

**UNIVERSIDADE FEDERAL DO PARANÁ
CNPQ – PROCESSO 442132/2018-2**

Stephanie Caroline de Souza Pereira

**INTRODUÇÃO ÀS TÉCNICAS DE REGULARIZAÇÃO
EM PROBLEMAS INVERSOS DISCRETOS**

Relatório Final de Iniciação Científica associado ao projeto Meninas na Matemática: Procuram-se Arletes, desenvolvido no período de Março/2019 a Fevereiro/2020, sob a orientação da Prof^a Dra Ana Gabriela Martinez.

CURITIBA

2020

RESUMO

O presente trabalho de Iniciação Científica na UFPR foi desenvolvido dentro do projeto intitulado “Meninas na Matemática: Procuram-se Arletes”, que possui financiamento do CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (processo 442132/2018-2) e é coordenado pela Professora Elizabeth Wegner Karas. O projeto, que deseja motivar e influenciar mulheres a seguir carreiras dentro das ciências exatas, conta com a participação de três bolsistas da graduação da Universidade Federal do Paraná (UFPR) e quinze bolsistas de ICJúnior das cinco escolas públicas participantes, além de parcerias com outros projetos da UFPR.

Este trabalho tem como objetivo principal introduzir o leitor ao universo dos problemas inversos e apresentar algumas técnicas para encontrar a sua solução.

Serão apresentadas algumas das técnicas de regularização mais usadas para resolver problemas inversos discretos e lineares, entre elas a regularização de Tikhonov e a regularização denominada de decomposição em valores singulares truncada ou TSVD. Esta última técnica é de uso estendido em processamento de imagens digitais. Em particular, será aplicado o método de regularização da TSVD no problema da compressão de imagens.

Palavras-chave: Métodos de Regularização. Problemas Inversos. Decomposição em Valores Singulares.

1. INTRODUÇÃO

Os Problemas Inversos aparecem em diversas áreas do conhecimento, como, por exemplo, em imagens médicas, em geofísica e ciências ambientais, em engenharias e, de forma geral, estão relacionados com a determinação de causas desconhecidas através de efeitos observados.

Uma das principais características de um problema inverso é que, em geral, ele não satisfaz pelo menos um dos postulados do matemático Jacques Hadamard para problemas bem postos. Para Hadamard, um problema bem-posto é aquele que cumpre as condições de existência, unicidade e estabilidade. Isto significa que nos problemas inversos pode ocorrer que não exista solução, que no caso de existir esta não seja única ou que a solução não dependa continuamente dos dados do problema. A estabilidade é necessária para garantir que pequenas variações nos dados resultem em pequenas mudanças na solução, logo a perda da estabilidade implica sérias dificuldades na hora de resolver numericamente um problema inverso.

De acordo com [5], o problema de estabilidade pode ser facilmente exemplificado por uma equação algébrica do 2º grau:

$$\alpha x^2 - 2x + 1 = 0 \quad (1)$$

Para $\alpha = 1$, temos as soluções: $x_1 = x_2 = 1$. Introduzindo um erro de 1% no coeficiente α , isto é, $\alpha = 1,01$, a solução da equação (1) torna-se: $x_{1,2} = 1 \pm 0,1 * i$, ou seja, com 1% de ruído em α a equação (1) não tem mais solução dentro dos números reais.

Outro exemplo a ser tratado é problema do tipo $Ax = b$, tal que

$$A = \begin{bmatrix} 1,0001 & 1 \\ 1 & 1 \end{bmatrix} \quad (2)$$

Se tivermos o vetor b da forma $b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, teremos a solução $x = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$. Por outro lado, para $b = \begin{bmatrix} 2 \\ 2.02 \end{bmatrix}$, teremos a solução $x = \begin{bmatrix} -200 \\ 200.02 \end{bmatrix}$.

Isso nos mostra que não basta apenas resolver o problema $Ax = b$ da forma $x = A^{-1}b$, mesmo que a matriz seja invertível, pois, se ela estiver muito próxima de ser singular, a estabilidade da solução não é garantida.

Uma das estratégias mais usadas para mitigar os problemas que resultam da perda da estabilidade na solução dos Problemas Inversos é o conjunto de

técnicas matemáticas chamadas de Métodos de Regularização. A ideia central destes métodos consiste em aproximar a solução do problema inverso mal-posto por uma sequência de problemas bem-postos aproximados.

Formalmente, o problema inverso consiste em determinar os parâmetros x a partir dos dados observados b , relacionados através do modelo descrito pelo operador A , isto é: $Ax = b$. Quando o operador A é linear, o problema inverso é dito linear. A maior parte das técnicas de regularização consiste em construir uma família de operadores dependentes de um parâmetro de regularização R_λ , tal que no limite quando o parâmetro λ tende a zero, estes aproximem à inversa do operador A , quando esta inversa existir, ou, em caso contrário, à pseudo-inversa de A . A escolha do parâmetro de regularização λ é crucial neste tipo de método, já que ele estabelece um equilíbrio entre a fidelidade da solução aos dados observados e a estabilidade. Entre os métodos mais destacados na literatura para a determinação de λ encontram-se o princípio da discrepância de Morozov, a validação cruzada generalizada e o método da curva L.

2. METODOLOGIA

2.1 Decomposição em valores singulares

Para poder descrever os métodos de regularização estudados, foi fundamental o estudo de uma fatoração muito conhecida: a decomposição em valores singulares.

De acordo com Poole [4], temos a seguinte definição:

Definição 1: *Seja A uma matriz $m \times n$. Os valores singulares de A são as raízes quadradas dos autovalores de $A^t A$ e são denotados por $\sigma_1, \dots, \sigma_n$. É convencional ordenar os valores singulares de modo que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$.*

Uma vez introduzidos os valores singulares, pode-se enunciar o seguinte teorema da decomposição em valores singulares (SVD).

Teorema 1: *Toda matriz A , $m \times n$, com valores singulares $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$, pode ser fatorada como produto de três matrizes:*

$$A = U \Sigma V^t$$

onde

- V é uma matriz ortogonal $n \times n$ construída a partir de um conjunto ortonormal de autovetores associados aos autovalores da matriz $A^t A$, dado por $\{v_1, v_2, \dots, v_n\}$;
- U é uma matriz ortogonal $m \times m$, cujos elementos são determinados por $u_i = \frac{1}{\sigma_i} A v_i$, $i = 1, \dots, m$;

Observação 1: *Se A tem r valores singulares não nulos, encontram-se r colunas de U . Caso $r < m$, não haverá uma base de \mathbb{R}^m . Nesse caso, será necessário determinar os $m - r$ vetores de \mathbb{R}^m restantes de forma conveniente, como por exemplo, recorrendo à ortogonalização de Gram-Schmidt tendo por base os versores canônicos.*

- Σ é uma matriz “diagonal” $m \times n$. Supondo que A tem r valores singulares não nulos, teremos Σ da forma:

$$\Sigma = \left[\begin{array}{c|ccc} \overbrace{}^r & & & \\ \hline & D & & \\ \hline & & 0 & \dots & 0 \\ & & \vdots & \ddots & \vdots \\ & & 0 & \dots & 0 \\ \hline 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{array} \right] \begin{array}{l} r \\ m-r \end{array}, \text{ em que } D = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}$$

Observação 2: Os valores singulares encontrados são únicos. O mesmo não pode ser garantido para os vetores u_i e v_i .

É sempre bom lembrar que matrizes ortogonais são muito boas, pois elas são estáveis, ou seja, não amplificam os erros do nosso problema.

Outra forma muito conveniente de escrever a SVD é a forma de produto, na qual a matriz A é representada como soma de matrizes de posto um.

Teorema 2: Seja A uma matriz $m \times n$, com r valores singulares não nulos. Sejam u_1, \dots, u_r as colunas da matriz U , denominados vetores singulares a esquerda de A e v_1, \dots, v_r as colunas da matriz V , denominados vetores singulares a direita de A . Então, a matriz A pode ser escrita na forma:

$$A = \sigma_1 u_1 v_1^t + \dots + \sigma_r u_r v_r^t = \sum_{i=1}^r \sigma_i u_i v_i^t \quad (3)$$

Esta forma de representar a matriz A é denominada forma de produto da SVD.

Observe que r , ou seja, o número de valores singulares não nulos de A , corresponde ao posto da matriz.

A forma de produto (3) da SVD será muito útil no estudo dos métodos de regularização da seção 3.

As demonstrações dos teoremas 1 e 2 podem ser vistas em Poole [4].

2.1.1 Número de condição

Ao lidar com matrizes, existe a possibilidade de se ter uma matriz invertível ou uma matriz não invertível, chamada de matriz singular. No entanto, observa-se certa dificuldade em resolver alguns problemas lineares onde os erros crescem de forma mais relevante do que em outros problemas. Para poder diferenciar estas situações, é introduzido o número de condição da matriz do sistema. Um problema

com um número de condição pequeno é chamado de bem condicionado, ao passo que um problema que possui um número de condição elevado é chamado de mal condicionado.

Problemas mal condicionados são aqueles onde pequenas modificações nos coeficientes resultam em grandes modificações na solução, como visto em (2). Matricialmente, isso ocorre quando a matriz A está muito próxima de ser singular.

Através da SVD, pode-se calcular o número de condição de uma matriz A . Segundo [4], ele é dado por:

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_n}$$

onde σ_1 é o maior valor singular e σ_n o menor valor singular de A . A partir da definição, resulta claro que sempre ocorre que $\text{cond}(A) \geq 1$.

A importância deste número pode ser mostrada na seguinte situação: dado o problema $Ax = b$ e escrevendo-o como:

$$A(x + \delta x) = b + \delta b$$

tem-se a seguinte relação:

$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta b\|}{\|b\|}$$

Observe que, se o número de condição for muito grande, ele irá predominar grande parte do valor da parcela da direita. Com isso, não há como garantir que pequenas variações em b impliquem em pequenas variações em x . Assim, matrizes com números de condição pequenos garantem estabilidade.

2.1.2 Pseudo-inversa

De acordo com Boos [2], temos a seguinte definição:

Definição 2: Sejam A uma matriz $m \times n$ e X uma matriz $n \times m$. Considere as seguintes condições, conhecidas como condições de Penrose:

- I. $AXA = A$;
- II. $XAX = X$;
- III. $(AX)^t = AX$;
- IV. $(XA)^t = XA$.

Se X satisfaz essas quatro propriedades, então X é conhecida como a inversa de Moore-Penrose, ou simplesmente pseudo-inversa, e é denotada por A^\dagger .

Dentre as diversas maneiras de computar a pseudo-inversa, uma delas é usando a SVD da matriz A , e esta é feita de acordo com a seguinte definição dada em [5]:

Definição 3: Seja $A = U\Sigma V^t$ a fatoração SVD de uma matriz A , $m \times n$, onde $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ e D é uma matriz diagonal $r \times r$ que contém os valores singulares diferentes de zero $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ de A .

A pseudo-inversa (ou inversa de Moore-Penrose) de A é a matriz A^\dagger , $n \times m$, definida por

$$A^\dagger = V \Sigma^{-1} U^t$$

onde Σ^{-1} é uma matriz $n \times m$

$$\Sigma^{-1} = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \text{ em que } D^{-1} = \begin{bmatrix} \frac{1}{\sigma_r} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_1} \end{bmatrix}$$

Temos que esta matriz satisfaz as quatro condições de Penrose descritas anteriormente e sua demonstração pode ser encontrada em [2].

3. MÉTODOS DE REGULARIZAÇÃO

Inicialmente, ao falar de métodos de regularização, pergunta-se qual a razão de regularizar um problema. Considere a seguinte situação: existe um conjunto de observações e deve-se recuperar a função que produziu aquelas informações. No entanto, essa informação não é exata, logo métodos tradicionais como $x = A^{-1}b$ não funcionam.

Neste caso, o que deve ser feito? Ao falar de um problema inverso, não interessa a solução exata do problema $Ax = b$, pois os valores observados em b não são exatos, ou seja, na regularização, pode-se permitir uma variação nos dados. Um exemplo é o caso dos quadrados mínimos, onde os dados observados são perturbados e, ao procurar uma solução, não busca-se uma reta que passe por todos os pontos, já que ela não existe, mas sim a reta que melhor aproxima a solução.

A ideia destes métodos é conseguir analisar e encontrar a solução de um problema mal-posto através da solução de um problema associado que é bem-posto. O método de regularização define um operador regularizador que depende de um parâmetro de regularização λ , de forma que, no limite, quando o parâmetro λ tende a zero, é possível garantir a convergência para a solução desejada.

3.1 Regularização nos quadrados mínimos

Um dos métodos de regularização mais antigos é o dos quadrados mínimos. Este método independente de parâmetros, mas também funciona como regularização, pois ao invés de resolver o problema exato, minimiza-se o quadrado do resíduo e é feita uma aproximação.

Dado o problema $Ax = b$, deseja-se encontrar

$$x = \operatorname{argmin} \|Ax - b\|_2^2$$

onde $b = b_{\text{exato}} + \delta b$.

Para solucionar o problema dos quadrados mínimos, usam-se as equações normais:

$$Ax = b$$

$$A^t A x = A^t b$$

$$x = (A^t A)^{-1} A^t b$$

Usando a forma de produto (3) da SVD, tem-se

$$x = A^\dagger b = \sum_{i=1}^r \frac{u_i^t b}{\sigma_i} v_i$$

onde A^\dagger é a pseudo-inversa de A .

Como já foi observado, problemas podem aparecer ao considerar $b = b_{exato} + \delta b$:

$$\begin{aligned} x &= A^\dagger(b_{exato} + \delta b) = A^\dagger b_{exato} + A^\dagger \delta b \\ &= \sum_{i=1}^r \frac{u_i^t b_{exato}}{\sigma_i} v_i + \sum_{i=1}^r \frac{u_i^t \delta b}{\sigma_i} v_i \end{aligned}$$

O erro corresponde à segunda parcela da soma e torna-se dominante quando a componente do erro dos dados está na direção do vetor singular associado ao menor valor singular, pois se o menor valor singular for pequeno, o valor do erro cresce e contaminará a solução. Portanto, deve-se reduzir este efeito indesejável estabilizando a solução usando alguma técnica de regularização.

3.2 Regularização de Tikhonov

Conforme Bazan [1], a regularização de Tikhonov substitui o problema

$$x = \operatorname{argmin} \|Ax - b\|_2^2$$

por

$$x = \operatorname{argmin} \{ \|Ax - b\|_2^2 + \lambda^2 \|L(x - x_0)\|_2^2 \}$$

onde $\|Ax - b\|_2^2$ é o resíduo, $\|L(x - x_0)\|_2^2$ é o tamanho da solução e λ é o parâmetro de regularização. O vetor x_0 é uma aproximação inicial para a solução, caso esteja disponível. Aqui, assume-se o caso mais comum, que é quando não há nenhuma informação inicial, ou seja, tem-se $x_0 = 0$.

O desafio é escolher um parâmetro λ correto tal que x_λ aproxime a solução exata x_{exato} .

Ao procurar o mínimo desta equação, tem-se à equação normal regularizada:

$$x_\lambda = (A^t A + \lambda^2 L^t L)^{-1} A^t b$$

A escolha do operador L depende do problema analisado. Ao buscar a solução de norma mínima, usa-se $L = I$. Deste modo, a equação normal utilizada é:

$$x_\lambda = (A^t A + \lambda^2 I)^{-1} A^t b$$

Note que, para $\lambda = 0$, tem-se a regularização dos mínimos quadrados. Note também que poderia não existir inversa da matriz $(A^t A + \lambda^2 I)$. No entanto, ao escolher $\lambda > 0$, mesmo que λ seja pequeno, essa matriz sempre terá inversa e sempre é possível encontrar uma solução, chamada solução de Tikhonov regularizada.

A solução regularizada pode ser escrita usando a forma de produto (3) da SVD da seguinte forma:

$$x_\lambda = \sum_{i=1}^r f_i \frac{u_i^t b}{\sigma_i} v_i$$

onde $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$ são chamados de filtros de regularização de Tikhonov e r é o posto da matriz A .

A partir da expressão que define f_i , resulta que: $f_i \cong \begin{cases} 1, & \sigma_i \gg \lambda \\ \frac{\sigma_i^2}{\lambda^2}, & \sigma_i \ll \lambda \end{cases}$

Se λ é muito grande, a solução pode não ter incorporado boa parte das informações do problema. Mas se λ é muito pequeno, tem-se $f_i \cong 1$ e provavelmente pouco ruído será filtrado. Deste modo, observa-se a importância de escolher o parâmetro de regularização λ corretamente.

3.3 Regularização através da TSVD

O método da *decomposição em valores singulares truncada (TSVD)* é muito eficiente para reduzir o tamanho de uma matriz e guardar apenas as informações correspondentes aos seus k primeiros valores singulares.

Segundo [4], seja A uma matriz $m \times n$ com $\text{posto}(A) = r$. Escrevendo A na forma de produto (3), tem-se

$$A = \sigma_1 u_1 v_1^t + \dots + \sigma_r u_r v_r^t$$

Defina $k < r$ e escreva:

$$A_k = \sigma_1 u_1 v_1^t + \dots + \sigma_k u_k v_k^t \quad (4)$$

Tendo em vista que cada vetor u_i tem m entradas e cada vetor v_i tem n entradas, este método precisará de um espaço de armazenamento para apenas

$$km + kn + k = k(m + n + 1)$$

números. Neste caso, supõe-se que os valores singulares $\sigma_1, \sigma_2, \dots, \sigma_n$ sejam suficientemente pequenos, ou seja, mesmo sendo ignorados, é possível produzir uma boa aproximação de A .

É possível afirmar que:

- A_k é uma aproximação de A que mantém apenas os primeiros k valores e os vetores singulares correspondentes;
- A_k é a melhor aproximação de posto(k) da matriz A : $\|A - A_k\| = \sigma_{k+1}$.

3.4 Aplicação da TSVD em compressão de imagens

As imagens digitais são extremamente importantes em várias áreas, como na medicina, biologia e astronomia. Sua transmissão é feita na forma de matrizes enormes e elas precisam de muito espaço de armazenamento.

Nesse contexto surge uma das mais impressionantes aplicações da TSVD, que é o seu uso em compressão de imagens. O objetivo principal é reduzir o espaço de armazenamento da imagem e fazer com que sua transmissão seja feita de forma mais rápida sem perder informações essenciais.

Como visto em [3] e [4], o primeiro passo na compressão de uma imagem visual é representá-la como uma matriz numérica. Por exemplo, suponha que você tem uma imagem em preto e branco com um tamanho de 626×421 pixels. Cada pixel é um dos 256 tons de cinza, que podem ser representados por um número entre 0 e 255. Armazenando essa informação em uma matriz A , tem-se uma matriz de tamanho 626×421 , ou seja, são 263.546 números armazenados.

Os menores valores singulares da SVD da matriz A detêm as partes menos importantes da imagem e muitas delas podem ser ignoradas. Isto é feito através da TSVD.

Como visto em (4), pode-se escrever a matriz A em sua forma truncada para algum $k \leq r$:

$$A_k = \sigma_1 u_1 v_1^t + \dots + \sigma_k u_k v_k^t$$

Para o exemplo de 626×421 , ao escolher $k = 20$, são transmitidos apenas os dados correspondentes aos primeiros 20 valores singulares. Então, ao invés de transmitir 263.546 números, é necessário apenas enviar 20 valores singulares, 20 vetores u_1, u_2, \dots, u_{20} em \mathbb{R}^{626} e os 20 vetores v_1, v_2, \dots, v_{20} em \mathbb{R}^{421} , ou seja, um total de

$$20 \times (1 + 626 + 421) = 20.960$$

números. Assim, tem-se uma economia de 92% em relação à matriz original.

3.4.1 Exemplo

A Figura 1 é uma imagem de 626×421 pixels, da mesma forma que foi tratado anteriormente. Esta imagem tem 256 tons de cinza, logo, sua matriz A correspondente é de tamanho 626×421 , com elementos entre 0 e 255.



Figura 1

Ao aproximar A de A_k , obtem-se uma imagem que corresponde aos primeiros k valores singulares de A . A Figura 2 mostra várias dessas imagens para valores de k de 4 a 54. No início, a imagem é muito embaçada, mas rapidamente toma forma. Note que a matriz A_{34} já produz uma boa aproximação da imagem real.

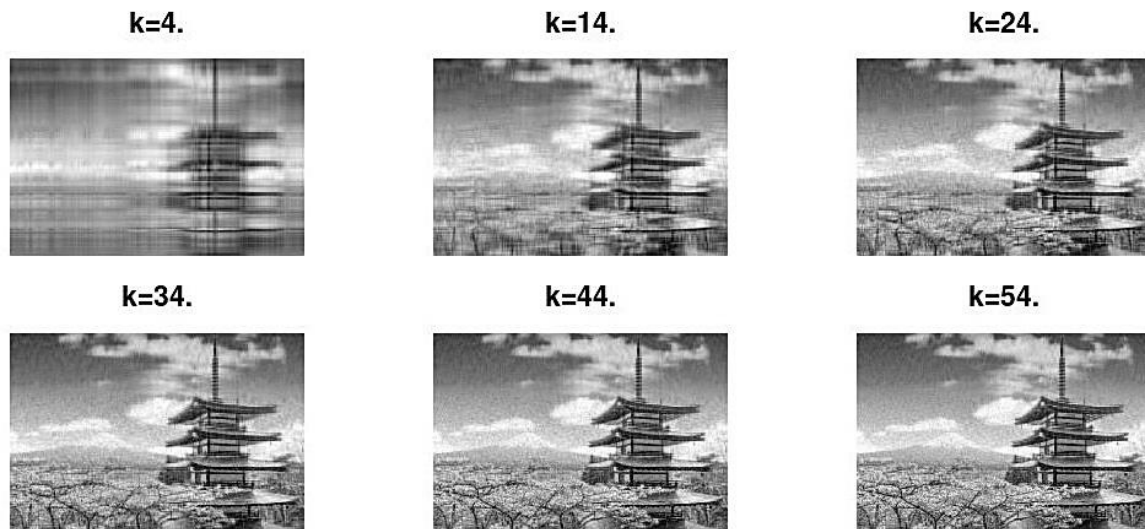


Figura 2

Os primeiros valores singulares de A são $\sigma_1 \cong 77013.83$, $\sigma_2 \cong 12311.74235$, $\sigma_3 \cong 6365.72103$, enquanto os últimos são $\sigma_{420} \cong 3.36397$ e $\sigma_{421} \cong 3.28147$. Os valores singulares menores contribuem muito pouco para a imagem, razão pela qual as aproximações rapidamente parecem tão próximas da imagem original.

O processamento digital da imagem foi realizado em Matlab de acordo com os passos apresentados em [3].

4. CONSIDERAÇÕES FINAIS

Este trabalho pode ser dividido em três partes essenciais. Na primeira parte, foram apresentadas ferramentas fundamentais de álgebra linear como a decomposição em valores singulares e a pseudo-inversa de uma matriz. Na segunda parte, foram introduzidos os métodos de regularização de Tikhonov e a TSVD e, por último, o método da TSVD foi aplicado em um problema de compressão de imagens.

Observa-se que o estudo da fatoração SVD foi de extrema importância para a formulação dos métodos de regularização, pois sua forma de produto permite representar estes métodos de uma forma muito eficiente.

A regularização de Tikhonov é certamente uma das mais populares. Mostrou-se que ela consiste em minimizar um funcional apropriado, visando encontrar uma aproximação para a solução do problema inverso.

Ao falar de imagens digitais, percebe-se que elas são muito importantes em diversas áreas. Como elas são transmitidas na forma de matrizes extensas, é exigido muito espaço de armazenamento. Por isso, torna-se necessário o estudo de métodos que acelerem a transmissão eletrônica de imagens e utilizem menos espaço de armazenamento. Ao analisar esse problema, vantagens foram observadas na reconstrução de uma imagem digital usando um conjunto de dados reduzidos e a regularização através da TSVD se mostrou muito eficiente como ferramenta de compressão, armazenamento e transmissão de imagens digitais.

É evidente que o trabalho realizado exigiu uma série de conhecimentos, de áreas diversas, para que os objetivos fossem alcançados.

Por fim, vale ressaltar que esta experiência de pesquisa, produção de relatórios, orientação e apresentação, juntamente com todos os conteúdos aprendidos no decorrer do trabalho, foram de grande importância para minha formação acadêmica.

REFERÊNCIAS

- [1] BAZAN, F.S.; BORGES, L. **Métodos para problemas inversos de grande porte**. Notas em Matemática Aplicada, Sbmec, Curitiba, v. 39, 2009.
- [2] BOOS, E. **Métodos Iterativos para a Pseudo-Inversa de Moore-Penrose e Aplicações na Resolução de Sistemas Lineares**. Trabalho de Conclusão de Curso – Universidade Federal de Santa Catarina, Florianópolis - SC, 2015.
- [3] COGUBUM, M. C. T. H. **Transmissão de Imagens Utilizando SVD**. Trabalho de Conclusão de Curso – Universidade Federal de São Carlos, Sorocaba - SP, 2013.
- [4] POOLE, D. **Álgebra linear: uma introdução moderna**. Cengage Learning BR, 2da edição, 2016.
- [5] VELHO, H. F. C. **Introdução aos problemas inversos e aplicações em pesquisa espacial**. Escola de verão em Computação Aplicada – LAC-INPE, São José dos Campos-SP, 2008.