). A matriz $D = [d_{ij}] \in \mathbb{R}^{n \times n}$, cujas entradas representam o quadrado da distância entre n pontos, colunas de X , (em \mathbb{R}^d) é denominada matriz de distâncias Euclidiana e denotada por $edm(X)$.

Definição 2.2 (Matriz de Gram). A matriz $G = X^T X \in \mathbb{R}^{n \times n}$, cujas entradas representam o produto interno entre n pontos em \mathbb{R}^d é denominada matriz de Gram associada a X .

Com base em (1) pode-se mostrar que

$$D \triangleq \mathbf{1}diag(X^T X)^T - 2X^T X + diag(X^T X)\mathbf{1}^T, \quad (2)$$



onde $\mathbf{1}$ representa o vetor coluna com todas as entradas iguais a 1, e $\text{diag}(A)$ representa o vetor coluna, composto pelos elementos da diagonal de A .

Dois resultados básicos, porém fundamentais para este trabalho são apresentados nos dois lemas que se seguem, e mostram que os postos das matrizes D e G dependem apenas da dimensão d e não do número de pontos n .

Lema 2.3. *O posto de uma matriz de Gram associada a X é no máximo d .*

Demonstração. Como X tem d linhas, temos que $\text{posto}(X) \leq d$. Pelas propriedades:

- $\text{posto}(AB) \leq \min(\text{posto}(A), \text{posto}(B))$,
- $\text{posto}(A) = \text{posto}(A^T)$,

concluimos que:

$$\text{posto}(G) = \text{posto}(X^T X) \leq \min(\text{posto}(X^T), \text{posto}(X)) = \text{posto}(X) \leq d.$$

□

Lema 2.4. *O posto de uma $\text{edm}(X)$ é no máximo $d + 2$.*

Demonstração. Pelo lema 2.3, temos que o posto de $X^T X$ é d . Por outro lado, tanto $\mathbf{1}\text{diag}(X^T X)^T$ como $\text{diag}(X^T X)\mathbf{1}^T$ são matrizes de posto 1.

Portanto, pela propriedade, $\text{posto}(A + B) \leq \text{posto}(A) + \text{posto}(B)$ e por (2), concluimos que:

$$\text{posto}(\text{edm}(X)) \leq 1 + d + 1 = d + 2.$$

□

Outro resultado fundamental no contexto deste trabalho é a decomposição SVD de uma matriz, que pode ser enunciada da seguinte forma:

Teorema 2.5 (Decomposição em Valores Singulares). *Seja $A \in \mathbb{R}^{n \times m}$ uma matriz de posto r . Então, A pode ser escrita na forma*

$$A = U\Sigma V^T, \tag{3}$$

onde $U \in \mathbb{R}^{n \times n}$ e $V \in \mathbb{R}^{m \times m}$ são matrizes ortogonais e $\Sigma \in \mathbb{R}^{n \times m}$ é dada por

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0_{r \times (m-r)} \\ 0_{(n-r) \times r} & 0_{(n-r) \times (m-r)} \end{bmatrix}, \tag{4}$$

com $\Sigma_1 = \text{diag}\{\sigma_1, \dots, \sigma_r\}$.

A decomposição (3) é chamada de *decomposição em valores singulares* de A . As colunas de U são vetores ortonormais, chamados de *vetores singulares à direita* de A e as colunas de V também são ortonormais e chamados de *vetores singulares à esquerda* de A .

No caso particular de matrizes de distâncias de pontos em \mathbb{R}^d , como $\text{posto}(D) \leq d + 2$ teremos no máximo $d + 2$ valores singulares diferentes de 0.



3. Caracterização do problema e modelo proposto

3.1. O problema de Localização em Redes de Sensores

O problema de Localização em Redes de Sensores é um problema de completamento de matriz de distância Euclidiana em pequena dimensão, onde a posição de um subconjunto de nós é especificado. Nesse problema, tem-se $n - m$ pontos desconhecidos

$$x_1, \dots, x_{n-m} \in \mathbb{R}^d,$$

chamados *sensores*, e m pontos conhecidos

$$a_1, \dots, a_m \in \mathbb{R}^d,$$

denominados *âncoras*.

Em notação matricial, consideramos

$$X := \begin{bmatrix} x_1^T \\ \vdots \\ x_{n-m}^T \end{bmatrix}^T \in \mathbb{R}^{d \times (n-m)}, \quad A := \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}^T \in \mathbb{R}^{d \times m}, \quad e \quad P := [X \ A] \in \mathbb{R}^{d \times n}.$$

Com isso, escrevemos a matriz de distâncias Euclidiana, D , associada a P , da seguinte forma:

$$D := \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix},$$

onde $D_{11} \in \mathbb{R}^{(n-m) \times (n-m)}$, $D_{12} \in \mathbb{R}^{(n-m) \times m}$, $D_{21} \in \mathbb{R}^{m \times (n-m)}$ e $D_{22} \in \mathbb{R}^{m \times m}$.

O problema de Localização em Redes de Sensores consiste em encontrar um $X \in \mathbb{R}^{d \times (n-m)}$ de modo que as distâncias conhecidas em relação a P sejam preservadas. Neste trabalho nos concentramos em encontrar aproximações para os valores desconhecidos na matriz D , dado que, uma vez obtida uma matriz completa de distâncias, há algoritmos MDS eficientes para se determinar a localização dos pontos X [Dokmanic et al., 2015].

Podemos notar que D_{22} é completamente conhecida, uma vez que ela representa a distância entre as âncoras, que são pontos conhecidos. Dependendo da amplitude de alcance de cada sensor, outras distâncias também podem ser conhecidas. Se um sensor conseguir se comunicar com um outro, através da emissão/recepção de sinais, a distância entre os mesmos é considerada conhecida.

3.2. Completamento de matrizes de posto reduzido

Considere uma matriz de distâncias Euclidianas D , associada à um conjunto de pontos $X = \{x_1, x_2, \dots, x_n\}$, onde $x_i \in \mathbb{R}^d$, com $d \ll n$, na qual algumas entradas são desconhecidas. Seja $\Omega \subset \{1, \dots, n\} \times \{1, \dots, n\}$ o conjunto dos índices das entradas conhecidas. Podemos estimar as entradas faltantes em D , resolvendo o seguinte problema de otimização [Candès e Recht, 2009]:

$$\begin{cases} \min & \text{rank}(X) \\ \text{s.a.} & \sum_{(i,j) \in \Omega} (D_{ij} - X_{ij})^2 \leq \delta \end{cases} \quad (5)$$



No entanto, como o problema (5) é não-convexo e a busca de uma solução através de métodos exatos torna-se computacionalmente difícil, diferentes versões relaxadas tem sido propostas, em geral, explorando a norma nuclear de X . Neste trabalho utilizaremos a seguinte relaxação:

$$\begin{cases} \min & \|X\|_* \\ \text{s.a} & \sum_{(i,j) \in \Omega} (D_{ij} - X_{ij})^2 \leq \delta, \end{cases} \quad (6)$$

onde $\|X\|_*$ é a *norma nuclear* de X (a soma dos valores singulares de X) e $\delta > 0$ é uma tolerância pré-estabelecida.

Alternativamente pode-se considerar uma versão equivalente do problema (6) na forma Lagrangiana [Mazumder et al., 2010]:

$$\underset{X}{\text{minimizar}} \quad \frac{1}{2} \sum_{(i,j) \in \Omega} (D_{ij} - X_{ij})^2 + \lambda \|X\|_* \quad (7)$$

Neste contexto, [Mazumder et al., 2010] propuseram uma heurística (Algoritmo 1) baseada na decomposição em valores singulares, denominada *soft-impute*, para encontrar uma solução do problema (7). O método consiste em tomar sucessivas decomposições SVD de uma matriz aproximada de forma a reduzir, iterativamente, o posto da matriz resultante.

Algorithm 1 soft-impute

- Entrada: Matriz com entradas em falta, X ($n \times n$), e a tolerância ϵ .

1. Inicializar $X^{antigo} = 0$.
2. Escolher uma sequência decrescente de escalares: $\lambda_1, \lambda_2, \dots, \lambda_K$.
3. Para $i = 1$ até K fazer:
 - (a) Calcule $X^{novo} \rightarrow S_{\lambda_i}(P_{\Omega}(X) + P_{\Omega}^{\perp}(X^{antigo}))$.
 - (b) Se $\frac{\|X^{novo} - X^{antigo}\|_F^2}{\|X^{antigo}\|_F^2} < \epsilon$, termine.
 - (c) Faça $X^{antigo} \leftarrow X^{novo}$.

- Saída: X^{novo}
-

No Algoritmo 1, $S_{\lambda}(X) = UD_{\lambda}V^T$, com $D_{\lambda} = \text{diag}[(d_1 - \lambda)_+, \dots, (d_r - \lambda)_+]$, UDV^T denota a decomposição em valores singulares de X , $D = \text{diag}[d_1, \dots, d_r]$, $t_+ = \max(t, 0)$, $r = \text{rank}(X)$, e

$$P_{\Omega}(X)(i, j) = \begin{cases} X_{i,j}, & \text{se } (i, j) \in \Omega \\ 0, & \text{se } (i, j) \notin \Omega \end{cases}$$

Uma dificuldade computacional na implementação do método soft-impute é o estabelecimento de uma sequência adequada de valores λ_i , no passo 2 do Algoritmo 1. Neste trabalho utilizamos uma variante do método soft-impute, proposta em [Moreira et al., 2017], que elimina o passo 2 do algoritmo 1 e considera $\lambda_i = \beta \sigma_{d+3}$, atualizado em cada iteração. Esta proposta, apesar de parecer pequena, tem se mostrado computacionalmente muito eficiente, como apresentado na próxima seção.



4. Resultados Numéricos

Nessa seção apresentamos alguns resultados computacionais para o problema de completamento de matrizes de distâncias associadas às redes de sensores. Os testes computacionais são divididos em 2 casos.

No primeiro, n pontos são distribuídos aleatoriamente sob uma região quadrangular $r \times r$ do plano e m pontos são escolhidos como âncoras, sendo os demais $n - m$ pontos designados sensores. Com base nesses pontos, construímos a matriz de distâncias Euclidianas associada a esses n pontos e de seguida apagamos, aleatoriamente, de acordo com uma distribuição uniforme, uma percentagem p dessas distâncias, preservando as distâncias entre as âncoras. Sob a matriz incompleta (com dados apagados) aplicamos o Algoritmo 1 e comparamos o resultado com a matriz original.

No segundo caso, executamos os mesmos passos até a construção da matriz de distâncias Euclidianas. Após essa etapa, apagamos todas as distâncias maiores do que a amplitude de alcance dos sensores (considerado constante), à exceção das distâncias entre as âncoras. A semelhança do caso anterior, aplicamos o Algoritmo 1 para estimar as distâncias em falta.

A figura 1 corresponde a uma simulação enquadrada no primeiro caso. Utilizamos 100 pontos (sensores e âncoras) no quadrado 1×1 , a partir dos quais determinamos a matriz de distâncias Euclidianas. Apagamos de maneira uniforme 73% dos dados e usamos o Algoritmo 1 para estimar os entradas em falta da EDM. Para a recuperação das entradas, usamos uma tolerância de 10^{-8} .

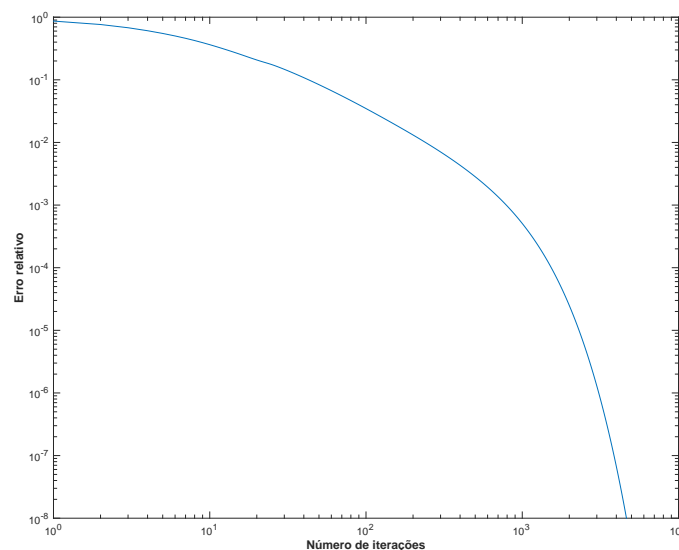


Figura 1: Comportamento do erro relativo em função do número de iterações, na recuperação das entradas de uma EDM associada a 100 pontos (sensores) em \mathbb{R}^2 . Foram apagadas 73% das entradas de uma maneira uniforme e foi usada uma tolerância de 10^{-8} .

Além da convergência do Algoritmo, a precisão absoluta foi muito boa, tendo no pior caso a recuperação de uma entrada com 7 casas decimais corretas.

A tabela 1 resume um conjunto de experimentos realizados para o segundo caso. Nota-se que a estimação dos dados não é tão boa quanto ao primeiro caso. Isto deve-se ao facto do



conjunto das posições dos dados conhecidos não seguirem uma distribuição uniforme, hipótese que é fundamental para que este método funcione bem [Candès e Recht, 2009].

# âncoras	Amplitude	% de apagamento	Erro relativo	Erro absoluto
30	0.8	15.6	1.51×10^{-1}	8.58×10^{-1}
30	0.9	6.1	9.99×10^{-5}	5.69×10^{-4}
30	1	3.36	9.93×10^{-5}	6.51×10^{-4}
50	0.8	11.72	3.16×10^{-2}	2.98×10^{-1}
50	0.9	6.16	9.97×10^{-5}	8.77×10^{-4}
50	1	1.5	9.87×10^{-5}	1.8×10^{-3}

Tabela 1: Simulação de 100 pontos em \mathbb{R}^2 , com um número máximo de iterações de 10000 e uma tolerância de 10^{-4} .

5. Conclusões

Neste trabalho estudamos o problema da localização em redes de sensores através de métodos de completamento de matrizes de posto reduzido. Apresentamos um algoritmo e diversos testes computacionais que mostram que, sob certas condições, é possível recuperar matriz de distâncias com alta percentagem de dados faltantes, com erro arbitrariamente pequeno. Além disso, como é esperado por razões teóricas, a precisão na recuperação de dados aumenta com o número de âncoras na rede e também se a posição dos dados faltantes for uniforme.

Referências

- Candès, E. J. e Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717. ISSN 1615-3383. URL <http://dx.doi.org/10.1007/s10208-009-9045-5>.
- Dokmanic, I., Parhizkar, R., Ranieri, J., e Vetterli, M. (2015). Euclidean distance matrices: Essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30. ISSN 1053-5888.
- Havel, T. F. e Wüthrich, K. (1985). An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *Journal of Molecular Biology*, 182(2):281 – 294. ISSN 0022-2836. URL <http://www.sciencedirect.com/science/article/pii/0022283685903468>.
- Mazumder, R., Hastie, T., e Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.*, 11:2287–2322. ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=1756006.1859931>.
- Moreira, N. J., Duarte, L. T., Lavor, C. C., e Torezzan, C. (2017). A novel low-rank matrix completion approach to estimate missing entries in euclidean distance matrices. *Submitted to Journal of Global Optimization*.
- P. Biswas, K. C. T. T. C. W., T. C. Liang e Ye., Y. (2006). Semidefinite programming approaches for sensor network localization with noisy distance. *IEEE Transactions on Automation Science and Engineering*, 3(4):360–371.