



INSTITUTO FEDERAL DA BAHIA PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO

PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

ÁREA DO CONHI ()BIOLÓGICAS	ECIMENTO:	(X) EXATAS	()HUMANAS
PROGRAMA:	() PIBIC	(X) PIVIC	

Título do Projeto: Estudo e Aplicação de Big Data e Machine Learning em

Ciência de Dados.

Nome do Grupo de Pesquisa Cadastrado no Diretório de Grupos de Pesquisa do CNPq: SSMART Research Group.

Orientador: Lauro Cássio Martins de Paula.

Unidade Acadêmica/Departamento: Departamento de Pesquisa Científica / Instituto Federal da Bahia – Campus Santo Antônio de Jesus.

PLANO DE TRABALHO

Edital 06/2017 – Fluxo Contínuo Período 2019/2020

Título do Plano de Trabalho: Implementação de Métodos Iterativos na Linguagem Computacional Python para Solução de Sistemas Lineares.

Aluno: Jorge Mendes de Almeida Filho

Matrícula: 20191TADSSAJ0015 - Aluno do curso TADS do IFBA-SAJ.

1. Introdução

Uma equação linear é um objeto de estudo matemático que contém constantes (coeficientes) e variáveis (incógnitas). Para que uma equação seja considerada linear, ela deve ter a forma $a_1x_1+a_2x_2+\cdots+a_nx_n=b$, onde a é um coeficiente, x é uma incógnita e b é o termo independente. Um determinado conjunto será a solução da equação linear se todos os elementos desse conjunto forem iguais às incógnitas da equação e, ao substituir os elementos desse conjunto nas incógnitas da equação linear, a igualdade $a_1x_1+a_2x_2+\cdots+a_nx_n=b$ for verdadeira.

Um sistema linear é um conjunto de duas ou mais equações lineares. Todo sistema linear contém m equações com n incógnitas cada. Para solucionar um sistema linear, é necessário solucionar todas as equações. Um sistema linear possui a forma $\mathbf{A}\mathbf{x} = \mathbf{b}$, onde \mathbf{A} é a matriz dos coeficientes, \mathbf{x} é o vetor das incógnitas e \mathbf{b} é o vetor dos termos independentes:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases}$$

Sistemas lineares esparsos (nos quais a maioria dos coeficientes são iguais a zero) e de grande porte podem surgir como resultado da modelagem de diversos problemas na ciência da computação e nas engenharias. Existem na literatura diferentes tipos de métodos utilizados para solucionar sistemas lineares. Métodos iterativos são mais indicados para solucionar sistemas grandes e esparsos. Tais métodos realizam aproximações sucessivas em cada iteração para obter uma solução mais precisa para o sistema (FRANCO, 2006). Dois métodos iterativos considerados clássicos são: Jacobi (JUDICE et al., 1996) e Gauss-Seidel (CLAUDIO e MARINS, 1989). Esses métodos são relativamente simples e fáceis de implementar em alguma linguagem de programação. Entretanto, ambos podem apresentar uma convergência lenta ou mesmo não convergirem para sistemas muito grandes. Além disso, eles não podem garantir uma convergência para o sistema linear quando o mesmo não é quadrado (número de linhas igual ao número de colunas).

Outros métodos iterativos têm se mostrado mais eficazes na solução de sistemas lineares grandes e esparsos. Por exemplo, o Gradiente Bi-Conjugado Estabilizado (BiCGStab) é um método iterativo desenvolvido por van der Vorst (VORST, 1992). Ele é uma extensão do método do Gradiente Conjugado e,

além de ser considerado robusto e eficiente, tem sido amplamente utilizado para a solução de sistemas lineares em dinâmica dos fluídos. Um outro método que tem sido bastante utilizado é o Gradiente Bi-Conjugado Estabilizado Híbrido (BiCGStab(2)). O BiCGStab(2) é uma versão estendida do BiCGStab, proposto pelo mesmo autor (SLEIJPEN e VORST, 1995). Ele é um método iterativo não-estacionário, que combina as vantagens do BiCGStab com as de um outro método chamado GMRES. Consequentemente, o BiCGStab(2) é considerado robusto e possui uma convergência superior a do BiCGStab.

Em comparação com métodos exatos, os métodos iterativos são mais indicados para solucionar sistemas lineares esparsos. No entanto, a execução desses métodos normalmente implica em um custo computacional elevado quando aplicados em grandes conjuntos de dados (Big Data). Big Data é um termo empregado na ciência de dados que consiste na análise e interpretação de grandes volumes de dados com vasta variedade. A ciência de dados é uma área de estudo que vai desde à concepção dos dados até à visualização das informações extraídas desses dados. Um dos principais objetivos consiste em extrair insights significativos dos dados e auxiliar na tomada de decisões por meio de análises preditivas. Portanto, o estudo, a implementação, execução e comparação de métodos iterativos para a solução de sistemas lineares grandes e esparsos compreendem uma tarefa importante a ser cientificamente investigada.

2. Objetivos

Pesquisar e selecionar métodos iterativos (possivelmente os métodos Jacobi, Gauss-Seidel, BiCGStab e BiCGStab(2)) para a implementação na linguagem de programação Python.

Simular e comparar tais métodos utilizando pelo menos um grande conjunto de dados de alguma área específica e, se possível, aplicar alguma técnica de análise preditiva nos resultados da simulação para extrair informações que auxiliem na tomada de decisões;

Obter resultados consideráveis e realizar comparações com trabalhos da literatura que compõem o estado da arte, demonstrando uma certa superioridade na qualidade dos resultados obtidos tanto em termos computacionais quanto em termos de convergência.

Publicar e apresentar tais resultados em forma de artigo científico em pelo menos uma conferência nacional na área de computação, além de publicar uma versão estendida do artigo em um periódico científico internacional;

Apresentar à comunidade acadêmica local os principais resultados obtidos com o objetivo de demonstrar a importância da iniciação científica na formação acadêmica do aluno.

3. Metodologia

A princípio, métodos iterativos em destaque na literatura serão pesquisados e investigados. Artigos científicos, livros, teses de doutorado, dissertações de mestrado e monografias serão as principais fontes para a pesquisa científica a ser realizada neste trabalho. Por ser considerada uma linguagem robusta e bastante utilizada por cientistas de dados, a linguagem de programação Python também será estudada e utilizada para a implementação dos métodos.

4. Etapas e Cronograma de Execução

As etapas do trabalho do aluno e seu cronograma estão resumidos na tabela abaixo.

Atividade \ Período	2019				2020							
	Α	S	0	N	D	J	F	М	Α	М	J	J
Revisão Bibliográfica.	Х	х	х	х	х							
Estudo e seleção de métodos iterativos.	х	х	х	х	х							
Estudo dos algoritmos dos métodos iterativos.				х	х	х	х	х				
Implementação dos métodos na linguagem Python.					х	х	х	х	х			
Pesquisa e obtenção de um conjunto de dados para execução dos métodos						х	Х	х	х	х		
Redação do Relatório de Acompanhamento.							х					
Comparação dos métodos							Х	х	X	х	х	
Redação do Relatório Final										х	х	х
Finalização do trabalho e escrita de artigo a ser submetido em algum congresso.												х

5. Resultados Esperados na Execução

Os resultados deste trabalho de iniciação científica compreendem:

- obtenção de resultados significativos por meio da execução de métodos iterativos implementados em linguagem computacional adequada para solucionar sistemas lineares grandes e esparsos;
- a iniciação do aluno na pesquisa e na publicação de trabalhos em congressos e periódicos com o rigor acadêmico-científico apropriado;
- uma possível extensão do tema para o trabalho de conclusão de curso do aluno e uma preparação do mesmo para ingressar num futuro mestrado na mesma área de pesquisa.

6. Referências Bibliográficas

FRANCO, N. B.. Cálculo Numérico. São Paulo: Pearson Prentice Hall, 2006.

JUDICE, J. J. *et al.*, **Sistemas de Equações Lineares**. Departamento de Matemática da Universidade de Coimbra, 1996.

CLAUDIO, D. M.; MARINS, J. M., **Cálculo Numérico Computacional**. São Paulo: Atlas, 1989.

VORST H. A., **Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems**, SIAM Journal of Scientific and Statistical Computing 13 (1992), no. 2, 631–644.

SLEIJPEN, G.; VORST H. A., **Hybrid Bi-Conjugate Gradient Methods for CFD Problems**, Computational Fluid Dynamics Review, 902 (1995).

PAULA, L. C. M., Parallelization of an Iterative Method for Solving Large and Sparse Linear Systems using the CUDA-Matlab Integration, International Conference on Parallel and Distributed Processing Techniques and Applications, (2014), 556-560.