



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E MÉTODOS
QUANTITATIVOS
MESTRADO ACADÊMICO EM MODELAGEM E MÉTODOS QUANTITATIVOS

LUCAS GONÇALVES MONTE

APRIMORAMENTO DO MÉTODO DE ANÁLISE DE COMPONENTES PRINCIPAIS
PARA MATRIZES COM DADOS INTERVALARES

FORTALEZA

2025

LUCAS GONÇALVES MONTE

APRIMORAMENTO DO MÉTODO DE ANÁLISE DE COMPONENTES PRINCIPAIS PARA
MATRIZES COM DADOS INTERVALARES

Dissertação apresentada ao Curso de Mestrado Acadêmico em Modelagem e Métodos Quantitativos do Programa de Pós-Graduação em Modelagem e Métodos Quantitativos do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Pesquisa Operacional.

Orientador: Prof. Dr. Ricardo Coêlho
Silva

Coorientador: Prof. Dr. Weldon Alexander
Lodwick

FORTALEZA

2025

LUCAS GONÇALVES MONTE

APRIMORAMENTO DO MÉTODO DE ANÁLISE DE COMPONENTES PRINCIPAIS PARA
MATRIZES COM DADOS INTERVALARES

Dissertação apresentada ao Curso de Mestrado Acadêmico em Modelagem e Métodos Quantitativos do Programa de Pós-Graduação em Modelagem e Métodos Quantitativos do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Pesquisa Operacional.

Aprovada em: 01/04/2025

BANCA EXAMINADORA

Prof. Dr. Ricardo Coêlho Silva (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Weldon Alexander Lodwick (Coorientador)
University of Colorado at Denver (CU Denver)

Profa. Dra. Marina Tuyako Mizukosh
Universidade Federal de Goiás (UFG)

Prof. Dr. Gualberto Segundo Agamez Montalvo
Universidade Federal do Ceará (UFC)

A todos que estiveram comigo durante esta caminhada, com contribuições diretas ou indiretas, e que me possibilitaram concluir essa etapa. A todas as pessoas que defendem a Universidade Pública, gratuita e com qualidade.

AGRADECIMENTOS

Ao Prof. Dr. Ricardo Coêlho Silva que me orientou nesta pesquisa e que nunca deixou que caísse no esquecimento a motivação inicial de todo o projeto. Sempre me lembrando, a cada reunião, de retornar à pesquisa da monografia e buscar com essa nova pesquisa dar as respostas para os questionamentos que tinham ficado abertos. Isso foi fundamental para que, ao final, não tivesse apenas uma pesquisa de qualidade, mas uma pesquisa que eu conseguisse me ver nela e ver o caminho que estou trilhando.

Ao Prof. Dr. Weldon Alexander Lodwick que coorientou a pesquisa, compartilhou seus conhecimentos e permitiu que eu me sentisse desafiado a dar passos mais ousados na pesquisa. Dessa forma, sinto que evoluí, consideravelmente, neste caminho para me tornar um cientista, construindo bases sólidas.

À Prof. Dra. Marina Tuyako Mizukosh e ao Prof. Dr. Gualberto Segundo Agamez Montalvo que trouxeram apontamentos fundamentais para que a pesquisa atingisse o nível de qualidade desejado para que seja uma pesquisa de referência. Todos os apontamentos geraram inquietações que me levaram a melhorar o rigor e a qualidade do trabalho.

Ao Dr. Hermany Rosa Vieira e ao Prof. Dr. Julio Francisco Barros Neto que fizeram questionamentos na defesa da monografia e a partir disto surgiu a proposta desta pesquisa. Logo, são fundamentais no desenvolvimento deste trabalho.

Ao corpo docente do Departamento de Estatística e Matemática Aplicada (DEMA) e aos servidores técnico-administrativos em educação (TAE) desta instituição que permitiram o desenvolvimento de cada conhecimento, a superação de cada etapa e que, ao disporem do seu tempo, proporcionaram a evolução deste trabalho e que este trabalho possa trazer para a sociedade o retorno investido por esses profissionais.

Agradeço a todos os familiares, as amigas e pessoas próximas que, ao longo desta jornada, não me deixaram desistir. Que pacientemente, mesmo sem entender a pesquisa, ouviam cada explicação de cada etapa e, assim, permitiam que eu me ouvisse e percebesse os pontos a melhorar. Por fim, agradeço à minha avó, D. Milicas, à minha tia Nena e, no coração e na saudade, à minha mãe, Iracema, que, com certeza, sonharam com a conclusão deste trabalho.

"Toda transformação é uma passagem da quantidade à qualidade" (Engels, 1883)

RESUMO

A Análise de Componente Principal, em inglês *Principal Component Analysis* (PCA), é uma técnica estatística multivariada que realiza uma transformação linear em um conjunto de dados composto por observações e variáveis inter-relacionadas, formando um novo conjunto de dados. Este novo conjunto possui variáveis não correlacionadas ou com menor correlação, podendo ter dimensão menor do que o conjunto original. Os principais objetivos da técnica de PCA são: (i) extrair informações mais pertinentes do conjunto observado; (ii) comprimir este conjunto; e (iii) simplificar a descrição do conjunto de dados. É uma técnica bastante utilizada em diversos contextos, como desenvolvimento de escalas psicométricas, neurociência, identificação de falhas em processos, agrupamento ou classificação de informação, dentre outros. Contudo, no dia a dia, é comum encontrarmos situações em que o conjunto de dados possui valores incertos. Dessa forma, o método da PCA Clássica terá dificuldade em fornecer uma solução satisfatória diante de dados incertos sem um tratamento adequado, sendo a matemática intervalar uma maneira de tratar essa incerteza nos dados. Assim, é interessante desenvolvermos técnicas para a PCA em seu contexto intervalar de modo que a incerteza possa ser considerada em sua análise de dados e isso precisa ser feito sem aumentar a dimensão da matriz de dados a fim de evitar um grande aumento do custo computacional. Logo, é proposta uma nova modelagem da matriz de dados intervalares para encontrar a matriz de covariância adequada e, a partir desta, aplicar os métodos para encontrar autovalores e autovetores intervalares de tal modo que possa aplicar uma transformação linear e obter componentes principais intervalares.

Palavras-chave: componentes principais; análise de intervalos (matemática); análise de componente principal intervalar; autovalor intervalar; autovetor intervalar.

ABSTRACT

Principal Component Analysis (PCA) is a multivariate statistical technique that performs a linear transformation on a data set composed of observations and interrelated variables, which forms a new data set. This new set has uncorrelated or less correlated variables, and its dimension may be smaller than the original set. The main objectives of the PCA technique are: (i) to extract more relevant information from the observed set; (ii) to compress this set; and (iii) to simplify the description of the data set. It is a technique widely used in several contexts, such as the development of psychometric scales, neuroscience, identification of process failures, grouping or classification of information, among others. However, in everyday life, it is common to encounter situations in which the data set has uncertain values. So, the Classical PCA method will have difficulty in providing a satisfactory solution without adequate treatment from uncertain data, which may be dealt with by using interval mathematics. Thus, it is interesting to develop techniques for PCA in its interval context so that uncertainty can be considered in its data analysis, and this needs to be done without increasing the dimension of the data matrix in order to avoid a large increase in computational cost. Therefore, a new modeling of the interval data matrix is proposed to find the appropriate covariance matrix and, from this, apply the methods to find interval eigenvalues and eigenvectors in such a way that a linear transformation can be applied and interval principal components obtained.

Keywords: principal component analysis; interval analysis (mathematics); interval principal component analysis; interval eigenvalue; interval eigenvector.

LISTA DE FIGURAS

Figura 1 – Distribuição de 100 observações e distribuição das componentes principais .	13
Figura 2 – Distribuição de 100 observações com três variáveis e distribuição das componentes principais	14
Figura 3 – Produção de artigos com tema PCA ao longo do ano	19
Figura 4 – Quantidade de citações por ano para PCA Intervalar ou Difuso	20
Figura 5 – Grafo das ocorrências das palavras-chave da PCA Intervalar ou Difuso . . .	21
Figura 6 – Gráfico das ocorrências da palavras-chave PCA Intervalar	22
Figura 7 – Grafo das ocorrências das palavras-chave da PCA Intervalar	23
Figura 8 – Gráfico das ocorrências da palavras-chave PCA Difuso	24
Figura 9 – Grafo das ocorrências das palavras-chave da PCA Difuso	25
Figura 10 – Autovetor Intervalar - Método SL	48
Figura 11 – Normalização do autovetor intervalar	49
Figura 12 – Autovetor Intervalar Normalizado - Método SL	49

LISTA DE TABELAS

Tabela 1 – Fontes mais relevantes para o termo PCA	19
Tabela 2 – Fontes mais relevantes da PCA Intervalar ou Difusa	20
Tabela 3 – Fontes mais relevantes da PCA Intervalar	22
Tabela 4 – Fontes mais relevantes da PCA Difuso	24
Tabela 5 – Assertividade de λ_1 e λ_2	45
Tabela 6 – Média das diferenças	73
Tabela 7 – Parâmetros apresentados no artigo	74
Tabela 8 – Comparativo de autovetor intervalar com clássico	76
Tabela 9 – Desempenho dos Candidatos por Indicadores	78
Tabela 10 – Comparação dos candidatos entre o dia 1 e o dia 2	80
Tabela 11 – Desempenho do Candidato 8	80
Tabela 12 – Desempenho do Candidato 11	80

LISTA DE ABREVIATURAS E SIGLAS

CIA *Constraint Interval Arithmetic*

IPCA *Interval Principal Component Analysis*

PACCE Programa de Aprendizagem Cooperativa em Células Estudantis

PCA *Principal Component Analysis*

SL *Single-level*

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivações	15
1.2	Estrutura	16
2	REVISÃO BIBLIOGRÁFICA	17
2.1	Análise para a palavra-chave <i>principal component analysis</i>	18
2.2	Análise para PCA intervalar ou difusa	19
2.2.1	<i>Artigos para PCA intervalar</i>	21
2.2.2	<i>Artigos para PCA difusa</i>	23
2.3	Trabalhos Importantes	25
3	ANÁLISE DE COMPONENTE PRINCIPAL	27
3.1	Contexto clássico	27
3.2	Contexto intervalar	29
3.2.1	<i>Método Deif</i>	35
3.2.2	<i>Método Deif com Leverrier-Faddeev</i>	37
3.2.3	<i>Método Single-level (SL)</i>	40
3.2.4	<i>Método Constraint Interval Arithmetic (CIA)</i>	41
3.2.5	<i>Comparação entre métodos</i>	43
3.2.6	<i>Escolha do Autovalor</i>	45
4	COMPONENTE PRINCIPAL INTERVALAR	47
4.1	Método Numérico	47
4.1.1	<i>SL - Autovetor</i>	47
4.1.2	<i>CIA - Autovetor</i>	50
4.1.3	<i>Deif com Leverrier-Faddeev - Autovetor</i>	51
4.2	Método Analítico	52
4.3	Método Analítico Generalizado	55
4.3.1	$a \leq b < 0$	57
4.3.2	$0 < a \leq b$	63
4.4	PCA Intervalar	69
5	APLICAÇÃO DE COMPONENTES PRINCIPAIS INTERVALARES	74
5.1	Avaliação de qualidade da água	74

5.2	Análise psicométrica	77
6	CONCLUSÕES E TRABALHOS FUTUROS	82
	REFERÊNCIAS	84

1 INTRODUÇÃO

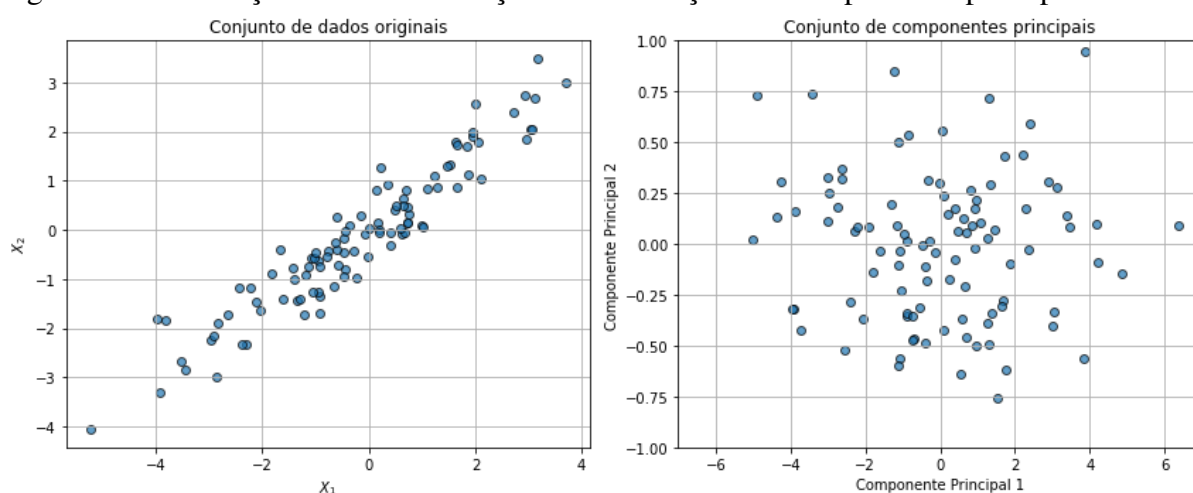
A Análise de Componentes Principais, em inglês *Principal Component Analysis* (PCA), é uma técnica de análise estatística multivariada proposta no começo do século XX, inicialmente, por Pearson (1901) e aprimorada por Hotelling (1933) (Abdi; Williams, 2010).

Esta técnica consiste em aplicar uma transformação linear em um conjunto de dados composto por observações e variáveis inter-relacionadas, formando um novo conjunto de dados. Este novo conjunto de dados tem variáveis não correlacionadas ou com menor correlação, de modo que esse novo conjunto pode possuir dimensão menor do que o conjunto original e as novas variáveis estejam ordenadas de forma decrescente representando da maior até a menor variação do conjunto de dados (Jolliffe, 2013). Ou seja, podemos realizar uma projeção do conjunto de dados originais em um novo conjunto de dados.

Em resumo, a técnica possui como objetivos: extrair as informações mais pertinentes do conjunto de dados observado, comprimir este conjunto de dados baseado nessas informações mais pertinentes, simplificar a descrição do conjunto de dados e permitir analisar a estrutura dessas observações (Abdi; Williams, 2010).

A fim de ilustrarmos a técnica, consideraremos o exemplo ilustrado na Figura 1. Nele, temos um conjunto de 100 observações com duas variáveis distribuídas e ao lado a distribuição do conjunto das componentes principais:

Figura 1 – Distribuição de 100 observações e distribuição das componentes principais



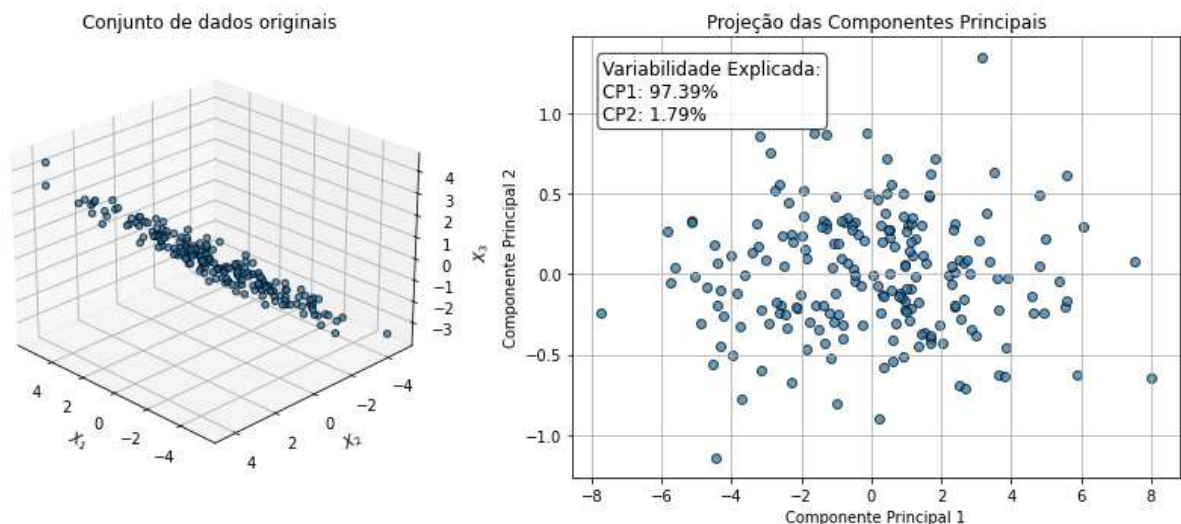
Fonte: próprio autor.

Perceba que os dados da primeira componente principal estão variando, aproximadamente, entre -6 e 6, enquanto os dados da segunda estão variando, aproximadamente, entre -1 e

1. Logo, a primeira componente representa a maior variabilidade dos dados, enquanto a segunda componente, a segunda maior variabilidade. Caso houvesse outras variáveis, poderíamos ter outras componentes principais.

Por exemplo, considere a distribuição a seguir com três variáveis e 100 observações. Perceba que estão sendo consideradas apenas duas componentes principais, porque a primeira representa 97,39% da variabilidade do conjunto original, enquanto a segunda representa 1,79% da variabilidade do conjunto original. Logo, as duas representam mais de 99% da variabilidade e, por isso, é possível reduzir a dimensão do conjunto original:

Figura 2 – Distribuição de 100 observações com três variáveis e distribuição das componentes principais



Fonte: próprio autor.

Assim, a PCA pode ser encontrada nas mais diversas aplicações como, por exemplo, o desenvolvimento de escalas psicométricas (Kaplan; Saccuzzo, 2012), teoria da ecologia social (Shevky *et al.*, 1949), genética populacional (Stephens; Novembre, 2008), neurociência (Chapin; Nicolelis, 1999), dentre outras. O potencial da técnica de redução de dimensão também auxilia na identificação de *outliers*, agrupamento, classificação, identificação de padrões e, conseqüentemente, é uma ferramenta interessante para atuar lado a lado com as diversas técnicas de computação flexível como, por exemplo, redes neurais e sistemas difusos.

Diante destas potencialidades e aplicações da PCA, é de se esperar a incerteza nos dados observados durante a modelagem de problemas reais, pois são aplicações muito próximas do cotidiano. Ou seja, no cotidiano iremos observar bases de dados que possuem parâmetros calculados/obtidos em intervalos. Então, no intuito de tratar matematicamente esse tipo de incerteza nos dados, pode-se usar algumas técnicas baseadas em teoria de conjuntos *fuzzy* e

matemática intervalar, em que as informações das variáveis estão contidas em intervalos, que será o foco deste trabalho.

Como possuímos uma base de dados intervalar, é interessante que a matriz de covariância considere os intervalos, logo será, também, uma matriz intervalar. Para realizar este processo e garantir a natureza intervalar, podemos utilizar a modelagem através da aritmética intervalar restrita e explorar as possíveis matrizes clássicas em uma matriz intervalar.

Então, para desenvolver a PCA aplicada à matriz intervalar, precisaremos encontrar seus autovalores e os autovetores associados. Por isso, aplicaremos técnicas apropriadas para obter autovalores e seus autovetores associados intervalares, a saber: os Métodos Deif, *Single-Level* e *Constraint Interval Arithmetic*.

1.1 Motivações

Dentre as diversas áreas de aplicação da PCA, a psicometria é uma área que merece atenção especial. Dentre as escalas psicométricas existentes na literatura, podemos destacar duas: Thurstone e Likert. A escala de Thurstone é uma das primeiras a utilizar a análise de componentes principais para realizar seus estudos (Kaplan; Saccuzzo, 2012). Contudo, por ser considerada muito trabalhosa e poder gerar dúvidas durante a utilização, foi proposta por Likert (1932) outra escala como aprimoramento para as mesmas aplicações. A escala Likert, como ficou conhecida, é mais intuitiva e tenta melhorar a captura da intensidade das atitudes mensuradas.

Contudo, apesar de ser muito utilizada, hoje a Escala Likert possui algumas desvantagens como, por exemplo, a incerteza entre ser uma escala ordinal ou intervalar durante o processo de análise. A necessidade de obter uma resposta clássica força que a incerteza da atitude mensurada seja aproximada (Monte, 2020). Diante desse cenário, Monte (2020) realizou levantamento de modelagens utilizando sistemas difusos para a Escala Likert. Apesar de existirem modelagens na literatura, pode-se questionar o contexto em que é preciso classificar ou categorizar o público respondente.

Assim, visando tratar de maneira adequada a incerteza, precisamos pensar em técnicas da PCA que considerem o contexto difuso. Entretanto, uma variável difusa pode ser, antes de tudo, modelada como uma variável intervalar com pertinência 1 para todos os elementos do intervalo. Então, é preciso primeiro aprimorar técnicas de PCA para o contexto intervalar e, assim, será possível dar um passo à frente para o contexto difuso, haja vista que intervalar e

difuso possuem relações bem próximas em suas aritméticas (Lodwick, 2003), apesar de serem áreas distintas.

1.2 Estrutura

Este trabalho divide-se em seis capítulos. No Capítulo 1, a Introdução traz um breve resumo sobre a técnica de PCA, suas potencialidades, o contexto da modelagem intervalar e as motivações para desenvolver a pesquisa. Após o conteúdo introdutório, encontra-se a Revisão Bibliográfica (Capítulo 2) com o desenvolvimento da busca por palavras-chave, análise bibliométrica dos conteúdos publicados ao longo do tempo e análise de trabalhos mais relevantes. Em seguida, no Capítulo 3, é apresentada a Análise de Componente Principal em seu contexto clássico e os métodos para resolução no contexto intervalar. O Capítulo 4 apresenta como construir o autovetor intervalar associado ao autovalor, também intervalar, e realiza uma transformação linear após definidas as componentes principais. No Capítulo 5, teremos duas aplicações em situações reais no intuito de ilustrar o funcionamento e a eficiência dos métodos apresentados nos capítulos 3 e 4. Por fim, o Capítulo 6 apresenta as conclusões e oportunidades para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

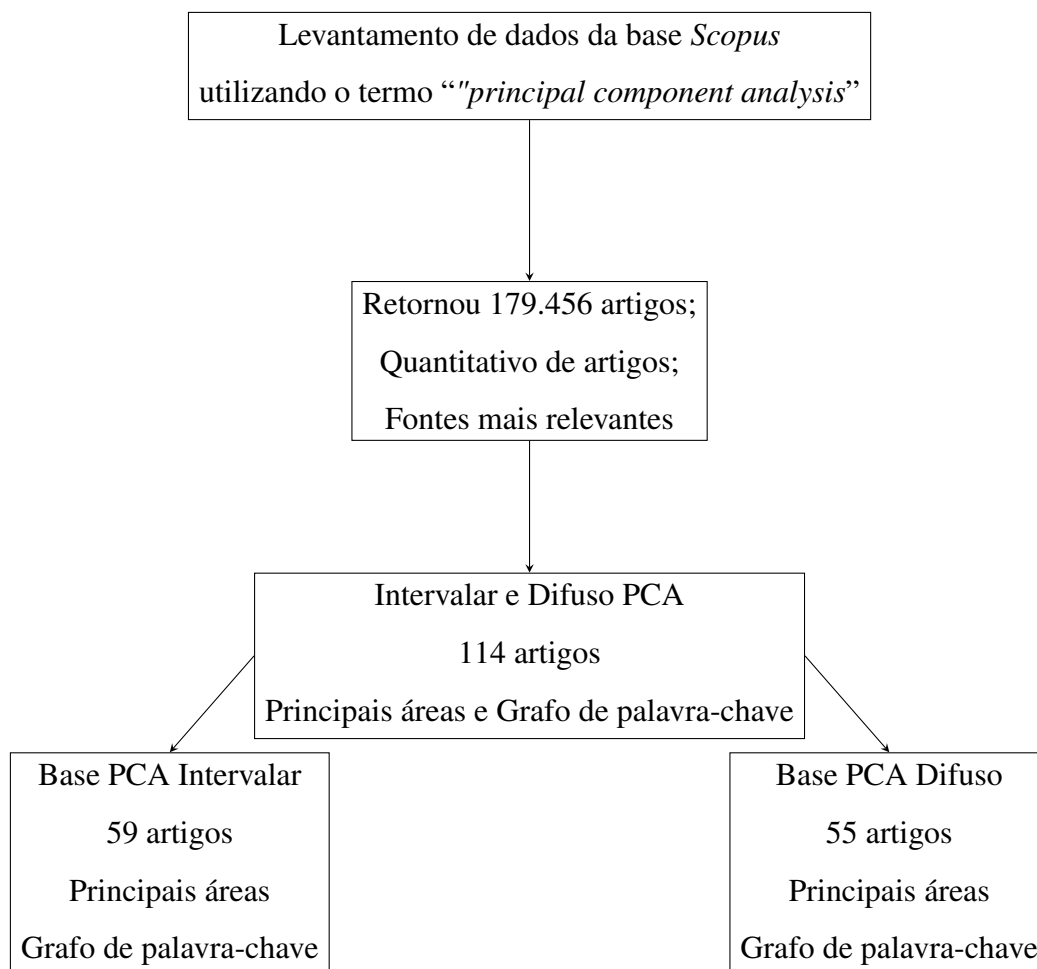
Como a PCA possui diversas aplicações nas mais variadas áreas e seu desenvolvimento inicial é do começo do século XX, é de se esperar que o volume de produções científicas seja bastante significativo. Na base de dados *Scopus*, considerando artigos publicados em língua inglesa e utilizando o termo “*principal component analysis*” é possível encontrar mais de 200.000 artigos nas mais variadas áreas nos últimos 20 anos. Diante deste volume, podemos pensar em realizar uma análise prévia inspirada nos modelos de Revisão Sistemática. Afinal, essa metodologia possui como objetivo “identificar, selecionar, avaliar, interpretar e sumarizar estudos disponíveis considerados relevantes para um tópico de pesquisa ou fenômeno de interesse” (Nakagawa *et al.*, 2017).

Para realizar este procedimento, foi realizada a busca por artigos disponíveis em língua inglesa na base de dados *Scopus* utilizando os termos “*principal component analysis*”, “*interval principal component analysis*” e “*fuzzy principal component analysis*”. A opção por estas palavras-chave deve-se ao fato de que estas palavras isoladamente, mesmo com a aplicação de operadores *booleanos*, poderiam retornar artigos muito destoantes do que é buscado para a PCA. A análise dos trabalhos encontrados foi conduzida através da ferramenta *Bibliometrix* disponível para a linguagem de programação R (Aria; Cuccurullo, 2017).

Como o volume de artigos para a palavra-chave “*principal component analysis*” é bastante extenso, a base de dados não nos permitiu exportar todo o conjunto de mais de 200.000 artigos. Por isso, no buscador, foi marcada a opção de ordenamento pela relevância dos trabalhos a fim de considerarmos os trabalhos mais relevantes para esta palavra-chave. Ao todo, foi possível analisar 179.456 artigos.

Como existe uma relação próxima entre conjuntos intervalares e conjuntos difusos (Lodwick, 2003) podemos verificar as palavras-chave “*interval principal component analysis*” e “*fuzzy principal component analysis*” conjuntamente através do operador *booleano OR*, a fim de avaliarmos possibilidades da técnica de PCA em contexto impreciso ou intervalar, e depois segmentar as duas palavras-chave. Dessa forma, encontramos 114 artigos, sendo 59 trabalhos associados à palavra-chave “*interval principal component analysis*” e 55 artigos associados à palavra-chave “*fuzzy principal component analysis*”.

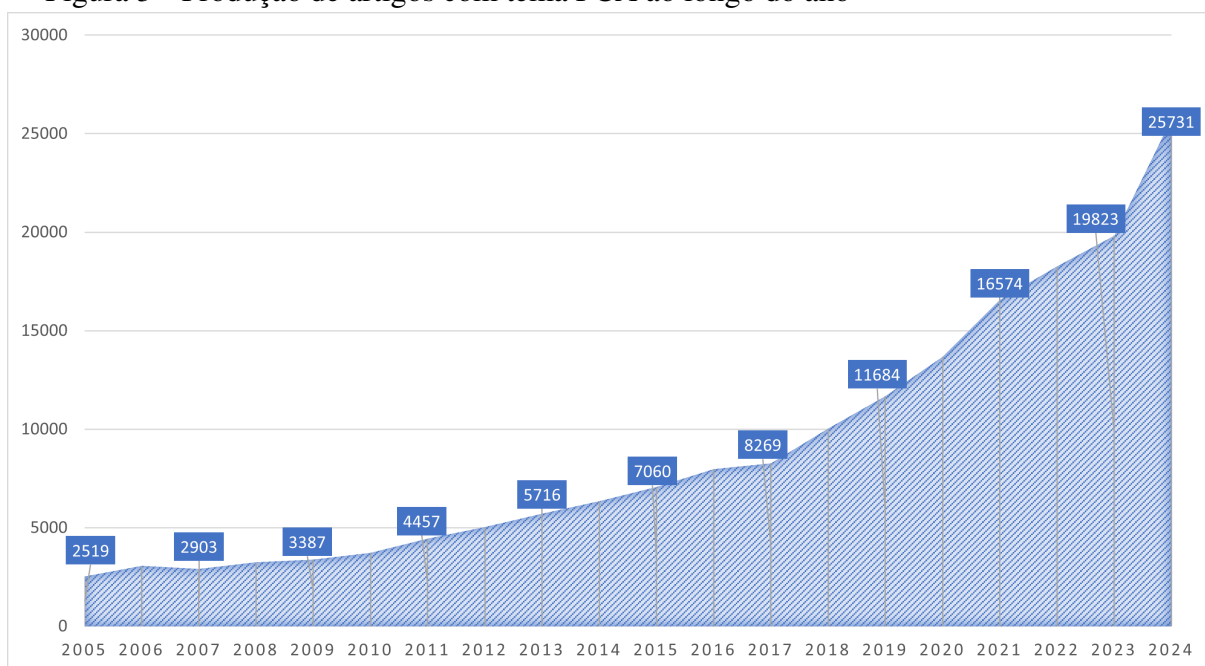
Assim, nossa análise pode ser melhor visualizada no fluxograma a seguir:



2.1 Análise para a palavra-chave *principal component analysis*

Ao todo foram analisados 179.456 artigos encontrados na base *Scopus* de 2005 a 2024. A análise foi realizada utilizando a ferramenta *Bibliometrix*. O resultado é de uma taxa de crescimento médio de 60,5% ao ano. Isso mostra o quanto a PCA é uma técnica bastante utilizada, conforme apresentado na Figura 3:

Figura 3 – Produção de artigos com tema PCA ao longo do ano



Fonte: Próprio autor

Ao olharmos as fontes mais relevantes na Tabela 1, encontramos a seguinte distribuição:

Tabela 1 – Fontes mais relevantes para o termo PCA

Fonte	%
SCIENCE OF THE TOTAL ENVIRONMENT	24,95%
FOOD CHEMISTRY	17,06%
CHEMOSPHERE	8,81%
FOOD RESEARCH INTERNATIONAL	8,36%
ENVIRONMENTAL POLLUTION	8,01%
ECOLOGICAL INDICATORS	7,14%
NEUROCOMPUTING	6,98%
LWT	6,54%
JOURNAL OF HAZARDOUS MATERIALS	6,27%
NEUROIMAGE	5,88%

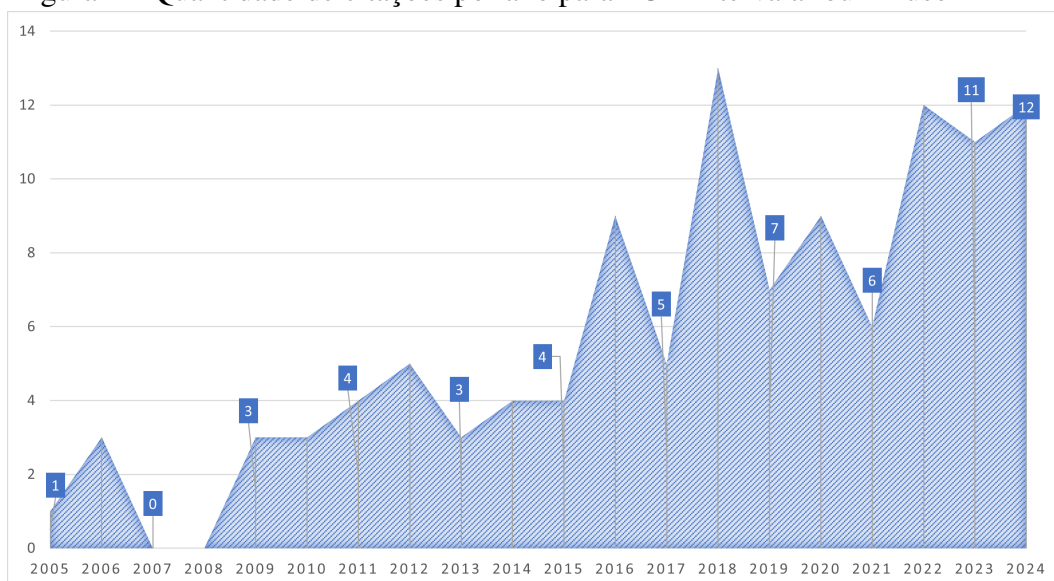
Fonte: Bibliometrix

2.2 Análise para PCA intervalar ou difusa

As palavras-chave *interval principal component analysis* ou *fuzzy principal component analysis* retornaram 114 artigos. Apesar de termos registros a partir de 2005, a área apresenta consolidação a partir de 2009 e demonstra estabilidade média nos últimos três anos, conforme apresentado na Figura 4. Segundo análise do *Bibliometrix*, a taxa de crescimento

médio é de 13,97%:

Figura 4 – Quantidade de citações por ano para PCA Intervalar ou Difuso



Fonte: Próprio autor.

As fontes mais relevantes para esta busca são apresentadas na Tabela 2:

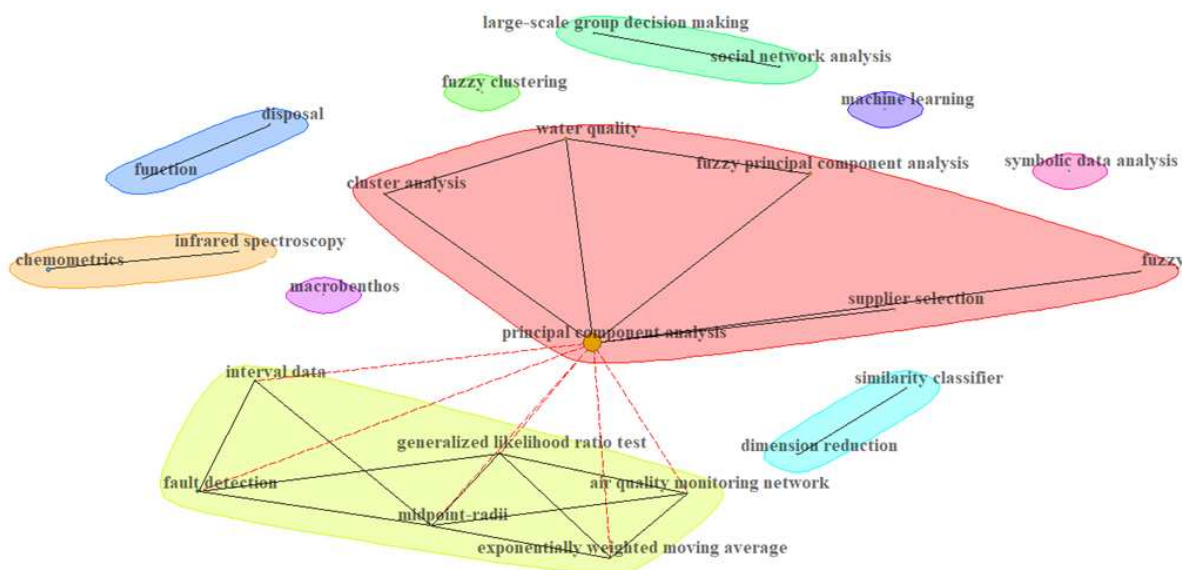
Tabela 2 – Fontes mais relevantes da PCA Intervalar ou Difusa

Fonte	Qtd.
SCIENCE OF THE TOTAL ENVIRONMENT	15,38%
APPLIED SOFT COMPUTING	12,82%
INFORMATION SCIENCES	12,82%
CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS	10,26%
DESALINATION AND WATER TREATMENT	10,26%
PROCEDIA COMPUTER SCIENCE	10,26%
KNOWLEDGE-BASED SYSTEMS	7,69%
MARINE POLLUTION BULLETIN	7,69%
TALANTA	7,69%
AUTOMATION IN CONSTRUCTION	5,13%

Fonte: Bibliometrix.

As revistas já nos dão fortes indícios das principais áreas que estão aplicando e desenvolvendo a PCA intervalar ou difusa. Contudo, podemos ver as correlações entre as palavras-chave. Então, utilizando o *Bibliometrix*, e considerando um grau de no mínimo dez, ou seja, palavras-chave que aparecem juntas pelo menos dez vezes, temos o seguinte grafo na Figura 5:

Figura 5 – Grafo das ocorrências das palavras-chave da PCA Intervalar ou Difuso



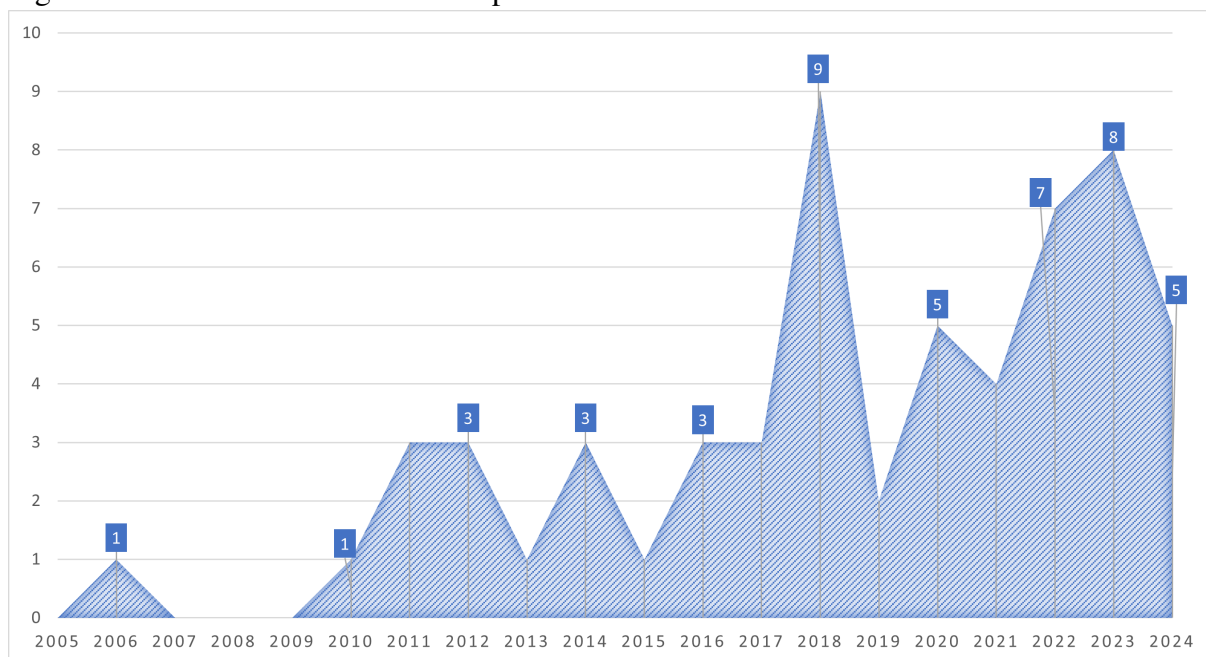
Fonte: Bibliometrix.

O grafo traz para cada cor um grupo de conexões de palavras-chave. É possível perceber, através do grafo, que o termo *principal component analysis* é o que mais aparece e conecta os dois principais grupos através das palavras dados intervalares, raio do ponto médio, classificação, detecção de falhas, difuso, análise de componentes principais difusas, dentre outros. Já os outros grupos evidenciam as outras áreas que possuem relevância como tratamento de quimioterapia, aprendizado de máquina, classificação difusa e análises de interações sociais.

2.2.1 Artigos para PCA intervalar

Avaliando os 59 trabalhos associados à palavra-chave PCA intervalar, temos a seguinte distribuição na Figura 6:

Figura 6 – Gráfico das ocorrências da palavras-chave PCA Intervalar



Fonte: Bibliometrix.

Percebemos que os trabalhos associados à palavra-chave de PCA intervalar começaram a ter maior frequência a partir de 2010. E, entre oscilações nos anos, o tema vem ganhando espaço, principalmente nas revistas apresentadas na Tabela 3:

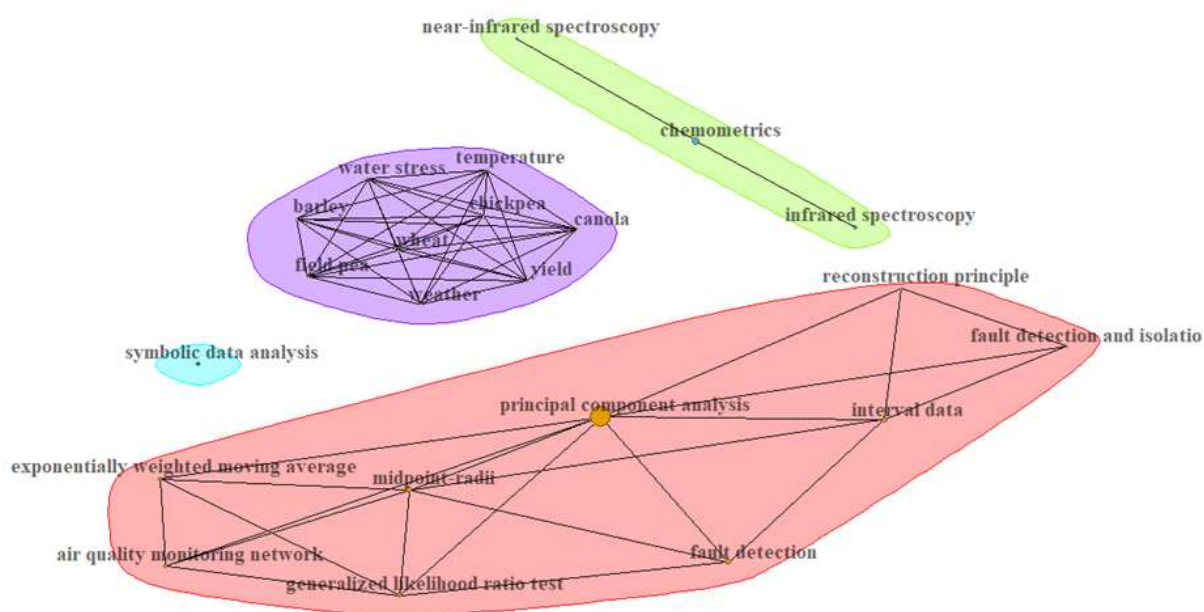
Tabela 3 – Fontes mais relevantes da PCA Intervalar

Fonte	Qtd.
SCIENCE OF THE TOTAL ENVIRONMENT	16,67%
BIORESOURCE TECHNOLOGY	11,11%
CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS	11,11%
FOOD QUALITY AND PREFERENCE	11,11%
JOURNAL OF PROCESS CONTROL	11,11%
TALANTA	11,11%
THE JOURNAL OF SUPERCRITICAL FLUIDS	11,11%
AGRICULTURAL AND FOREST METEOROLOGY	5,56%
ANIMAL BEHAVIOUR	5,56%
ANNALS OF NUCLEAR ENERGY	5,56%

Fonte: Bibliometrix.

Utilizando o *Bibliometrix*, e considerando um grau de no mínimo nove, ou seja, palavras-chave que aparecem juntas pelo menos nove vezes, temos o seguinte grafo na Figura 7:

Figura 7 – Grafo das ocorrências das palavras-chave da PCA Intervalar



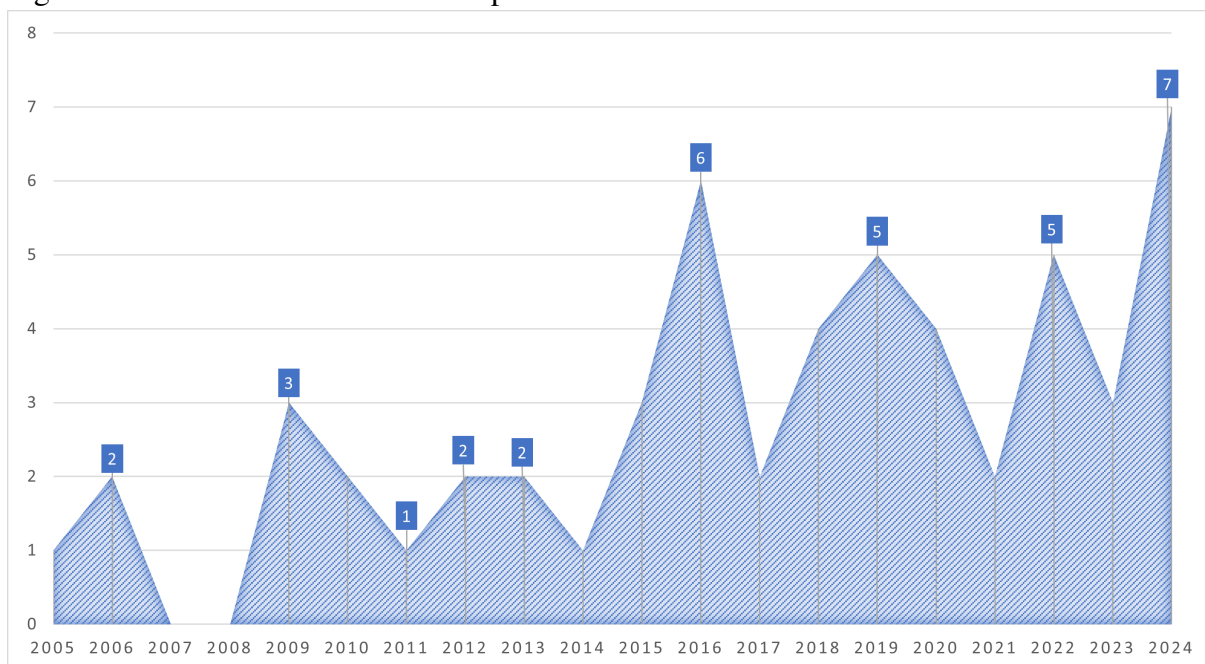
Fonte: Bibliometrix.

Assim como no anterior, a palavra-chave *principal component analysis* possui grande relevância e está associada à detecção de falhas, dados intervalares, dentre outros. Também é possível verificar que a categoria associada à pesquisa com quimioterapia aparece novamente e há destaque para a categoria associada à produção de alimentos que possui vários elementos interligados.

2.2.2 Artigos para PCA difusa

Avaliando os 55 trabalhos associados à palavra-chave PCA difuso, temos a seguinte distribuição na Figura 8:

Figura 8 – Gráfico das ocorrências da palavras-chave PCA Difuso



Fonte: Bibliometrix.

Percebemos que os trabalhos associados à palavra-chave de PCA difuso começaram a ter maior frequência a partir de 2009. E, mesmo apresentando oscilações ao longo dos anos, o tema vem ganhando espaço, principalmente nas revistas apresentadas na Tabela 4:

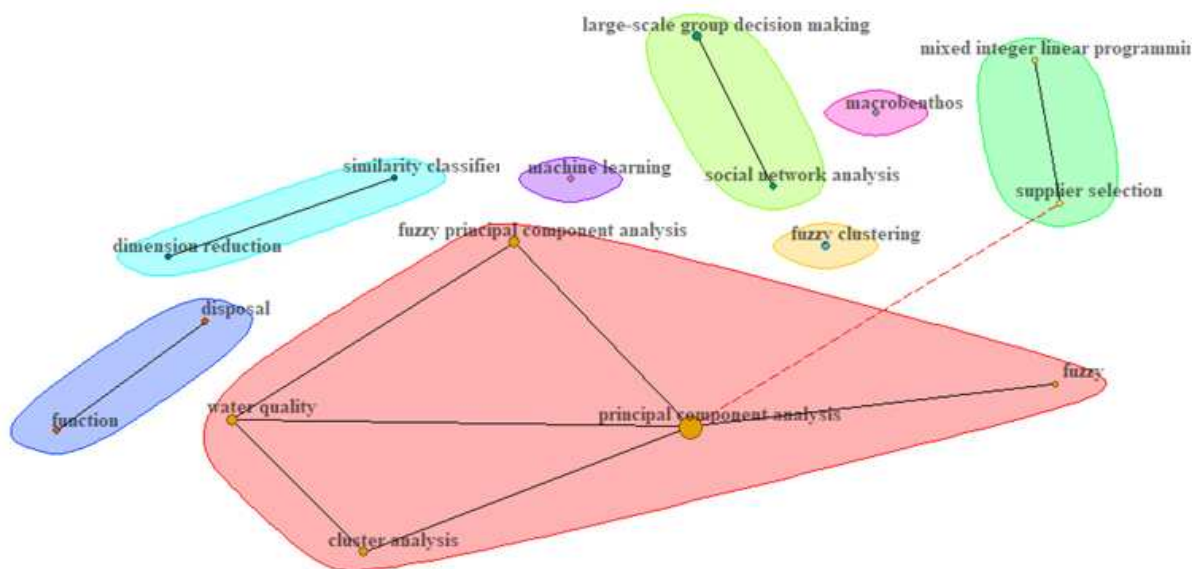
Tabela 4 – Fontes mais relevantes da PCA Difuso

Fonte	Qtd.
APPLIED SOFT COMPUTING	16,13%
DESALINATION AND WATER TREATMENT	12,90%
INFORMATION SCIENCES	12,90%
KNOWLEDGE-BASED SYSTEMS	9,68%
MARINE POLLUTION BULLETIN	9,68%
PROCEDIA COMPUTER SCIENCE	9,68%
SCIENCE OF THE TOTAL ENVIRONMENT	9,68%
AUTOMATION IN CONSTRUCTION	6,45%
CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS	6,45%
COMPUTERS & INDUSTRIAL ENGINEERING	6,45%

Fonte: Bibliometrix.

Utilizando o *Bibliometrix*, e considerando um grau de no mínimo nove, ou seja, palavras-chave que aparecem juntas pelo menos nove vezes, temos o seguinte grafo na Figura 9:

Figura 9 – Grafo das ocorrências das palavras-chave da PCA Difuso



Fonte: Bibliometrix.

Assim como no anterior, a palavra-chave *principal component analysis* possui grande relevância e está associada à versão difusa da técnica de PCA, análise de agrupamento, identificação da qualidade de água, dentre outros. Também é possível verificar que a categoria associada à análise de redes sociais aparece novamente e há destaque nas áreas de agrupamento difuso, aprendizado de máquina, dentre outros, mostrando as possibilidades de áreas de pesquisa relacionadas com o tema.

2.3 Trabalhos Importantes

Um dos trabalhos que aparece como de grande relevância na base *Scopus* é o trabalho de Serrao *et al.* (2023). Este trabalho busca aprimorar a PCA para o contexto específico da análise simbólica de dados. Serrao *et al.* (2023) busca construir a matriz de covariância para a situação-problema e esta, por ser intervalar, é analisada utilizando os métodos de V-PCA e C-PCA. A situação-problema principal está relacionada com a detecção de ataques virtuais.

Outro trabalho relevante foi desenvolvido por Zhang e Wang (2022) que aplicou o método proposto na detecção de falhas em processos industriais. Como é mencionado no artigo, este é um processo que, por causa das falhas, irá apresentar valores intervalares, sendo necessário um método de PCA que capte essa situação. Em seu contexto, Zhang e Wang (2022), propõe remodelar a matriz de informações através do cálculo do valor esperado associado à Distribuição

Estatística que modela o processo industrial observado.

Ainda pensando no contexto de detecção de falhas, podemos mencionar o trabalho de Basha *et al.* (2018). Neste artigo são mencionadas três estratégias utilizadas em trabalhos anteriores, que os autores buscam aplicar em simulação para detecção de falhas e classificação. Um ponto chave deste trabalho está em afirmar que estas tentativas incorporam uma grande área que chamam de *Interval Principal Component Analysis* (IPCA). Como menciona Basha *et al.* (2018), as tentativas de generalizar a PCA já são um pouco mais eficientes que a PCA clássica, mas concordam que é necessário haver mais estudos e mais desenvolvimento nestas estratégias de IPCA.

Em IZEM *et al.* (2015) também é abordado o problema de detecção e isolamento de falhas, aplicando os métodos V-PCA e C-PCA, além do *Midpoint-Radii*, descrito em Palumbo e Lauro (2003). Este último busca encontrar a PCA para a matriz centro e matriz raio no contexto intervalar. Como veremos mais à frente, a ideia de utilizar matriz centro e matriz raio também faz parte do Método Deif, proposto por Deif (1991). O método Deif tem como objetivo encontrar autovalores de matrizes intervalares e é bem exemplificado no trabalho de Maiti e Chakraverty (2023). No próximo capítulo será mostrado que uma das estratégias para encontrar a PCA é utilizando os autovalores e autovetores da matriz de covariância das variáveis. Por isso, são relevantes as estratégias para encontrar autovalores intervalares, como é mostrado em Deif (1991) e Mizukoshi e Lodwick (2021).

Durante a revisão bibliográfica, dois métodos são bastante mencionados. O primeiro é o C-PCA, que considera apenas o valor central dos intervalos. Contudo, Giordani e Kiers (2006) pondera que, desta forma, a incerteza é descartada. O outro método é o V-PCA, proposto por Cazes *et al.* (1997), que considera projetar hiper-retângulo, sendo as extremidades dos intervalos os vértices. Para este método, Giordani e Kiers (2006) diz que esta estratégia aumenta a dimensão da matriz de observações e isso torna o problema mais custoso computacionalmente, além de, no final, não ser possível apresentar componentes principais intervalares.

3 ANÁLISE DE COMPONENTE PRINCIPAL

Neste capítulo iremos apresentar uma breve descrição da técnica PCA e como são encontradas as componentes principais na versão clássica, formalizar o contexto intervalar e apresentar a metodologia para o desenvolvimento da PCA em seu contexto intervalar.

3.1 Contexto clássico

Como a ideia principal da PCA é condensar o máximo de informações através da aplicação de uma transformação linear, então considere o vetor de variáveis aleatórias $\mathbf{x} \in \mathbb{R}^p$ e que a matriz de covariância Σ associada a este vetor seja conhecida. Queremos encontrar um vetor $\alpha_1 \in \mathbb{R}^p$ que, ao ser aplicado em \mathbf{x} , maximize a variância de \mathbf{x} :

$$\alpha_1^t \mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j \quad (3.1)$$

Para isso, precisamos encontrar o vetor α_1 que, ao ser aplicado na matriz de covariância, maximize a variância:

$$\begin{aligned} \max \quad & \frac{1}{2} \alpha_1^t \Sigma \alpha_1 \\ \text{s.a.:} \quad & \alpha_1^t \alpha_1 = 1 \end{aligned} \quad (3.2)$$

A restrição de normalização $\alpha_1^t \alpha_1 = 1$ garante que o vetor tenha tamanho limitado, evitando que o problema seja ilimitado. Para o problema (3.2), utilizando o método dos Multiplicadores de Lagrange, temos:

$$\max \quad \frac{1}{2} \alpha_1^t \Sigma \alpha_1 - \lambda (\alpha_1^t \alpha_1 - 1), \text{ com } \lambda \in \mathbb{R} \quad (3.3)$$

Derivando em relação a α_1 e igualando a 0, temos:

$$\frac{1}{2} \cdot 2 \Sigma \alpha_1 - \lambda \alpha_1 = 0 \Rightarrow (\Sigma - \lambda \cdot \mathbf{I}_p) \alpha_1 = 0, \quad (3.4)$$

sendo \mathbf{I}_p a matriz identidade de dimensão p . Dessa forma, λ é o autovalor associado à matriz de covariância e α_1 o autovetor associado. Ou seja, encontrar a PCA torna-se o problema de encontrar o autovalor e o autovetor associado ao autovalor de uma matriz.

A fim de ilustrarmos o procedimento, suponha o exemplo numérico clássico em que a matriz \mathbf{A} possui duas variáveis, representadas pelas colunas 1 e 2 da matriz \mathbf{A} , sendo cada linha uma observação realizada de um processo qualquer.

$$\mathbf{A} = \begin{bmatrix} 2 & 1,03 \\ 4 & 1,97 \\ 6 & 3,00 \\ 8 & 3,98 \\ 10 & 4,98 \end{bmatrix} \quad (3.5)$$

Associada a matriz \mathbf{A} , temos a matriz de covariância abaixo:

$$\Sigma = \begin{bmatrix} 10 & 4,9550 \\ 4,9550 & 2,4556 \end{bmatrix} \quad (3.6)$$

Para obtermos uma solução não nula de $(\Sigma - \lambda \cdot \mathbf{I}_p)\alpha_1 = \mathbf{0}$, é necessário que $\det(\Sigma - \lambda \cdot \mathbf{I}_p) = 0$. Temos:

$$\det(\Sigma - \lambda \cdot \mathbf{I}_2) = \begin{vmatrix} 10 - \lambda & (4,9550) \\ (4,9550) & (2,4556 - \lambda) \end{vmatrix} = \lambda^2 - 12,4556\lambda + 0,0040 = 0 \quad (3.7)$$

para

$$\lambda_1 = 0,0003 \quad \text{e} \quad \lambda_2 = 12,4553 \quad (3.8)$$

Optando pelo maior autovalor, por representar aproximadamente 100% da variabilidade do processo, utilizaremos λ_2 para encontrar o autovetor associado. Assim, temos:

$$\Sigma - \lambda_2 \cdot \mathbf{I}_2 = \begin{bmatrix} -2,4553 & 4,9550 \\ 4,9550 & -9,9997 \end{bmatrix} \Rightarrow \begin{bmatrix} -2,4553 & 4,9550 \\ 4,9550 & -9,9997 \end{bmatrix} \alpha_1 = \mathbf{0} \quad (3.9)$$

gerando um sistema de equações lineares. A solução do sistema (3.9) é dada por:

$$\alpha_1 = \begin{bmatrix} 2,0181 \\ 1 \end{bmatrix} \quad (3.10)$$

Aplicando o processo de normalização, temos:

$$\alpha_1 = \begin{bmatrix} 0,8960 \\ 0,4440 \end{bmatrix} \quad (3.11)$$

Assim, podemos realizar a transformação linear da equação (3.1) para encontrar as componentes principais associadas ao autovetor associado ao maior autovalor:

$$PCA = \begin{bmatrix} 2,2494 \\ 4,4588 \\ 6,7082 \\ 8,9353 \\ 11,1714 \end{bmatrix} \quad (3.12)$$

Contudo, sabemos que diversos problemas da vida real são representados através de informações intervalares. Palumbo e Lauro (2003) menciona, por exemplo, análise comportamental, análise e previsão de condições meteorológicas, controle estatístico de qualidade, análise de dados financeiros como áreas que podem se beneficiar da análise intervalar. Serrao *et al.* (2023) também menciona a emergente área de análise simbólica de dados e a importância do desenvolvimento da PCA intervalar. Além das diversas áreas mencionadas na revisão bibliográfica.

Diante disso, desenvolver técnicas para encontrar a PCA em seu contexto intervalar, preservando a incerteza das informações e não aumentando a dimensão da matriz, torna-se um objetivo a ser alcançado.

3.2 Contexto intervalar

O contexto intervalar para a análise de componente principal se dá quando alguma informação da matriz de dados é intervalar, a essa matriz chamaremos de matriz intervalar. Então, considere $[\mathbf{A}]$, com m linhas e n colunas, uma matriz qualquer que possui valores intervalares e é dada da seguinte forma:

$$[\mathbf{A}] = \begin{bmatrix} [a_{11}; \bar{a}_{11}] & \dots & [a_{1n}; \bar{a}_{1n}] \\ \vdots & \ddots & \vdots \\ [a_{m1}; \bar{a}_{m1}] & \dots & [a_{mn}; \bar{a}_{mn}] \end{bmatrix} \quad (3.13)$$

O nosso objetivo é encontrar a matriz de covariância Σ desta matriz. Para isso, podemos reescrever a estrutura da matriz $[A]$ usando combinação entre os limitantes inferior e superior dos intervalos através do tamanho do intervalo multiplicado por uma variável γ . A estratégia de intervalo de restrições (Mizukoshi; Lodwick, 2021) pode ser utilizada. Logo, temos:

$$\mathbf{A}(\gamma) = \begin{bmatrix} \underline{a}_{11} + \gamma_{11}d_{11} & \dots & \underline{a}_{11} + \gamma_{1n}d_{1n} \\ \vdots & \ddots & \vdots \\ \underline{a}_{m1} + \gamma_{m1}d_{m1} & \dots & \underline{a}_{m1} + \gamma_{mn}d_{mn} \end{bmatrix}, \text{ com } 0 \leq \gamma_j \leq 1 \text{ e } d_{ij} = |\bar{a}_{ij} - \underline{a}_{ij}|. \quad (3.14)$$

Do processo anterior, temos:

$$\underline{\mathbf{A}} = \begin{bmatrix} \underline{a}_{11} & \dots & \underline{a}_{1n} \\ \vdots & \ddots & \vdots \\ \underline{a}_{m1} & \dots & \underline{a}_{mn} \end{bmatrix} \quad (3.15)$$

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1n} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \dots & \gamma_{mn} \end{bmatrix} \quad (3.16)$$

$$\mathbf{\Delta} = \begin{bmatrix} d_{11} & \dots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{m1} & \dots & d_{mn} \end{bmatrix} \quad (3.17)$$

Então:

$$\mathbf{A}(\gamma) = \underline{\mathbf{A}} + \mathbf{\Gamma} \odot \mathbf{\Delta}, \text{ com } 0 \leq \gamma_j \leq 1, d_{ij} = |\bar{a}_{ij} - \underline{a}_{ij}| \text{ e } \odot \text{ o produto de Hadamard.} \quad (3.18)$$

Como,

$$\Sigma = cov(A_i, A_j) = E[(A_i - E[A_i])(A_j - E[A_j])^t], \forall i, j \quad (3.19)$$

Sendo A_j uma coluna j qualquer da matriz de dados e \underline{a}_j o limitante inferior de um elemento qualquer da coluna j , temos que:

$$E[A_j] = \mu(A_j) = \frac{1}{m} \sum_i^m (a_{ij} + \gamma_j d_{ij}) = \frac{1}{m} \sum_i^m a_{ij} + \frac{1}{m} \sum_i^m \gamma_j d_{ij}. \quad (3.20)$$

Portanto,

$$\mu(A_j) = \mu(\underline{a}_j) + \frac{1}{m} \sum_i^m \gamma_j d_{ij} \Rightarrow \mu(A) = M_{1n}, \quad (3.21)$$

o vetor linha com a média de cada coluna.

Com o intuito de manter a notação matricial para uma execução algorítmica e sabendo que $A_j - E[A_j]$ é para todo j , podemos compor uma matriz M_{mn} , sendo todas as linhas iguais a M_{1n} .

Temos a matriz $A(\gamma)_{mn}$ e a matriz média M_{mn} . Ao subtrair uma pela outra teremos uma matriz que representa a dispersão em torno da média:

$$A_c = A(\gamma) - M_{mn} \quad (3.22)$$

O termo geral da matriz A_c será dado por:

$$c_{ij} = a_{ij} + \gamma_j d_{ij} - \mu(\underline{a}_j) - \frac{1}{m} \sum_i^m \gamma_j d_{ij} \quad (3.23)$$

Agora, para determinar um elemento qualquer da matriz de covariância, precisamos:

$$E[\langle c_{.j}^t, c_{.w} \rangle], \quad (3.24)$$

sendo j e w colunas quaisquer.

Deste modo, temos que, para um elemento i qualquer deste vetor:

$$\langle c_{.j}^t, c_{.w} \rangle = \left(\underline{a}_{ij} + \gamma_j d_{ij} - \mu(\underline{a}_j) - \frac{1}{m} \sum_i^m \gamma_j d_{ij} \right) \cdot \left(\underline{a}_{iw} + \gamma_w d_{iw} - \mu(\underline{a}_w) - \frac{1}{m} \sum_i^m \gamma_w d_{iw} \right) \quad (3.25)$$

$$\begin{aligned}
\langle c_{.j}^t, c_{.w} \rangle &= \underline{a}_{ij} \underline{a}_{iw} + \underline{a}_{ij} \gamma_{iw} d_{iw} - \underline{a}_{ij} \mu(\underline{a}_w) - \frac{\underline{a}_{ij}}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&+ \gamma_{ij} d_{ij} \underline{a}_{iw} + \gamma_{ij} d_{ij} \gamma_{iw} d_{iw} - \gamma_{ij} d_{ij} \mu(\underline{a}_w) - \frac{\gamma_{ij} d_{ij}}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&- \mu(\underline{a}_j) \underline{a}_{iw} - \mu(\underline{a}_j) \gamma_{iw} d_{iw} + \mu(\underline{a}_j) \mu(\underline{a}_w) + \frac{\mu(\underline{a}_j)}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&- \frac{\underline{a}_{iw}}{m} \sum_i^m \gamma_{ij} d_{ij} - \frac{\gamma_{iw} d_{iw}}{m} \sum_i^m \gamma_{ij} d_{ij} + \frac{\mu(\underline{a}_w)}{m} \sum_i^m \gamma_{ij} d_{ij} + \frac{1}{m^2} \left(\sum_i^m \gamma_{ij} d_{ij} \right) \left(\sum_i^m \gamma_{iw} d_{iw} \right)
\end{aligned} \tag{3.26}$$

Para obtermos a média dos termos do produto interno é necessário somar todas as possibilidades de i e dividir por $(m-1)$ a fim de termos $E[\langle c_{.j}^t, c_{.w} \rangle]$. Note também que a única variável que existe para (3.26) é $\gamma_{ij} d_{ij}$ ou $\gamma_{iw} d_{iw}$, ou seja, γ modela todas as possibilidades. Desta forma, podemos encontrar uma matriz de covariância intervalar que modele todas as possibilidades da matriz original.

Para isso temos que maximizar e minimizar $\frac{1}{m-1} \sum_{i=1}^m \langle c_{.j}^t, c_{.w} \rangle$. Note que, neste caso, podemos desconsiderar as constantes. Abaixo, para um i qualquer, temos as constantes destacadas em negrito e depois as mesmas são removidas:

$$\begin{aligned}
\langle c_{.j}^t, c_{.w} \rangle &= \mathbf{\underline{a}_{ij}} \mathbf{\underline{a}_{iw}} + \underline{a}_{ij} \gamma_{iw} d_{iw} - \mathbf{\underline{a}_{ij}} \mu(\underline{\mathbf{a}}_w) - \frac{\underline{a}_{ij}}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&+ \gamma_{ij} d_{ij} \underline{a}_{iw} + \gamma_{ij} d_{ij} \gamma_{iw} d_{iw} - \gamma_{ij} d_{ij} \mu(\underline{a}_w) - \frac{\gamma_{ij} d_{ij}}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&- \mu(\mathbf{\underline{a}_j}) \underline{a}_{iw} - \mu(\underline{a}_j) \gamma_{iw} d_{iw} + \mu(\mathbf{\underline{a}_j}) \mu(\underline{\mathbf{a}}_w) + \frac{\mu(\underline{a}_j)}{m} \sum_i^m \gamma_{iw} d_{iw} \\
&- \frac{\underline{a}_{iw}}{m} \sum_i^m \gamma_{ij} d_{ij} - \frac{\gamma_{iw} d_{iw}}{m} \sum_i^m \gamma_{ij} d_{ij} + \frac{\mu(\underline{a}_w)}{m} \sum_i^m \gamma_{ij} d_{ij} + \frac{1}{m^2} \left(\sum_i^m \gamma_{ij} d_{ij} \right) \left(\sum_i^m \gamma_{iw} d_{iw} \right)
\end{aligned} \tag{3.27}$$

$$\begin{aligned}
\langle c_{.j}^t, c_{.w} \rangle &= \gamma_{iw} d_{iw} (\underline{a}_{ij} - \mu(\underline{a}_j)) - \sum_i^m \gamma_{iw} d_{iw} \left(\frac{\underline{a}_{ij} - \mu(\underline{a}_j)}{m} \right) \\
&+ \gamma_{ij} d_{ij} (\underline{a}_{iw} - \mu(\underline{a}_w)) - \sum_i^m \gamma_{ij} d_{ij} \left(\frac{\underline{a}_{iw} - \mu(\underline{a}_w)}{m} \right) \\
&+ \gamma_{ij} d_{ij} \gamma_{iw} d_{iw} - \frac{1}{m} (\gamma_{ij} d_{ij} \sum_i^m \gamma_{iw} d_{iw} + \gamma_{iw} d_{iw} \sum_i^m \gamma_{ij} d_{ij}) + \frac{1}{m^2} \left(\sum_i^m \gamma_{ij} d_{ij} \right) \left(\sum_i^m \gamma_{iw} d_{iw} \right)
\end{aligned} \tag{3.28}$$

A *priori*, podemos pensar que precisaremos resolver um problema de otimização para cada elemento da matriz de covariância, mas lembremo-nos em um intervalo fechado $[a, b]$, todo elemento x satisfaz $a \leq x \leq b$, ou seja, os γ que maximizam ou minimizam serão iguais. E, como estamos considerando j e w quaisquer, temos, em especial, o caso em que $j = w$, o elemento da diagonal principal da matriz de covariância. Assim, podemos resolver os problemas associados a esses elementos e utilizar o γ encontrado para construir toda a matriz de covariância intervalar.

Como $j = w$, para um i qualquer, temos:

$$\begin{aligned}
\langle c_{.j}^t, c_{.j} \rangle &= \gamma_j d_{ij} (\underline{a}_{ij} - \mu(\underline{a}_j)) - \sum_i^m \gamma_j d_{ij} \left(\frac{\underline{a}_{ij} - \mu(\underline{a}_j)}{m} \right) \\
&+ \gamma_j d_{ij} (\underline{a}_{ij} - \mu(\underline{a}_j)) - \sum_i^m \gamma_j d_{ij} \left(\frac{\underline{a}_{ij} - \mu(\underline{a}_j)}{m} \right) \\
&+ \gamma_j d_{ij} \gamma_j d_{ij} - \frac{1}{m} \left(\gamma_j d_{ij} \sum_i^m \gamma_j d_{ij} + \gamma_j d_{ij} \sum_i^m \gamma_j d_{ij} \right) + \frac{1}{m^2} \left(\sum_i^m \gamma_j d_{ij} \right) \left(\sum_i^m \gamma_j d_{ij} \right) \\
&= 2\gamma_j d_{ij} (\underline{a}_{ij} - \mu(\underline{a}_j)) - 2 \sum_i^m \gamma_j d_{ij} \left(\frac{\underline{a}_{ij} - \mu(\underline{a}_j)}{m} \right) \\
&+ (\gamma_j d_{ij})^2 - \frac{1}{m} \left(2\gamma_j d_{ij} \sum_i^m \gamma_j d_{ij} \right) + \frac{1}{m^2} \left(\sum_i^m \gamma_j d_{ij} \right)^2 \\
&= 2\gamma_j d_{ij} (\underline{a}_{ij} - \mu(\underline{a}_j)) - 2 \sum_i^m \gamma_j d_{ij} \left(\frac{\underline{a}_{ij} - \mu(\underline{a}_j) - \gamma_j d_{ij}}{m} \right) \\
&+ (\gamma_j d_{ij})^2 + \frac{1}{m^2} \left(\sum_i^m \gamma_j d_{ij} \right)^2 \\
&= c_{ij}^*
\end{aligned} \tag{3.29}$$

Para termos uma matriz de covariância intervalar, precisamos encontrar os valores de γ_j que maximizem e minimizem os elementos da diagonal principal. Dessa forma, para cada elemento da diagonal principal, teremos que resolver um problema de maximização e de minimização. Por exemplo:

$$\begin{aligned} \min \quad & c_{ij}^* \\ \text{s.a. : } & 0 \leq \gamma_j \leq 1 \end{aligned} \quad \text{e} \quad \begin{aligned} \max \quad & c_{ij}^* \\ \text{s.a. : } & 0 \leq \gamma_j \leq 1 \end{aligned} \quad (3.30)$$

Dessa forma, podemos construir a matriz de covariância intervalar:

$$\text{Cov}([\mathbf{A}]) = \begin{bmatrix} [\underline{\sigma}_{11}; \bar{\sigma}_{11}] & \dots & [\underline{\sigma}_{1n}; \bar{\sigma}_{1n}] \\ \vdots & \ddots & \vdots \\ [\underline{\sigma}_{n1}; \bar{\sigma}_{n1}] & \dots & [\underline{\sigma}_{nn}; \bar{\sigma}_{nn}] \end{bmatrix} \quad (3.31)$$

A fim de facilitar a compreensão, vamos aplicar a proposta apresentada da PCA clássica no aso em que uma das variáveis aleatórias tem entradas intervalares devido a uma perturbação nos dados. Dessa forma, temos:

$$[\mathbf{A}] = \begin{bmatrix} 2 & [0,95; 1,11] \\ 4 & [1,92; 2,01] \\ 6 & [2,94; 3,05] \\ 8 & [3,93; 4,03] \\ 10 & [4,92; 5,04] \end{bmatrix} \quad (3.32)$$

Remodelando a matriz de acordo com a estratégia de intervalo de restrições (MIZUKOSHI; LODWICK, 2021):

$$\mathbf{A} = \begin{bmatrix} 2 & 0,95 + 0,16\gamma_1 \\ 4 & 1,92 + 0,09\gamma_2 \\ 6 & 2,94 + 0,11\gamma_3 \\ 8 & 3,93 + 0,10\gamma_4 \\ 10 & 4,92 + 0,12\gamma_5 \end{bmatrix} \quad (3.33)$$

Calculando a média das colunas e repetindo as linhas para termos a matriz \mathbf{M} , temos:

$$\mathbf{M} = \begin{bmatrix} 6 & 2,93 + \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 6 & 2,93 + \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 6 & 2,93 + \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 6 & 2,93 + \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 6 & 2,93 + \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \end{bmatrix} \quad (3.34)$$

Assim, a matriz centro \mathbf{A}_c é:

$$\mathbf{A}_c = \begin{bmatrix} -4 & 0,16\gamma_1 - 1,98 - \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ -2 & 0,09\gamma_2 - 1,01 - \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 0 & 0,11\gamma_3 + 0,01 - \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 2 & 0,10\gamma_4 + 1,00 - \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \\ 4 & 0,12\gamma_5 + 1,99 - \frac{1}{5}(0,16\gamma_1 + 0,09\gamma_2 + 0,11\gamma_3 + 0,10\gamma_4 + 0,12\gamma_5) \end{bmatrix} \quad (3.35)$$

Como não há incerteza na primeira variável, o primeiro elemento da matriz de covariância não depende de γ_i . Sendo esse valor igual a 10. Contudo, para encontrar o outro elemento da diagonal principal, precisamos resolver o problema de otimização para o segundo elemento da diagonal principal. Fazendo:

$$\gamma_{min} = \begin{bmatrix} 1 \\ 1 \\ 0,5 \\ 0 \\ 0 \end{bmatrix} \quad e \quad \gamma_{max} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (3.36)$$

Temos o limite inferior do intervalo gerado por γ_{min} e o limite superior por γ_{max} . Assim, temos a seguinte matriz de covariância:

$$\Sigma = Cov([\mathbf{A}]) = \begin{bmatrix} 10 & [4,77;5,1450] \\ [4,77;5,1450] & [2,2758;2,6485] \end{bmatrix} \quad (3.37)$$

Para encontrar a primeira componente principal, precisamos encontrar o maior autovalor associado à matriz Σ , ou seja, precisamos encontrar autovalores intervalares e, para isso, iremos abordar três métodos: *Deif*, *CIA* e *Single-Level*.

3.2.1 Método Deif

Deif (1991) propõe um método que permite encontrar intervalos disjuntos de autovalores intervalares. Suponha, por exemplo, uma matriz intervalar 2x2:

$$[A] = \begin{bmatrix} [a, \bar{a}] & [b, \bar{b}] \\ [c, \bar{c}] & [d, \bar{d}] \end{bmatrix} \quad (3.38)$$

Podemos segmentar esta matriz em uma matriz com os valores centrais dos intervalos, A_c , e uma matriz raio, A_Δ , que representa as perturbações aplicadas em A_c :

$$A_c = \begin{bmatrix} \frac{a+\bar{a}}{2} & \frac{b+\bar{b}}{2} \\ \frac{c+\bar{c}}{2} & \frac{d+\bar{d}}{2} \end{bmatrix} \text{ e } A_\Delta = \begin{bmatrix} \frac{\bar{a}-a}{2} & \frac{\bar{b}-b}{2} \\ \frac{\bar{c}-c}{2} & \frac{\bar{d}-d}{2} \end{bmatrix} \quad (3.39)$$

Deif (1991) diz que, se $[A] = A_c \pm A_\Delta$ for uma matriz intervalar simétrica real e $S_i = \text{diag}(\text{sgn}(x_1^{(i)}), \dots, \text{sgn}(x_n^{(i)}))$, $i = 1, \dots, n$, retirado dos autovetores de A_c , for constante em $[A]$, então λ_i é o i -ésimo autovalor de A , $A \in [A]$ varia no intervalo:

$$[\lambda_i] = [\lambda_i(A_c - S_i A_\Delta S_i), \lambda_i(A_c + S_i A_\Delta S_i)], \quad i = 1, \dots, n \quad (3.40)$$

Abaixo, segue a decomposição da matriz de covariância intervalar do exemplo anterior:

$$\Sigma_c = \begin{bmatrix} \frac{10+10}{2} & \frac{4,77+5,1450}{2} \\ \frac{4,77+5,1450}{2} & \frac{2,2758+2,6485}{2} \end{bmatrix} = \begin{bmatrix} 10 & 4,9575 \\ 4,9575 & 2,4622 \end{bmatrix} \quad (3.41)$$

$$\Sigma_\Delta = \begin{bmatrix} \frac{10-10}{2} & \frac{5,1450-4,77}{2} \\ \frac{5,1450-4,77}{2} & \frac{2,6485-2,2758}{2} \end{bmatrix} = \begin{bmatrix} 0 & 0,1875 \\ 0,1875 & 0,1864 \end{bmatrix} \quad (3.42)$$

Após calcular as duas matrizes, precisamos calcular as matrizes S_i , a matriz de sinais dos autovetores de Σ_c , que nos auxiliará na escolha dos autovalores. Como nossa matriz Σ possui duas colunas, teremos duas matrizes S_j . A saber:

$$S_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \text{ e } S_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 \quad (3.43)$$

Os determinantes das matrizes S_i indicam o sentido da orientação considerada na análise realizada pelo método de *Deif*. Quando $\det(S_i) = 1$, preserva-se o sentido original da base, e o método seleciona os maiores autovalores da matriz $\Sigma_c \pm S_i \Sigma_\Delta S_i$. Por outro lado, quando $\det(S_i) = -1$, há uma inversão no sentido da base, e o método seleciona os menores autovalores. Para S_1 , por exemplo, $\det(S_1) = -1$, o método escolherá os menores autovalores da matriz $\Sigma_c \pm S_1 \Sigma_\Delta S_1$.

$$\Sigma_c \pm \mathbf{S}_1 \Sigma_\Delta \mathbf{S}_1 \rightarrow \begin{bmatrix} 10 & 5,145 \\ 5,145 & 2,2758 \end{bmatrix} \text{ e } \begin{bmatrix} 10 & 4,77 \\ 4,77 & 2,6485 \end{bmatrix} \quad (3.44)$$

Os autovalores da primeira matriz são $-0,2953$ e $12,5712$; e os autovalores da segunda matriz são $0,3023$ e $12,3462$. Como $\det(\mathbf{S}_1) = -1$, então escolheremos os menores autovalores, ou seja, $\lambda_1 = [-0,2953; 0,3023]$. Contudo, estamos falando de uma matriz de covariância, logo os valores inferiores a zero não nos interessam, deste modo temos: $\lambda_1 = [0; 0,3023]$.

Para \mathbf{S}_2 , temos:

$$\Sigma_c \pm \mathbf{S}_2 \Sigma_\Delta \mathbf{S}_2 \rightarrow \begin{bmatrix} 10 & 4,77 \\ 4,77 & 2,2758 \end{bmatrix} \text{ e } \begin{bmatrix} 10 & 5,1450 \\ 5,1450 & 2,6485 \end{bmatrix} \quad (3.45)$$

Os autovalores da primeira matriz são $0,0004$ e $12,2754$; e os autovalores da segunda matriz são $0,0011$ e $12,6474$. Como $\det(\mathbf{S}_2) = 1$, então, conforme dito anteriormente, escolheremos os maiores autovalores, ou seja, $\lambda_2 = [12,2754; 12,6474]$. Em resumo, temos:

$$\lambda_1 = [0; 0,3023] \text{ e } \lambda_2 = [12,2754; 12,6474] \quad (3.46)$$

Perceba que, apesar do método apresentar os autovalores, é necessário construir o número de colunas mais duas matrizes, dessa forma, aumentando a complexidade operacional. Outro ponto importante é a atenção na matriz de sinais, pois a escolha errada irá apresentar outro conjunto de autovalores. Por isso, o interessante é a poder aplicar o método PCA para a matriz intervalar de modo os métodos numéricos já conhecidos possam ser aplicados. Essa ideia será apresentado nas sessões seguintes.

3.2.2 Método Deif com Leverrier-Faddeev

Uma proposta de aprimorar o Método *Deif* já apresentado é combiná-lo com algum método numérico, de modo que seja possível, através do método numérico, encontrar o autovalor através do polinômio característico da matriz raio e da matriz centro.

Uma forma de encontrar o polinômio característico é o método numérico de *Leverrier-Faddeev*. O método é utilizado para calcular os coeficientes do polinômio característico de matrizes

quadradas (Franco, 2006). Considere o polinômio característico de uma matriz quadrada \mathbf{A} qualquer. Este polinômio pode ser definido como:

$$P(\lambda) = (-1)^n [\lambda^n - c_1 \lambda^{n-1} - c_2 \lambda^{n-2} - \dots - c_{n-1} \lambda - c_n],$$

em que os coeficientes c_k são os valores que o método calcula.

O método de *Leverrier-Faddeev* diz:

1. Início:

$$\mathbf{A}_1 = \mathbf{A}, c_1 = tr(\mathbf{A}_1), \mathbf{B}_1 = \mathbf{A}_1 - c_1 \mathbf{I}, \text{ sendo } tr(\mathbf{A}) \text{ o traço da matriz.}$$

2. Para $k = 2, \dots, n$:

– Calcule:

$$\mathbf{A}_k = \mathbf{A} \mathbf{B}_{k-1}, c_k = \frac{tr(\mathbf{A}_k)}{k}, \mathbf{B}_k = \mathbf{A}_k - c_k \mathbf{I}$$

3. Polinômio característico:

Substitua os coeficientes c_k no polinômio característico:

$$P(\lambda) = (-1)^n [\lambda^n - c_1 \lambda^{n-1} - c_2 \lambda^{n-2} - \dots - c_{n-1} \lambda - c_n]$$

Para a matriz centro, apresentada anteriormente, temos:

$$\Sigma_c = \mathbf{A}_1 = \begin{bmatrix} 10 & 4,9575 \\ 4,9575 & 2,4621 \end{bmatrix}, c_1 = tr(\mathbf{A}_1) = 12,4622 \quad (3.47)$$

$$\mathbf{B}_1 = \mathbf{A}_1 - c_1 \mathbf{I} = \begin{bmatrix} -2,4622 & 4,9575 \\ 4,9575 & -10 \end{bmatrix} \quad (3.48)$$

$$\mathbf{A}_2 = \mathbf{A} \mathbf{B}_1 = \begin{bmatrix} -0,0447 & 0 \\ 0 & -0,0447 \end{bmatrix}, c_1 = \frac{tr(\mathbf{A}_2)}{2} = -0,0447 \quad (3.49)$$

Desse modo, o polinômio característico é:

$$P(\lambda_c) = \lambda_c^2 - (12,4622)\lambda_c - (-0,0447) \quad (3.50)$$

$$\lambda_{c_1} = 0,0036 \text{ e } \lambda_{c_2} = 12,4586 \quad (3.51)$$

Para a matriz raio, apresentada anteriormente, temos:

$$\Sigma_{\Delta} = \mathbf{A}_1 = \begin{bmatrix} 0 & 0,1875 \\ 0,1875 & 0,1864 \end{bmatrix}, c_1 = tr(\mathbf{A}_1) = 0,1864 \quad (3.52)$$

$$\mathbf{B}_1 = \mathbf{A}_1 - c_1 \mathbf{I} = \begin{bmatrix} -0,1864 & 0,1875 \\ 0,1875 & 0 \end{bmatrix} \quad (3.53)$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1 = \begin{bmatrix} 0,0352 & 0 \\ 0 & 0,0352 \end{bmatrix}, c_1 = \frac{tr(\mathbf{A}_2)}{2} = 0,0352 \quad (3.54)$$

Desse modo, o polinômio característico é:

$$P(\lambda_{\Delta}) = \lambda_{\Delta}^2 - (0,1864)\lambda_{\Delta} - (0,0352) \quad (3.55)$$

$$\lambda_{\Delta_1} = -0,1163 \text{ e } \lambda_{\Delta_2} = 0,3027 \quad (3.56)$$

Logo, temos que:

$$\begin{aligned} \lambda_1 &= \lambda_{c_1} \pm \lambda_{\Delta_1} \text{ e } \lambda_2 = \lambda_{c_2} \pm \lambda_{\Delta_2} \\ \lambda_1 &= 0,0036 \pm 0,1163 \text{ e } \lambda_2 = 12,4586 \pm 0,3027 \\ \lambda_1 &= [-0,1127; 0,1199] \text{ e } \lambda_2 = [12,1559; 12,7613] \end{aligned} \quad (3.57)$$

Como nos interessa apenas os autovalores não-negativos, então:

$$\lambda_1 = [0; 0,1199] \text{ e } \lambda_2 = [12,1559; 12,7613] \quad (3.58)$$

Que não possui o mesmo resultado que o Método *Deif* clássico, mas que já não há a problemática da matriz de sinais, uma vez que a direção já está contida no processo. Como o método de *Leverrier-Faddeev* foi utilizado acima, então será usado, também, nos métodos a seguir para garantir algum critério de comparação.

3.2.3 Método *Single-level* (SL)

Para aplicar o Método *Single-level* (SL) proposto por Mizukoshi e Lodwick (2021) precisamos remodelar a matriz Σ . Esse processo é igual ao processo de remodelagem que fizemos anteriormente na matriz intervalar $[A]$ para encontrar a matriz de covariância. A diferença está na variável γ que é única para toda a matriz. Assim, temos:

$$\Sigma(\gamma) = \begin{bmatrix} 10 & 4,77 + 0,375\gamma \\ 4,77 + 0,375\gamma & 2,2758 + 0,3727\gamma \end{bmatrix} \text{ com } 0 \leq \gamma \leq 1. \quad (3.59)$$

Desta forma, podemos encontrar os autovalores através do polinômio característico utilizando o método numérico de *Leverrier-Faddeev*:

$$\Sigma(\gamma) = A_1 = \begin{bmatrix} 10 & 4,77 + 0,375\gamma \\ 4,77 + 0,375\gamma & 2,2758 + 0,3727\gamma \end{bmatrix}, c_1 = tr(A_1) = 12,2758 + 0,3727\gamma \quad (3.60)$$

$$B_1 = A_1 - c_1 I = \begin{bmatrix} -2,2758 - 0,3727\gamma & 4,77 + 0,375\gamma \\ 4,77 + 0,375\gamma & -10 \end{bmatrix} \quad (3.61)$$

$$A_2 = AB_1 = \begin{bmatrix} 10(-2,2758 - 0,3727\gamma) + (4,77 + 0,375\gamma)^2 = (*) & 0 \\ 0 & (*) \end{bmatrix} \quad (3.62)$$

$$A_2 = \begin{bmatrix} -0,0051 - 1,9383\gamma + 0,1406\gamma^2 & 0 \\ 0 & -0,0051 - 1,9383\gamma + 0,1406\gamma^2 \end{bmatrix} \quad (3.63)$$

$$c_2 = \frac{tr(A_2)}{2} = -0,0051 - 0,1495\gamma + 0,1406\gamma^2 \quad (3.64)$$

Desse modo, o polinômio característico é:

$$P(\lambda) = \lambda^2 - c_1\lambda - c_2 \quad (3.65)$$

$$P(\lambda) = \lambda^2 - (12,2758 + 0,3727\gamma)\lambda - (-0,0051 - 0,1495\gamma + 0,1406\gamma^2) \quad (3.66)$$

Resolvendo o polinômio, temos:

$$\lambda = 0,00005(3727\gamma + 122758 \pm \sqrt{70130529\gamma^2 + 855238132\gamma + 15067486564}) \quad (3.67)$$

Dessa forma temos um problema min/max para encontrar os dois autovalores intervalares relacionados à equação anterior e podemos já incluir como restrição de que a solução precisará ser superior a zero, por ser autovalor oriundo de uma matriz de covariância. A saber:

$$\lambda_1 = [0,0004; 0,0036], \text{ com } \gamma_{min} = 0 \text{ e } \gamma_{max} = 0,52688 \quad (3.68)$$

$$\lambda_2 = [12,2754; 12,6474], \text{ com } \gamma_{min} = 0 \text{ e } \gamma_{max} = 1 \quad (3.69)$$

3.2.4 Método *Constraint Interval Arithmetic* (CIA)

Para aplicar o Método *Constraint Interval Arithmetic* (CIA) proposto por Mizukoshi e Lodwick (2021) precisamos remodelar a matriz Σ similar à etapa anterior, mas cada intervalo possui um parâmetro γ associado. Assim, temos:

$$\Sigma(\gamma) = \begin{bmatrix} 10 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & 2,2758 + 0,3727\gamma_2 \end{bmatrix} \text{ com } 0 \leq \gamma \leq 1 \quad (3.70)$$

Aplicando o método de *Leverrier-Faddeev* para a aproximação via CIA:

$$\Sigma(\gamma) = A_1 = \begin{bmatrix} 10 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & 2,2758 + 0,3727\gamma_2 \end{bmatrix}, c_1 = tr(A_1) = 12,2758 + 0,3727\gamma_2 \quad (3.71)$$

$$B_1 = A_1 - c_1 I = \begin{bmatrix} -2,2758 - 0,3727\gamma_2 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & -10 \end{bmatrix} \quad (3.72)$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1 = \begin{bmatrix} 10(-2,2758 - 0,3727\gamma_2) + (4,77 + 0,375\gamma_1)^2 = (*) & 0 \\ 0 & (*) \end{bmatrix} \quad (3.73)$$

$$\mathbf{A}_2 = \begin{bmatrix} -0,0051 + 3,5775\gamma_1 - 3,727\gamma_2 + 0,1406\gamma_1^2 & 0 \\ 0 & -0,0051 + 3,5775\gamma_1 - 3,727\gamma_2 + 0,1406\gamma_1^2 \end{bmatrix} \quad (3.74)$$

$$c_2 = \frac{tr(\mathbf{A}_2)}{2} = -0,0051 + 3,5775\gamma_1 - 3,727\gamma_2 + 0,1406\gamma_1^2 \quad (3.75)$$

Desse modo, o polinômio característico é:

$$P(\lambda) = \lambda^2 - c_1\lambda - c_2 \quad (3.76)$$

$$P(\lambda) = \lambda^2 - (12,2758 + 0,3727\gamma_2)\lambda - (-0,0051 + 3,5775\gamma_1 - 3,727\gamma_2 + 0,1406\gamma_1^2) \quad (3.77)$$

Logo, as raízes do polinômio são dadas por:

$$\lambda = 0,00005(3727\gamma_2 + 122758 \pm \sqrt{56240000\gamma_1^2 + 1431000000\gamma_1 + 13890529\gamma_2^2 - 575761868\gamma_2 + 15067486564}) \quad (3.78)$$

Dessa forma temos um problema min/max para encontrar os dois autovalores intervalares relacionados à equação (3.78) e considerando a restrição de que a solução deve ser maior que zero, por ser autovalor de uma matriz de covariância. A saber:

$$\lambda_1 = [0; 0,3023], \text{ com os pares de soluções } \gamma_{min} = (1; 0,9962) \text{ e } \gamma_{max} = (0; 1) \quad (3.79)$$

$$\lambda_2 = [12,2754; 12,6474], \text{ com os pares de soluções } \gamma_{min} = (0; 0) \text{ e } \gamma_{max} = (1; 1) \quad (3.80)$$

3.2.5 Comparação entre métodos

Para comparar os três métodos, é interessante termos alguma informação sobre os possíveis autovalores da matriz intervalar. Para isto, foi realizada uma simulação das possíveis matrizes. Ao todo, foram geradas 442 matrizes com a simulação abaixo:

```
1 import numpy as np
2
3 matriz = np.array([[10, 4.77],
4                   [4.77, 2.2758]])
5
6 valor_inicial = matriz[1, 1]
7 autovalores = np.linalg.eigvals(matriz)
8
9 lambda1_min = autovalores[1]
10 lambda1_max = autovalores[1]
11
12 lambda2_min = autovalores[0]
13 lambda2_max = autovalores[0]
14
15 i = 0
16 while matriz[0, 1] <= 5.145:
17     matriz[1, 1] = valor_inicial
18     while matriz[1, 1] <= 2.6485 + 0.0186350:
19         i += 1
20         print("matriz: ", i)
21
22         autovalores = np.linalg.eigvals(matriz)
23
24         print(matriz)
25         matriz[1, 1] += 0.0186350
26
27         if autovalores[1] < lambda1_min:
28             lambda1_min = autovalores[1]
29
```

```

30     if autovalores[1] > lambda1_max:
31         lambda1_max = autovalores[1]
32
33     if autovalores[0] < lambda2_min:
34         lambda2_min = autovalores[0]
35
36     if autovalores[0] > lambda2_max:
37         lambda2_max = autovalores[0]
38
39     matriz[0, 1] += 0.01875
40     matriz[1, 0] += 0.01875

```

Considerando os menores e os maiores autovalores encontrados, temos:

$$\lambda_1 = [0; 0,3023] \text{ e } \lambda_2 = [12,2754; 12,6474] \quad (3.81)$$

Assim, podemos utilizar estes intervalos como referência para encontrar a taxa de cobertura e assertividade. A taxa de cobertura pode ser encontrada da seguinte forma: suponha que temos um intervalo de referência [a;b] e um intervalo a ser comparado [c;d], a assertividade é dada:

$$\text{Assertividade} = \text{Cobertura Relativa} - \text{Excedente} - \text{Falta} \quad (3.82)$$

$$\text{Excedente} = \frac{\text{Comprimento da intersecção entre intervalos}}{\text{Comprimento do intervalo referência}} \quad (3.83)$$

$$\text{Cobertura Relativa} = \frac{\text{Comprimento do excesso do limite superior avaliado}}{\text{Comprimento do intervalo referência}} \quad (3.84)$$

$$\text{Falta} = \frac{\text{Comprimento da falta do limite inferior avaliado}}{\text{Comprimento do intervalo referência}} \quad (3.85)$$

$$\text{Assertividade} = \frac{\max(0; \min(b; d) - \max(a; c))}{b - a} - \max\left(0; \frac{d - b}{b - a}\right) - \max\left(0, \frac{a - c}{b - a}\right) \quad (3.86)$$

Assim, temos a assertividade verificada na Tabela 5:

Tabela 5 – Assertividade de λ_1 e λ_2

Método	Autovalor	Assertividade
Simulação	[0,0000; 0,3023]	–
Deif	[0,0000; 0,3023]	100,00%
Deif com Leverrier-Faddeev	[0,0000; 0,1199]	39,66%
Single-Level	[0,0004; 0,0036]	1,06%
CIA	[0,0000; 0,3023]	100,00%
Simulação	[12,2754; 12,6474]	–
Deif	[12,2754; 12,6474]	100,00%
Deif com Leverrier-Faddeev	[12,1559; 12,7613]	37,26%
Single-Level	[12,2754; 12,6474]	100,00%
CIA	[12,2754; 12,6474]	100,00%

Fonte: próprio autor.

Os métodos apresentaram valores muito próximos do simulado. Apenas o Método SL, para o primeiro autovalor, teve um resultado com pouca assertividade. Apesar dos Métodos *Deif* e CIA terem apresentado os mesmos resultados, o primeiro depende da análise da matriz de sinais referente aos autovetores da matriz centro e essa necessidade exige maior atenção durante sua aplicação, além de gerar mais matrizes. O método CIA, no entanto, associa a matriz intervalar a matrizes clássicas, permitindo utilizar resoluções convencionais ou métodos numéricos conhecidos.

3.2.6 Escolha do Autovalor

Caso não estejamos interessados em considerar todos os autovalores para a definição das componentes principais, será necessário avaliar quais autovalores são mais significativos. Para isso, é importante considerar o comprimento do autovalor e a magnitude de seus valores.

Por exemplo, considere dois intervalos $\alpha = [a; b]$ e $\beta = [c; d]$. Devemos considerar tanto o comprimento do intervalo, quanto o valor do intervalo. Assim, considere $d_\alpha = b - a$ e $d_\beta = d - c$ os comprimentos intervalares e $m_\alpha = \frac{b+a}{2}$ e $m_\beta = \frac{d+c}{2}$ os pontos médios.

Logo, a significância dos intervalos pode ser definida como:

$$S_\alpha = d_\alpha \cdot m_\alpha \text{ e } S_\beta = d_\beta \cdot m_\beta \quad (3.87)$$

Para exemplificar, considere os intervalos $\lambda_1 = [0; 0,3023]$ e $\lambda_2 = [12,2754; 12,6474]$. Temos, $d_{\lambda_1} = 0,3023$ e $m_{\lambda_1} = 0,1511$ para o primeiro intervalo e $d_{\lambda_2} = 0,372$ e $m_{\lambda_2} = 12,4614$ para o segundo intervalo. Assim,

$$S_{\lambda_1} = d_{\lambda_1} \cdot m_{\lambda_1} \text{ e } S_{\lambda_2} = d_{\lambda_2} \cdot m_{\lambda_2} \quad (3.88)$$

$$S_{\lambda_1} = 0,3023 \cdot 0,1511 = 0,0458 \text{ e } S_{\lambda_2} = 0,372 \cdot 12,4614 = 4,6356 \quad (3.89)$$

Sendo a significância relativa dada:

$$\Delta S_{\lambda_1} = \frac{S_{\lambda_1}}{S_{\lambda_1} + S_{\lambda_2}} \text{ e } \Delta S_{\lambda_2} = \frac{S_{\lambda_2}}{S_{\lambda_1} + S_{\lambda_2}} \quad (3.90)$$

$$\Delta S_{\lambda_1} = \frac{0,0458}{4,6814} = 0,9783\% \text{ e } \Delta S_{\lambda_2} = \frac{4,6356}{4,6814} = 99,0217\% \quad (3.91)$$

Dessa forma, podemos dizer que, considerando a magnitude do intervalo e o comprimento, o autovalor λ_2 é mais significativo. Nos resta agora determinar o autovetor, pois este será a componente principal.

4 COMPONENTE PRINCIPAL INTERVALAR

Para avançar na definição da análise das componentes principais, é necessário definir os autovetores, pois eles formam as componentes principais. Neste capítulo, iremos apresentar como identificar um autovetor intervalar associado ao autovalor intervalar. De forma prática, vamos considerar o autovalor λ_2 , por ter apresentado maior significância, segundo o critério adotado, para apresentar a modificação realizada no método clássico de *Leverrier-Faddeev*, descrito em Franco (2006).

4.1 Método Numérico

O método de *Leverrier-Faddeev* apresentado no capítulo anterior também permite identificar o autovetor através das etapas do algoritmo. Para identificar o autovetor através do método:

1. Faça:

$$u_0 = e, \text{ sendo } e \text{ uma coluna da matriz identidade.}$$

2. Calcule:

$$u_i = \lambda_k u_{i-1} + b_i, i = 1, 2, \dots, n-1$$

3. Considere b_i a coluna da matriz B_i correspondente a coluna escolhida da matriz identidade.

4. $u = u_{n-1}$ é o autovetor correspondente ao autovalor λ_k . Caso, u_{n-1} seja o vetor nulo, será necessário escolher outra coluna da matriz identidade e repetir o processo.

Vamos aplicar o algoritmo nos resultados do capítulo anterior.

4.1.1 SL - Autovetor

Para o resultado do *Single-level* considere:

$$u_0 = e = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.1)$$

$$u_1 = \lambda_2 u_0 + b_1 = (12,2754 + 0,372\gamma) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2,2758 - 0,3727\gamma \\ 4,77 + 0,375\gamma \end{bmatrix} \quad (4.2)$$

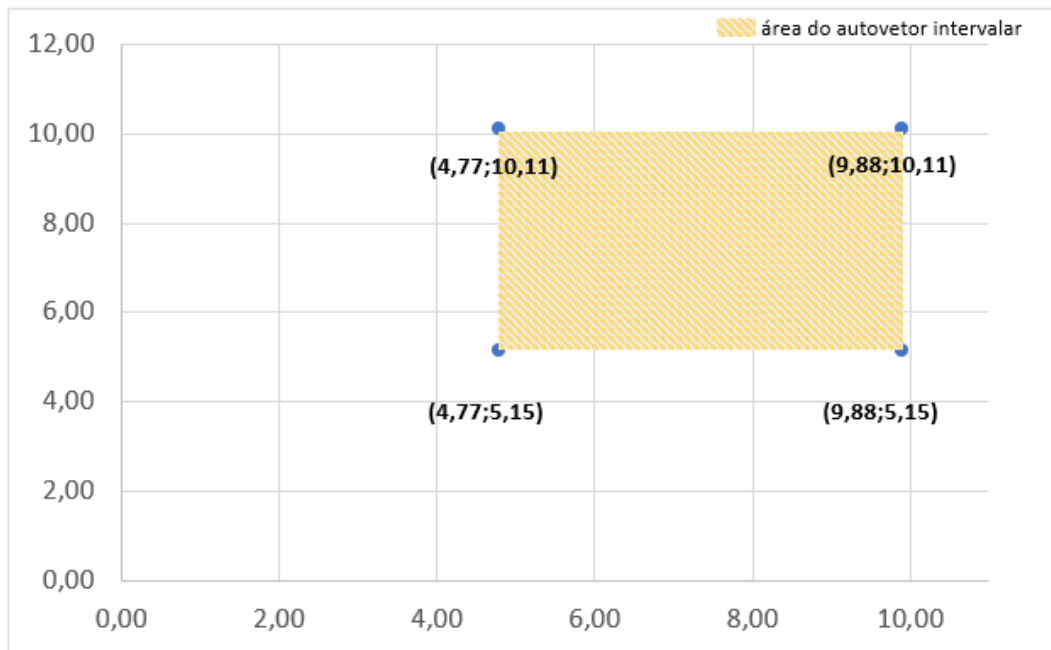
$$u_1 = \begin{bmatrix} 9,9996 - 0,0007\gamma \\ 4,77 + 0,375\gamma \end{bmatrix} \quad (4.3)$$

Para obter a solução na forma intervalar, basta encontrar os valores de γ que dão os limites inferiores e superiores. Os limites serão obtidos para quando $\gamma=0$ e $\gamma=1$. Logo, temos o autovetor:

$$[v_{sl}] = \begin{bmatrix} [9,9989; 9,9996] \\ [4,77; 5,145] \end{bmatrix} \quad (4.4)$$

Compreender o autovetor intervalar em seu contexto geométrico é um ponto importante para podermos realizar outros procedimentos como, por exemplo, a normalização. Um autovetor qualquer $[a, b]$ representa a direção da origem até o ponto (a, b) . Logo, se as componentes do autovetor são intervalos, então o autovetor intervalar representará a direção da origem a todos os pontos possíveis do intervalo. Ou seja, para $[v_{sl}]$ ilustrado na Figura 10, temos:

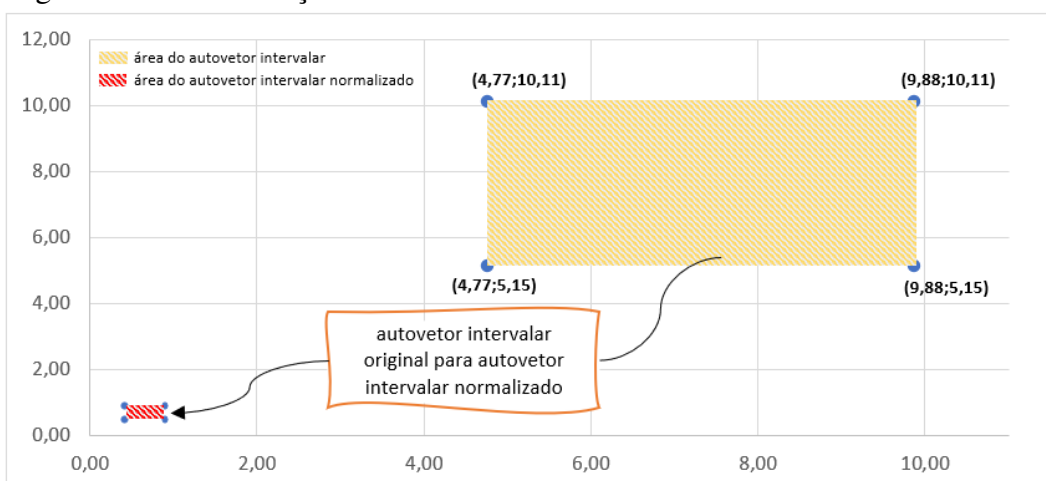
Figura 10 – Autovetor Intervalar - Método SL



Fonte: Próprio autor.

Para termos toda esta área normalizada, então os pontos precisam ter, no máximo, norma igual a um. Graficamente, o que queremos é apresentado na Figura 11:

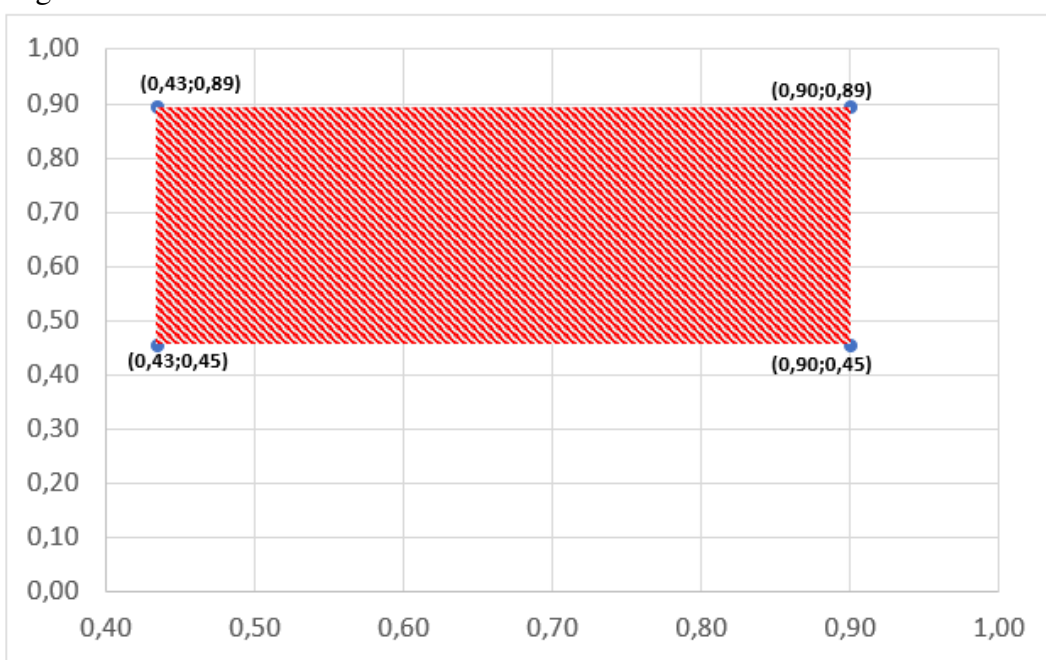
Figura 11 – Normalização do autovetor intervalar



Fonte: Próprio autor.

Expandindo a área do autovetor intervalar normalizado na Figura 12:

Figura 12 – Autovetor Intervalar Normalizado - Método SL



Fonte: Próprio autor.

Ou seja, podemos normalizar da seguinte forma:

$$[v_{sl}] = \begin{bmatrix} \left[\frac{9,9989}{\sqrt{9,9989^2 + 4,77^2}}; \frac{9,9996}{\sqrt{9,9996^2 + 5,145^2}} \right] \\ \left[\frac{4,77}{\sqrt{9,9989^2 + 4,77^2}}; \frac{5,145}{\sqrt{9,9996^2 + 5,145^2}} \right] \end{bmatrix} = \begin{bmatrix} [0,9026; 0,8892] \\ [0,4306; 0,4575] \end{bmatrix} \quad (4.5)$$

Que, geometricamente, é a área apresentada no gráfico. Logo, não há perda de informação ajustarmos as extremidades da primeira componente para termos um intervalo convencional. Assim:

$$[v_{sl}] = \begin{bmatrix} [0, 8892; 0, 9026] \\ [0, 4306; 0, 4575] \end{bmatrix} \quad (4.6)$$

4.1.2 CIA - Autovetor

Para o resultado do *constraint interval arithmetic* considere:

$$u_0 = e = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.7)$$

$$u_1 = \lambda_2 u_0 + b_1 = (12, 2754 + 0, 372 \gamma_3) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2, 2758 - 0, 3727 \gamma_2 \\ 4, 77 + 0, 375 \gamma_1 \end{bmatrix} \quad (4.8)$$

$$u_1 = \begin{bmatrix} 9, 9996 + 0, 372 \gamma_3 - 0, 3727 \gamma_2 \\ 4, 77 + 0, 375 \gamma_1 \end{bmatrix} \quad (4.9)$$

Para obter na forma intervalar, basta encontrar os valores de γ_i que dão os limites inferiores e superiores. Os limites inferiores serão dados quando $\gamma_1 = 0$, $\gamma_2 = 1$ e $\gamma_3 = 0$; limites superiores serão dados quando $\gamma_1 = 1$, $\gamma_2 = 0$ e $\gamma_3 = 1$. Logo, temos o autovetor:

$$[v_{cia}] = \begin{bmatrix} [9, 6269; 10, 3716] \\ [4, 77; 5, 145] \end{bmatrix} \quad (4.10)$$

Normalizando, temos:

$$[v_{cia}] = \begin{bmatrix} \left[\frac{9, 6269}{\sqrt{9, 6269^2 + 4, 77^2}}; \frac{10, 3716}{\sqrt{10, 3716^2 + 5, 145^2}} \right] \\ \left[\frac{4, 77}{\sqrt{9, 6269^2 + 4, 77^2}}; \frac{5, 145}{\sqrt{10, 3716^2 + 5, 145^2}} \right] \end{bmatrix} = \begin{bmatrix} [0, 8960; 0, 8958] \\ [0, 4440; 0, 4444] \end{bmatrix} \quad (4.11)$$

Conforme apresentado anteriormente, estes intervalos representam uma área no plano, então, por mais que o $[0, 8960; 0, 8958]$ não seja um intervalo convencional, irá formar a mesma área no plano que o intervalo $[0, 8958; 0, 8960]$, logo, sem perda de informação, podemos ajustar a primeira linha do autovetor associado ao autovalor e assim termos:

$$[v_{cia}] = \begin{bmatrix} [0, 8958; 0, 8960] \\ [0, 4440; 0, 4444] \end{bmatrix} \quad (4.12)$$

4.1.3 Deif com Leverrier-Faddeev - Autovetor

Nesta proposta, aplicamos anteriormente o método numérico de *Leverrier-Faddeev* tanto na matriz centro, quanto na matriz raio. Desta forma, precisaremos verificar ambos os processos, mas avaliando apenas a informação relacionada com o maior autovalor.

Para a matriz centro, temos:

$$u_{0_c} = e = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.13)$$

$$u_{1_c} = \lambda_{c_2} u_0 + b_1 = 12,4586 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2,4622 \\ 4,9575 \end{bmatrix} = \begin{bmatrix} 9,9964 \\ 4,9575 \end{bmatrix} \quad (4.14)$$

Para a matriz raio, temos:

$$u_{0_\Delta} = e = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.15)$$

$$u_{1_\Delta} = \lambda_{\Delta_2} u_0 + b_1 = 0,3027 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -0,1864 \\ 0,1875 \end{bmatrix} = \begin{bmatrix} 0,1163 \\ 0,1875 \end{bmatrix} \quad (4.16)$$

Logo,

$$[v_{deif-lv}] = u_{1_c} \pm u_{1_\Delta} = \begin{bmatrix} 9,9964 \\ 4,9575 \end{bmatrix} \pm \begin{bmatrix} 0,1163 \\ 0,1875 \end{bmatrix} = \begin{bmatrix} [9,8801; 10,1127] \\ [4,77; 5,145] \end{bmatrix} \quad (4.17)$$

Normalizando, temos:

$$[v_{deif-lv}] = \begin{bmatrix} \left[\frac{9,8801}{\sqrt{9,8801^2 + 4,77^2}}; \frac{10,1127}{\sqrt{10,1127^2 + 5,145^2}} \right] \\ \left[\frac{4,77}{\sqrt{9,8801^2 + 4,77^2}}; \frac{5,145}{\sqrt{10,1127^2 + 5,145^2}} \right] \end{bmatrix} = \begin{bmatrix} [0,9005; 0,8913] \\ [0,4348; 0,4535] \end{bmatrix} \quad (4.18)$$

Novamente, estes intervalos representam uma área no plano, então, por mais que o $[0,9005; 0,8913]$ não seja um intervalo convencional, irá formar a mesma área no plano que o intervalo $[0,8913; 0,9005]$, logo, sem perda de informação, podemos ajustar a primeira linha do autovetor associado ao autovalor e assim termos:

$$[v_{deif-lv}] = \begin{bmatrix} [0,8913;0,9005] \\ [0,4348;0,4535] \end{bmatrix} \quad (4.19)$$

4.2 Método Analítico

A proposta para o método *Deif* clássico não possui as matrizes B_i para identificar o autovetor associado ao autovalor através do método numérico de *Leverrier-Faddeev*. Por isso, apresentamos a seguir uma proposta de identificar analiticamente o autovetor associado ao autovalor utilizando, também, a estratégia do método CIA.

Para encontrar o autovetor, queremos um vetor $v \neq \vec{0}$, tal que $(A - \lambda I)v = 0$. Considerando a matriz intervalar Σ e o autovalor $[\lambda_2]$, temos:

$$\Sigma = \begin{bmatrix} 10 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & 2,2758 + 0,3727\gamma_2 \end{bmatrix}, \text{ com } 0 \leq \gamma_1 \leq 1 \quad (4.20)$$

$$[\lambda_2] = [12,2754;12,6474] = 12,2754 + 0,372\gamma_3, \text{ com } 0 \leq \gamma_3 \leq 1 \quad (4.21)$$

$$\Sigma - [\lambda_2]I = \begin{bmatrix} 10 - 12,2754 - 0,372\gamma_3 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & 2,2758 + 0,3727\gamma_2 - 12,2754 - 0,372\gamma_3 \end{bmatrix} \quad (4.22)$$

$$\Sigma - [\lambda_2]I = \begin{bmatrix} -2,2754 - 0,372\gamma_3 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & -9,9996 + 0,3727\gamma_2 - 0,372\gamma_3 \end{bmatrix} \quad (4.23)$$

Por definição, a matriz $\Sigma - [\lambda_2]I$ é singular, logo suas linhas são linearmente dependentes e isso implica que a solução do sistema $(\Sigma - [\lambda_2]I)[v] = \vec{0}$ é:

$$(-2,2754 - 0,372\gamma_3)[v_1] + (4,77 + 0,375\gamma_1)[v_2] = 0 \quad (4.24)$$

$$[v_1] = \frac{-(4,77 + 0,375\gamma_1)[v_2]}{(-2,2754 - 0,372\gamma_3)} \quad (4.25)$$

$$[v] = \begin{bmatrix} [v_1] \\ [v_2] \end{bmatrix} = \begin{bmatrix} \frac{-(4,77+0,375\gamma_1)[v_2]}{(-2,2754-0,372\gamma_3)} \\ [v_2] \end{bmatrix} \quad (4.26)$$

Se $[v_2] = [4, 77; 5, 145]$, então

$$[v_1] = [\underline{v}_1, \overline{v}_1] \Rightarrow \begin{cases} \underline{v}_1 = \frac{-(4,77+0,375\gamma_1)4,77}{(-2,2754-0,372\gamma_3)} = 8,5944, \text{ com } \gamma_1 = 0 \text{ e } \gamma_3 = 1 \\ \overline{v}_1 = \frac{-(4,77+0,375\gamma_1)5,145}{(-2,2754-0,372\gamma_3)} = 11,6335, \text{ com } \gamma_1 = 1 \text{ e } \gamma_3 = 0 \end{cases} \quad (4.27)$$

Assim, o autovetor $[v]$ será:

$$[v] = \begin{bmatrix} [1, 7988; 4, 5251] \\ [1; 2] \end{bmatrix} \quad (4.28)$$

Agora precisamos verificar se o autovetor encontrado é, de fato, um autovetor válido.

Sabemos que $A v = \lambda v$, assim queremos que $\Sigma[v] = [\lambda][v]$. Organizando as informações:

$$\Sigma = \begin{bmatrix} 10 & 4,77 + 0,375\gamma_1 \\ 4,77 + 0,375\gamma_1 & 2,2758 + 0,372\gamma_2 \end{bmatrix},$$

$$[\lambda_2] = 12,2754 + 0,372\gamma_3, \quad (4.29)$$

$$[v] = \begin{bmatrix} 8,5944 + 3,0391\gamma_4 \\ 4,77 + 0,375\gamma_5 \end{bmatrix} \text{ com } 0 \leq \gamma_i \leq 1$$

$$\Sigma[v] = \begin{bmatrix} 30,391\gamma_4 + (0,375\gamma_1 + 4,77)(0,375\gamma_5 + 4,77) + 85,944 \\ (0,375\gamma_1 + 4,77)(3,0391\gamma_4 + 8,5944) + (0,372\gamma_2 + 2,2758)(0,375\gamma_5 + 4,77) \end{bmatrix} \quad (4.30)$$

$$[\lambda_2][v] = \begin{bmatrix} (0,372\gamma_3 + 12,2754)(3,0391\gamma_4 + 8,5944) \\ (0,372\gamma_3 + 12,2754)(0,375\gamma_5 + 4,77) \end{bmatrix} \quad (4.31)$$

O que estamos buscando mostrar são os valores de $[v]$ que $\Sigma[v] = [\lambda_2][v]$. Como $\Sigma[v] = [\lambda_2][v]$, então $\Sigma[v] - [\lambda_2][v] = 0$ e os valores de $[v]$ são oriundos da variação de γ_4 e γ_5 . Logo, queremos:

$$\min f(\gamma) = \gamma_4 + \gamma_5$$

s.a. :

$$(30,391\gamma_4 + (0,375\gamma_1 + 4,77)(0,375\gamma_5 + 4,77) + 85,944) - ((0,372\gamma_3 + 12,2754)(3,0391\gamma_4 + 8,5944)) = 0$$

$$((0,375\gamma_1 + 4,77)(3,0391\gamma_4 + 8,5944) + (0,3727\gamma_2 + 2,2758)(0,375\gamma_5 + 4,77)) - ((0,372\gamma_3 + 12,2754)(0,375\gamma_5 + 4,77)) = 0$$

$$0 \leq \gamma_i \leq 1, \forall i = 1, \dots, 5.$$

(4.32)

e

$$\max f(\gamma) = \gamma_4 + \gamma_5$$

s.a. :

$$(30,391\gamma_4 + (0,375\gamma_1 + 4,77)(0,375\gamma_5 + 4,77) + 85,944) - ((0,372\gamma_3 + 12,2754)(3,0391\gamma_4 + 8,5944)) = 0$$

$$((0,375\gamma_1 + 4,77)(3,0391\gamma_4 + 8,5944) + (0,3727\gamma_2 + 2,2758)(0,375\gamma_5 + 4,77)) - ((0,372\gamma_3 + 12,2754)(0,375\gamma_5 + 4,77)) = 0$$

$$0 \leq \gamma_i \leq 1, \forall i = 1, \dots, 5.$$

(4.33)

Para o mínimo, temos $\gamma_4 = 0,3309$ e $\gamma_5 = 0$; e para o máximo, $\gamma_4 = 0,77210$ e $\gamma_5 = 1$.

Logo,

$$[v_{analítico}] = \begin{bmatrix} 8,5944 + 3,0391\gamma_4 \\ 4,77 + 0,375\gamma_5 \end{bmatrix} = \begin{bmatrix} [9,60; 10,7857] \\ [4,77; 5,145] \end{bmatrix} \quad (4.34)$$

Normalizando, temos:

$$[v]_{\text{analítico}} = \begin{bmatrix} \left[\frac{9,60}{\sqrt{9,60^2+4,77^2}}; \frac{10,7857}{\sqrt{10,7857^2+5,145^2}} \right] \\ \left[\frac{4,77}{\sqrt{9,60^2+4,77^2}}; \frac{5,145}{\sqrt{10,7857^2+5,145^2}} \right] \end{bmatrix} = \begin{bmatrix} [0,8955;0,9026] \\ [0,4450;0,4305] \end{bmatrix} = \begin{bmatrix} [0,8955;0,9026] \\ [0,4305;0,4450] \end{bmatrix} \quad (4.35)$$

4.3 Método Analítico Generalizado

Na seção anterior, definimos $[v_2] = [4,77;5,145]$, porém é interessante generalizar-mos essa escolha. Então, consideremos $[v_2] = [a;b]$, assim, temos que:

$$[v_1] = [\underline{v}_1, \bar{v}_1] \Rightarrow \begin{cases} \underline{v}_1 = \frac{-(4,77+0,375\chi)a}{(-2,2754-0,372\beta)} = c \\ \bar{v}_1 = \frac{-(4,77+0,375\chi)b}{(-2,2754-0,372\beta)} = d \end{cases} \quad (4.36)$$

Como os valores de χ e β variam no intervalo de $[0; 1]$, então os valores de c e d irão variar de $[1,8018; 2,2611]$ em função de a e b , respectivamente. É importante ressaltar, também, que se $a \leq 0$ e $b > 0$, teremos o autovetor nulo clássico presente no autovetor intervalar. Logo, para evitarmos isso precisamos que $a \leq b < 0$ ou $0 < a \leq b$.

Assim, o autovetor $[v]$ será:

$$[v] = \begin{bmatrix} [2,2611a;1,8018b] \\ [a;b] \end{bmatrix} \text{ caso } a \leq b < 0. \quad (4.37)$$

para garantir que, sendo $a \leq b < 0$, teremos $c \leq d < 0$. E

$$[v] = \begin{bmatrix} [1,8018a;2,2611b] \\ [a;b] \end{bmatrix} \text{ caso } 0 < a \leq b. \quad (4.38)$$

para garantir que, sendo $0 < a \leq b$, teremos $0 < c \leq d$.

Agora o que queremos é verificar se $[\Sigma][v] - [\lambda][v] = 0$. Por questão de organização, vamos primeiro realizar os cálculos para a primeira linha e depois para a segunda. Também, visando a organização, iremos considerar, inicialmente, $[v]$ como:

$$[v] = \begin{bmatrix} [c;d] \\ [a;b] \end{bmatrix} = \begin{bmatrix} c + (d-c)\chi \\ a + (b-a)\chi \end{bmatrix} \quad (4.39)$$

sendo o mesmo η para ambos os intervalos, devido a um estar em função do outro. Ou seja, a cada passo η que for dado no intervalo $[a; b]$, o mesmo deverá ser dado no intervalo $[c; d]$.

Assim, na primeira linha, temos:

$[\Sigma][v]$:

$$\begin{aligned} & 10(c + d\eta - c\eta) + (4,77 + 0,375\eta_2)(a + b\eta - a\eta) \\ & = 10(c + d\eta - c\eta) + 4,77(a + b\eta - a\eta) + 0,375(a\eta_2 + b\eta_2 - a\eta_2) \end{aligned} \quad (4.40)$$

$[\lambda][v]$:

$$\begin{aligned} & (12,2754 + 0,372\eta_3)(c + d\eta - c\eta) \\ & = 12,2754(c + d\eta - c\eta) + 0,372(c\eta_3 + d\eta_3 - c\eta_3) \end{aligned} \quad (4.41)$$

$[\Sigma][v] - [\lambda][v]$:

$$\begin{aligned} & -2,2754(c + d\eta - c\eta) + 4,77(a + b\eta - a\eta) + 0,375(a\eta_2 + b\eta_2 - a\eta_2) \\ & - 0,372(c\eta_3 + d\eta_3 - c\eta_3) = f_1(\eta, \eta_2, \eta_3) \end{aligned} \quad (4.42)$$

Na segunda linha, temos:

$[\Sigma][v]$:

$$\begin{aligned} & (4,77 + 0,375\eta_2)(c + d\eta - c\eta) + (2,2758 + 0,372\eta_4)(a + b\eta - a\eta) \\ & = 4,77(c + d\eta - c\eta) + 0,375(c\eta_2 + d\eta_2 - c\eta_2) + 2,2758(a + b\eta - a\eta) \\ & + 0,3727(a\eta_4 + b\eta_4 - a\eta_4) \end{aligned} \quad (4.43)$$

$[\lambda][v]$:

$$\begin{aligned} & (12,2754 + 0,372\eta_3)(a + b\eta - a\eta) \\ & = 12,2754(a + b\eta - a\eta) + 0,372(a\eta_3 + b\eta_3 - a\eta_3) \end{aligned} \quad (4.44)$$

$[\Sigma][v] - [\lambda][v]$:

$$\begin{aligned}
&= 4,77(c + d\gamma_1 - c\gamma_1) + 0,375(c\gamma_2 + d\gamma_1\gamma_2 - c\gamma_1\gamma_2) - 9,9996(a + b\gamma_1 - a\gamma_1) \\
&+ 0,3727(a\gamma_4 + b\gamma_1\gamma_4 - a\gamma_1\gamma_4) - 0,372(a\gamma_3 + b\gamma_1\gamma_3 - a\gamma_1\gamma_3) = f_2(\gamma_1, \gamma_2, \gamma_3, \gamma_4)
\end{aligned} \tag{4.45}$$

4.3.1 $a \leq b < 0$

Sendo $a < 0$ e $b < 0$, então $c = 2,2611a$ e $d = 1,8018b$. Logo, para $f_1(\gamma_1)$, temos:

$$\begin{aligned}
f_1(\gamma_1) = &-2,2754(2,2611a + 1,8018b\gamma_1 - 2,2611a\gamma_1) + 4,77(a + b\gamma_1 - a\gamma_1) + \\
&0,375(a\gamma_2 + b\gamma_1\gamma_2 - a\gamma_1\gamma_2) - 0,372(2,2611a\gamma_3 + 1,8018b\gamma_1\gamma_3 - 2,2611a\gamma_1\gamma_3)
\end{aligned} \tag{4.46}$$

Temos agora um problema de minimização e outro de maximização para verificar se os autovetores intervalares são válidos. Caso haja solução, será um intervalo válido. Dessa forma, temos:

$$\begin{aligned}
\min f_1(\gamma_1) = &-2,2754(2,2611a + 1,8018b\gamma_1 - 2,2611a\gamma_1) + 4,77(a + b\gamma_1 - a\gamma_1) + \\
&0,375(a\gamma_2 + b\gamma_1\gamma_2 - a\gamma_1\gamma_2) - 0,372(2,2611a\gamma_3 + 1,8018b\gamma_1\gamma_3 - 2,2611a\gamma_1\gamma_3)
\end{aligned}$$

s.a. :

$$0 \leq \gamma_1 \leq 1$$

$$f_1(\gamma_1) = 0$$

A priori, podemos relaxar o problema retirando a restrição associada ao γ_1 e avaliaremos a solução apresentada pelo SymPy:

```

1 import sympy as sp
2
3 # Definindo as variaveis simbolicas
4 a, b, x1, x2, x3, lambda0 = sp.symbols('a b x1 x2 x3 lambda0')
5
6 # Definindo a funcao objetivo
7 f = -2.2754*(2.2611*a + 1.8018*b*x1 - 2.2611*a*x1) + \
8 4.77*(a + b*x1 - a*x1) + 0.375*(a*x2 + b*x1*x2 - a*x1*x2) - \
9 0.372*(2.2611*a*x3 + 1.8018*b*x1*x3 - 2.2611*a*x1*x3)
10

```

```

11 # Definindo as restricoes ativas
12 g0 = -2.2754*(2.2611*a + 1.8018*b*x1 - 2.2611*a*x1) + \
13 4.77*(a + b*x1 - a*x1) + 0.375*(a*x2 + b*x1*x2 - a*x1*x2) - \
14 0.372*(2.2611*a*x3 + 1.8018*b*x1*x3 - 2.2611*a*x1*x3)
15
16 # Funcao Lagrangiana
17 L = f + lambda0 * g0
18
19 # Derivadas parciais da funcao Lagrangiana
20 dL_dx1 = sp.diff(L, x1)
21 dL_dx2 = sp.diff(L, x2)
22 dL_dx3 = sp.diff(L, x3)
23 dL_dlambda0 = sp.diff(L, lambda0)
24
25 # Resolvendo o sistema de equacoes derivadas
26 sol = sp.solve([dL_dx1, dL_dx2, dL_dx3, dL_dlambda0], (x1, x2,
27               x3, lambda0))
28
29 # Exibindo a solucao
30 print(sol)

```

A solução encontrada indica que o η_1 e η_3 são variáveis livres e η_2 está em função de η_1 e η_3 . Podemos verificar se há valores para $a \leq b < 0$, $0 \leq \eta_1 \leq 1$ e $0 \leq \eta_3 \leq 1$ que satisfaçam $0 \leq \eta_2 \leq 1$ através de simulação numérica. Avaliando na linguagem Python através do pacote Numpy:

```

1 import numpy as np
2
3 # Definindo a funcao que representa a expressao
4 def check_expression(a, b, x1, x3):
5     expr = 1.6e-7*(-14018820.0*a*x1*x3 - 6248449.0*a*x1 +
6               14018820.0*a*x3 +
7               6248449.0*a + 11171160.0*b*x1*x3 -
8               11169738.0*b*x1)/ \
9     (-a*x1 + a + b*x1)

```

```
8     return 0 <= expr <= 1
9
10 # Definindo as condicoes
11 def check_conditions(a, b, x1, x3):
12     return a <= b < 0 and 0 <= x1 <= 1 and 0 <= x3 <= 1
13
14 # Gerando valores aleatorios para a, b, x1, x3
15 np.random.seed(42) # Para manter a consistencia dos resultados
16 num_samples = 1000 # Numero de amostras a serem testadas
17 results = []
18
19 for _ in range(num_samples):
20     # Gerar valores aleatorios para a, b, x1, x3
21     a = np.random.uniform(-10, 0) # Gera valores para a entre
22         -10 e 0
23     b = np.random.uniform(a, 0) # Gera valores para b entre a
24         e 0
25     x1 = np.random.uniform(0, 1) # Gera valores para x1 entre
26         0 e 1
27     x3 = np.random.uniform(0, 1) # Gera valores para x3 entre
28         0 e 1
29
30     # Verificar se o conjunto de valores satisfaz as condicoes e
31     # a expressao
32     if check_conditions(a, b, x1, x3) and check_expression(a, b
33         , x1, x3):
34         results.append((a, b, x1, x3))
35
36 # Exibindo os resultados validos
37 if results:
38     print("Valores validos encontrados: {results}")
39 else:
40     print("Nenhuma solucao valida encontrada.")
```

O código encontrou soluções que atendem as condições estabelecidas. Haver soluções válidas numericamente, é um forte indício de que os valores a e b escolhidos são válidos para serem os extremos dos autovetores. Para o problema de maximização, basta mudar o sinal da função objetivo para verificar se o autovetor intervalar obtido será solução ou não.

Agora precisamos verificar a segunda linha representada por $f_2(\gamma)$ dado por:

$$f_2(\gamma) = 4,77(2,2611a + 1,8018b\gamma_1 - 2,2611a\gamma_1) + 0,375(2,2611a\gamma_2 + 1,8018b\gamma_1\gamma_2 - 2,2611a\gamma_1\gamma_2) - 9,9996(a + b\gamma_1 - a\gamma_1) + 0,3727(a\gamma_4 + b\gamma_1\gamma_4 - a\gamma_1\gamma_4) - 0,372(a\gamma_3 + b\gamma_1\gamma_3 - a\gamma_1\gamma_3)$$

Realizando o mesmo procedimento anterior para verificar a validade do intervalo, teremos o seguinte problema de otimização:

$$\begin{aligned} \min f_2(\gamma) = & 4,77(2,2611a + 1,8018b\gamma_1 - 2,2611a\gamma_1) + 0,375(2,2611a\gamma_2 + 1,8018b\gamma_1\gamma_2 \\ & - 2,2611a\gamma_1\gamma_2) - 9,9996(a + b\gamma_1 - a\gamma_1) + 0,3727(a\gamma_4 + b\gamma_1\gamma_4 - a\gamma_1\gamma_4) \\ & - 0,372(a\gamma_3 + b\gamma_1\gamma_3 - a\gamma_1\gamma_3) \end{aligned}$$

s.a. :

$$0 \leq \gamma \leq 1$$

$$f_2(\gamma) = 0$$

Igualmente ao processo anterior, podemos relaxar o problema retirando a restrição associada ao γ e avaliaremos a solução apresentada pelo SymPy:

```

1 import sympy as sp
2
3 # Definindo as variaveis simbolicas
4 a, b, x1, x2, x3, x4, lambda0 = sp.symbols('a b x1 x2 x3 x4
      lambda0')
5
6 # Definindo a funcao objetivo
7 f = 4.77*(2.2611*a + 1.8018*b*x1 - 2.2611*a*x1) + 0.375*(2.2611*
      a*x2 +
8 1.8018*b*x1*x2 - 2.2611*a*x1*x2) - 9.9996*(a + b*x1 - a*x1) +
      0.3727*(a*x4 +
9 b*x1*x4 - a*x1*x4) - 0.372*(a*x3 + b*x1*x3 - a*x1*x3)

```

```

10
11 # Definindo as restricoes ativas
12 g0 = 4.77*(2.2611*a + 1.8018*b*x1 - 2.2611*a*x1) +
      0.375*(2.2611*a*x2 +
13 1.8018*b*x1*x2 - 2.2611*a*x1*x2) - 9.9996*(a + b*x1 - a*x1) +
      0.3727*(a*x4 +
14 b*x1*x4 - a*x1*x4) - 0.372*(a*x3 + b*x1*x3 - a*x1*x3)
15
16 # Funcao Lagrangiana
17 L = f + lambda0 * g0
18
19 # Derivadas parciais da funcao Lagrangiana
20 dL_dx1 = sp.diff(L, x1)
21 dL_dx2 = sp.diff(L, x2)
22 dL_dx3 = sp.diff(L, x3)
23 dL_dx4 = sp.diff(L, x4)
24 dL_dlambda0 = sp.diff(L, lambda0)
25
26 # Resolvendo o sistema de equacoes derivadas
27 sol = sp.solve([dL_dx1, dL_dx2, dL_dx3, dL_dx4, dL_dlambda0], (
      x1, x2, x3, x4,
28 lambda0))
29
30 # Exibindo a solucao
31 print(sol)

```

A solução encontrada indica que o η_1 , η_3 e η_4 são variáveis livres e η_2 está em função delas. Podemos verificar se há valores para $a \leq b < 0$, $0 \leq \eta_1, \eta_3, \eta_4 \leq 1$ que satisfaçam $0 \leq \eta_2 \leq 1$ através de simulação numérica. Avaliando na linguagem Python através do pacote Numpy:

```

1 import numpy as np
2
3 # Definindo a funcao que representa a expressao
4 def check_expression(a, b, x1, x3, x4):
5     expr = 8.88888888888889e-12*\

```

```

6      (-372000000000000.0*a*x1*x3 + 372700000000000.0*a*x1*x4 +
7      785846999999999.0*a*x1 + 372000000000000.0*a*x3 -
          372700000000000.0*a*x4
8      - 785846999999999.0*a + 372000000000000.0*b*x1*x3
9      - 372700000000000.0*b*x1*x4 + 1.405014e+15*b*x1)/\
10     (-7537.0*a*x1 + 7537.0*a + 6006.0*b*x1)
11     return 0 <= expr <= 1
12
13 # Definindo as condicoes
14 def check_conditions(a, b, x1, x3, x4):
15     return a <= b < 0 and 0 <= x1 <= 1 and 0 <= x3 <= 1 and 0 <=
          x4 <= 1
16
17 # Gerando valores aleatorios para a, b, x1, x3, x4
18 np.random.seed(42) # Para manter a consistencia dos resultados
19 num_samples = 1000 # Numero de amostras a serem testadas
20 results = []
21
22 for _ in range(num_samples):
23     # Gerar valores aleatorios para a, b, x1, x3, x4
24     a = np.random.uniform(-10, 0) # Gera valores para a entre
          -10 e 0
25     b = np.random.uniform(a, 0) # Gera valores para b entre a
          e 0
26     x1 = np.random.uniform(0, 1) # Gera valores para x1 entre
          0 e 1
27     x3 = np.random.uniform(0, 1) # Gera valores para x3 entre
          0 e 1
28     x4 = np.random.uniform(0, 1) # Gera valores para x4 entre
          0 e 1
29
30     # Verificar se o conjunto de valores satisfaz as condicoes e
          a expressao
31     if check_conditions(a, b, x1, x3, x4) and check_expression(a

```

```

    , b, x1, x3, x4):
32     results.append((a, b, x1, x3, x4))
33
34 # exibindo os resultados validos
35 if results:
36     print("Valores validos encontrados: {results}")
37 else:
38     print("Nenhuma solucao valida encontrada.")

```

O código encontrou soluções que atendem as condições estabelecidas. Haver soluções válidas numericamente, é um forte indício de que os valores a e b escolhidos são válidos para serem os extremos dos autovetores. Para o problema de maximização, basta mudar o sinal da função objetivo para verificar se o autovetor intervalar obtido será solução ou não.

4.3.2 $0 < a \leq b$

Sendo $a > 0$ e $b > 0$, então $c = 1,8018a$ e $d = 2,2611b$. Logo, para $f_1(\gamma)$, temos:

$$f_1(\gamma) = -2,2754(1,8018a + 2,2611b\gamma_1 - 1,8018a\gamma_1) + 4,77(a + b\gamma_1 - a\gamma_1) + 0,375(a\gamma_2 + b\gamma_1\gamma_2 - a\gamma_1\gamma_2) - 0,372(1,8018a\gamma_3 + 2,2611b\gamma_1\gamma_3 - 1,8018a\gamma_1\gamma_3) \quad (4.47)$$

Temos agora um problema de minimização e outro de maximização para verificar se os intervalos dos autovetores são válidos. Caso haja solução, será um intervalo válido. Dessa forma, temos:

$$\min f_1(\gamma) = -2,2754(1,8018a + 2,2611b\gamma_1 - 1,8018a\gamma_1) + 4,77(a + b\gamma_1 - a\gamma_1) + 0,375(a\gamma_2 + b\gamma_1\gamma_2 - a\gamma_1\gamma_2) - 0,372(1,8018a\gamma_3 + 2,2611b\gamma_1\gamma_3 - 1,8018a\gamma_1\gamma_3)$$

s.a. :

$$0 \leq \gamma_i \leq 1$$

$$f_1(\gamma) = 0$$

A priori, podemos relaxar o problema retirando a restrição associada ao γ e avaliaremos a solução apresentada pelo SymPy:

```

1 import sympy as sp
2
3 # Definindo as variaveis simbolicas
4 a, b, x1, x2, x3, lambda0 = sp.symbols('a b x1 x2 x3 lambda0')
5
6 # Definindo a funcao objetivo
7 f = -2.2754*(1.8018*a + 2.2611*b*x1 - 1.8018*a*x1) + \
8 4.77*(a + b*x1 - a*x1) + 0.375*(a*x2 + b*x1*x2 - a*x1*x2) - \
9 0.372*(1.8018*a*x3 + 2.2611*b*x1*x3 - 1.8018*a*x1*x3)
10
11 # Definindo as restricoes ativas
12 g0 = -2.2754*(1.8018*a + 2.2611*b*x1 - 1.8018*a*x1) + \
13 4.77*(a + b*x1 - a*x1) + 0.375*(a*x2 + b*x1*x2 - a*x1*x2) - \
14 0.372*(1.8018*a*x3 + 2.2611*b*x1*x3 - 1.8018*a*x1*x3)
15
16 # Funcao Lagrangiana
17 L = f + lambda0 * g0
18
19 # Derivadas parciais da funcao Lagrangiana
20 dL_dx1 = sp.diff(L, x1)
21 dL_dx2 = sp.diff(L, x2)
22 dL_dx3 = sp.diff(L, x3)
23 dL_dlambda0 = sp.diff(L, lambda0)
24
25 # Resolvendo o sistema de equacoes derivadas
26 sol = sp.solve([dL_dx1, dL_dx2, dL_dx3, dL_dlambda0], (x1, x2,
27 x3, lambda0))
28 # Exibindo a solucao
29 print(sol)

```

A solução encontrada indica que o η e β são variáveis livres e γ está em função de η e β . Podemos verificar se há valores para $a \leq b < 0$, $0 \leq \eta \leq 1$ e $0 \leq \beta \leq 1$ que satisfaçam

$0 \leq \eta \leq 1$ através de simulação numérica. Avaliando na linguagem Python através do pacote Numpy:

```

1 import numpy as np
2
3 # Definindo a funcao que representa a expressao
4 def check_expression(a, b, x1, x3):
5     expr = 1.6e-7*(-11171160.0*a*x1*x3 + 11169738.0*a*x1 +
6             11171160.0*a*x3
7             - 11169738.0*a + 14018820.0*b*x1*x3 +
8             6248449.0*b*x1)\
9
10    /(-a*x1 + a + b*x1)
11    return 0 <= expr <= 1
12
13
14 # Definindo as condicoes
15 def check_conditions(a, b, x1, x3):
16     return a <= b < 0 and 0 <= x1 <= 1 and 0 <= x3 <= 1
17
18
19 # Gerando valores aleatorios para a, b, x1, x3
20 np.random.seed(42) # Para manter a consistencia dos resultados
21 num_samples = 1000 # Numero de amostras a serem testadas
22 results = []
23
24 for _ in range(num_samples):
25     # Gerar valores aleatorios para a, b, x1, x3
26     a = np.random.uniform(-10, 0) # Gera valores para a entre
27         -10 e 0
28     b = np.random.uniform(a, 0) # Gera valores para b entre a
29         e 0
30     x1 = np.random.uniform(0, 1) # Gera valores para x1 entre
31         0 e 1
32     x3 = np.random.uniform(0, 1) # Gera valores para x3 entre
33         0 e 1

```

```

27     # Verificar se o conjunto de valores satisfaz as condicoes e
        a expressao
28     if check_conditions(a, b, x1, x3) and check_expression(a, b
        , x1, x3):
29         results.append((a, b, x1, x3))
30
31 # Exibindo os resultados validos
32 if results:
33     print(f"Valores validos encontrados: {results}")
34 else:
35     print("Nenhuma solucao valida encontrada.")

```

O código encontrou soluções que atendem às condições estabelecidas. Haver soluções válidas numericamente, é um forte indício de que os valores a e b escolhidos são válidos como autovetores. Para o problema de maximização, mudaríamos o sinal da função objetivo e isso nos iria direcionar, também, a uma solução, devido à característica monótona dos intervalos.

Agora precisamos verificar a segunda linha representada por $f_2(\gamma)$. Temos:

$$f_2(\gamma) = 4,77(1,8018a + 2,2611b\gamma_1 - 1,8018a\gamma_1) + 0,375(1,8018a\gamma_2 + 2,2611b\gamma_1\gamma_2 - 1,8018a\gamma_1\gamma_2) - 9,9996(a + b\gamma_1 - a\gamma_1) + 0,3727(a\gamma_4 + b\gamma_1\gamma_4 - a\gamma_1\gamma_4) - 0,372(a\gamma_3 + b\gamma_1\gamma_3 - a\gamma_1\gamma_3)$$

Realizando o mesmo procedimento anterior para verificar a validade do intervalo.

Teremos o seguinte problema de otimização:

$$\begin{aligned} \min f_2(\gamma) = & 4,77(1,8018a + 2,2611b\gamma_1 - 1,8018a\gamma_1) + 0,375(1,8018a\gamma_2 + 2,2611b\gamma_1\gamma_2 \\ & - 1,8018a\gamma_1\gamma_2) - 9,9996(a + b\gamma_1 - a\gamma_1) + 0,3727(a\gamma_4 + b\gamma_1\gamma_4 - a\gamma_1\gamma_4) \\ & - 0,372(a\gamma_3 + b\gamma_1\gamma_3 - a\gamma_1\gamma_3) \end{aligned}$$

s.a. :

$$0 \leq \gamma_i \leq 1$$

$$f_2(\gamma) = 0$$

Igualmente ao processo anterior, podemos relaxar o problema retirando a restrição associada ao γ e avaliaremos a solução apresentada pelo SymPy:

```

1 import sympy as sp
2
3 # Definindo as variaveis simbolicas
4 a, b, x1, x2, x3, x4, lambda0 = sp.symbols('a b x1 x2 x3 x4
      lambda0')
5
6 # Definindo a funcao objetivo
7 f = 4.77*(1.8018*a + 2.2611*b*x1 - 1.8018*a*x1) + 0.375*(1.8018*
      a*x2 +
8 2.2611*b*x1*x2 - 1.8018*a*x1*x2) - 9.9996*(a + b*x1 - a*x1) +
      0.3727*(a*x4 +
9 b*x1*x4 - a*x1*x4) - 0.372*(a*x3 + b*x1*x3 - a*x1*x3)
10
11 # Definindo as restricoes ativas
12 g0 = 4.77*(1.8018*a + 2.2611*b*x1 - 1.8018*a*x1) +
      0.375*(1.8018*a*x2 +
13 2.2611*b*x1*x2 - 1.8018*a*x1*x2) - 9.9996*(a + b*x1 - a*x1) +
      0.3727*(a*x4 +
14 b*x1*x4 - a*x1*x4) - 0.372*(a*x3 + b*x1*x3 - a*x1*x3)
15
16 # Funcao Lagrangiana
17 L = f + lambda0 * g0
18
19 # Derivadas parciais da funcao Lagrangiana
20 dL_dx1 = sp.diff(L, x1)
21 dL_dx2 = sp.diff(L, x2)
22 dL_dx3 = sp.diff(L, x3)
23 dL_dx4 = sp.diff(L, x4)
24 dL_dlambda0 = sp.diff(L, lambda0)
25
26 # Resolvendo o sistema de equacoes derivadas
27 sol = sp.solve([dL_dx1, dL_dx2, dL_dx3, dL_dx4, dL_dlambda0], (
      x1, x2, x3, x4,
28 lambda0))

```

```

29
30 # Exibindo a solucao
31 print(sol)

```

A solução encontrada indica que o γ_1 , γ_3 e γ_4 são variáveis livres e γ_2 está em função delas. Podemos verificar se há valores para $a \leq b < 0$, $0 \leq \gamma_1, \gamma_3, \gamma_4 \leq 1$ que satisfaçam $0 \leq \gamma_2 \leq 1$ através de simulação numérica. Avaliando na linguagem Python através do pacote Numpy:

```

1 import numpy as np
2
3 # Definindo a funcao que representa a expressao
4 def check_expression(a, b, x1, x3, x4):
5     expr = 8.88888888888889e-12*\
6     (-372000000000000.0*a*x1*x3 + 372700000000000.0*a*x1*x4
7     - 1.405014e+15*a*x1 + 372000000000000.0*a*x3 -
8     372700000000000.0*a*x4
9     + 1.405014e+15*a + 372000000000000.0*b*x1*x3 -
10    372700000000000.0*b*x1*x4
11    - 785846999999999.0*b*x1)/(-6006.0*a*x1 + 6006.0*a +
12    7537.0*b*x1)
13    return 0 <= expr <= 1
14
15 # Definindo as condicoes
16 def check_conditions(a, b, x1, x3, x4):
17     return a <= b < 0 and 0 <= x1 <= 1 and 0 <= x3 <= 1 and 0 <=
18     x4 <= 1
19
20 # Gerando valores aleatorios para a, b, x1, x3, x4
21 np.random.seed(42) # Para manter a consistencia dos resultados
22 num_samples = 1000 # Numero de amostras a serem testadas
23 results = []
24
25 for _ in range(num_samples):
26     # Gerar valores aleatorios para a, b, x1, x3, x4
27     a = np.random.uniform(-10, 0) # Gera valores para a entre

```

```

-10 e 0
24  b = np.random.uniform(a, 0)      # Gera valores para b entre a
    e 0
25  x1 = np.random.uniform(0, 1)     # Gera valores para x1 entre
    0 e 1
26  x3 = np.random.uniform(0, 1)     # Gera valores para x3 entre
    0 e 1
27  x4 = np.random.uniform(0, 1)     # Gera valores para x4 entre
    0 e 1
28
29  # Verificar se o conjunto de valores satisfaz as condicoes e
    a expressao
30  if check_conditions(a, b, x1, x3, x4) and check_expression(a
    , b, x1, x3, x4):
31      results.append((a, b, x1, x3, x4))
32
33 # exibindo os resultados validos
34 if results:
35     print("Valores validos encontrados: {results}")
36 else:
37     print("Nenhuma solucao valida encontrada.")

```

O código encontrou soluções que atendem às condições estabelecidas. Haver soluções válidas numericamente, é um forte indício de que os valores a e b escolhidos são válidos como autovetores. Para o problema de maximização, mudaríamos o sinal da função objetivo e isso nos iria direcionar, também, a uma solução, devido à característica monótona dos intervalos.

4.4 PCA Intervalar

Como falado no capítulo anterior, a técnica de PCA realiza uma transformação no conjunto de dados avaliado. Inicialmente, temos a matriz intervalar $[A]$:

$$[\mathbf{A}] = \begin{bmatrix} 2 & [0,95;1,11] \\ 4 & [1,92;2,01] \\ 6 & [2,94;3,05] \\ 8 & [3,93;4,03] \\ 10 & [4,92;5,04] \end{bmatrix} \quad (4.48)$$

Que, de acordo com CIA, pode ser escrita da seguinte forma:

$$[\mathbf{A}] = \begin{bmatrix} 2 & 0,95 + 0,16\gamma_1 \\ 4 & 1,92 + 0,09\gamma_2 \\ 6 & 2,94 + 0,11\gamma_3 \\ 8 & 3,93 + 0,10\gamma_4 \\ 10 & 4,92 + 0,12\gamma_5 \end{bmatrix} \quad (4.49)$$

A componente principal a ser considerada são os autovetores $[v]_{sl}$, $[v]_{cia}$, $[v]_{deif-lv}$ e $[v]_{analitico}$ apresentados anteriormente.

Para $[v]_{sl}$:

$$[v]_{sl} = \begin{bmatrix} [0,8892;0,9026] \\ [0,4306;0,4575] \end{bmatrix} = \begin{bmatrix} 0,8892 + 0,0134\gamma_6 \\ 0,4306 + 0,0269\gamma_7 \end{bmatrix} \quad (4.50)$$

Dessa forma, temos:

$$[\mathbf{A}][v]_{sl} = \begin{bmatrix} 2(0,8892 + 0,0134\gamma_6) + (0,95 + 0,16\gamma_1)(0,4306 + 0,0269\gamma_7) \\ 4(0,8892 + 0,0134\gamma_6) + (1,92 + 0,09\gamma_2)(0,4306 + 0,0269\gamma_7) \\ 6(0,8892 + 0,0134\gamma_6) + (2,94 + 0,11\gamma_3)(0,4306 + 0,0269\gamma_7) \\ 8(0,8892 + 0,0134\gamma_6) + (3,93 + 0,10\gamma_4)(0,4306 + 0,0269\gamma_7) \\ 10(0,8892 + 0,0134\gamma_6) + (4,92 + 0,12\gamma_5)(0,4306 + 0,0269\gamma_7) \end{bmatrix} \quad (4.51)$$

$$[\mathbf{A}][v]_{sl} = PCA_{sl} = \begin{bmatrix} [2,1875;2,3130] \\ [4,3836;4,5300] \\ [6,6012;6,8110] \\ [8,8059;9,0645] \\ [11,0106;11,3318] \end{bmatrix}, \text{ com } \gamma_i = 0 \text{ e } \gamma_i = 1 \text{ para m\u00ednimo e m\u00e1ximo.} \quad (4.52)$$

Para $[v]_{cia}$:

$$[v]_{cia} = \begin{bmatrix} [0,8958;0,8960] \\ [0,4440;0,4444] \end{bmatrix} = \begin{bmatrix} 0,8958 + 0,0002\gamma_6 \\ 0,4440 + 0,0004\gamma_7 \end{bmatrix} \quad (4.53)$$

Dessa forma, temos:

$$[A][v]_{cia} = \begin{bmatrix} 2(0,8958 + 0,0002\gamma_6) + (0,95 + 0,16\gamma_1)(0,4440 + 0,0004\gamma_7) \\ 4(0,8958 + 0,0002\gamma_6) + (1,92 + 0,09\gamma_2)(0,4440 + 0,0004\gamma_7) \\ 6(0,8958 + 0,0002\gamma_6) + (2,94 + 0,11\gamma_3)(0,4440 + 0,0004\gamma_7) \\ 8(0,8958 + 0,0002\gamma_6) + (3,93 + 0,10\gamma_4)(0,4440 + 0,0004\gamma_7) \\ 10(0,8958 + 0,0002\gamma_6) + (4,92 + 0,12\gamma_5)(0,4440 + 0,0004\gamma_7) \end{bmatrix} \quad (4.54)$$

$$[A][v]_{cia} = PCA_{cia} = \begin{bmatrix} [2,2134;2,2853] \\ [4,4357;4,4772] \\ [6,6802;6,7314] \\ [8,9113;8,9589] \\ [11,1425;11,1998] \end{bmatrix}, \text{ com } \gamma_i = 0 \text{ e } \gamma_i = 1 \text{ para m\u00ednimo e m\u00e1ximo.} \quad (4.55)$$

Para $[v]_{deif-lv}$:

$$[v]_{deif-lv} = \begin{bmatrix} [0,8913;0,9005] \\ [0,4348;0,4535] \end{bmatrix} = \begin{bmatrix} 0,8913 + 0,0092\gamma_6 \\ 0,4348 + 0,0187\gamma_7 \end{bmatrix} \quad (4.56)$$

Dessa forma, temos:

$$[A][v]_{deif-lv} = \begin{bmatrix} 2(0,8913 + 0,0092\gamma_6) + (0,95 + 0,16\gamma_1)(0,4348 + 0,0187\gamma_7) \\ 4(0,8913 + 0,0092\gamma_6) + (1,92 + 0,09\gamma_2)(0,4348 + 0,0187\gamma_7) \\ 6(0,8913 + 0,0092\gamma_6) + (2,94 + 0,11\gamma_3)(0,4348 + 0,0187\gamma_7) \\ 8(0,8913 + 0,0092\gamma_6) + (3,93 + 0,10\gamma_4)(0,4348 + 0,0187\gamma_7) \\ 10(0,8913 + 0,0092\gamma_6) + (4,92 + 0,12\gamma_5)(0,4348 + 0,0187\gamma_7) \end{bmatrix} \quad (4.57)$$

$$[A][v]_{deif-lv} = PCA_{deif-lv} = \begin{bmatrix} [2, 1957; 2, 3044] \\ [4, 4000; 4, 5135] \\ [6, 6261; 6, 7862] \\ [8, 8392; 9, 0316] \\ [11, 0522; 11, 2906] \end{bmatrix}, \text{ com } \gamma_i = 0 \text{ e } \gamma_i = 1 \text{ para m\u00ednimo e m\u00e1ximo.} \quad (4.58)$$

Para $[v]_{analitico}$:

$$[v]_{analitico} = \begin{bmatrix} [0, 8955; 0, 9026] \\ [0, 4305; 0, 4450] \end{bmatrix} = \begin{bmatrix} 0, 8955 + 0, 0071\gamma_6 \\ 0, 4305 + 0, 0145\gamma_7 \end{bmatrix} \quad (4.59)$$

Dessa forma, temos:

$$[A][v]_{analitico} = \begin{bmatrix} 2(0, 8955 + 0, 0071\gamma_6) + (0, 95 + 0, 16\gamma_1)(0, 4305 + 0, 0145\gamma_7) \\ 4(0, 8955 + 0, 0071\gamma_6) + (1, 92 + 0, 09\gamma_2)(0, 4305 + 0, 0145\gamma_7) \\ 6(0, 8955 + 0, 0071\gamma_6) + (2, 94 + 0, 11\gamma_3)(0, 4305 + 0, 0145\gamma_7) \\ 8(0, 8955 + 0, 0071\gamma_6) + (3, 93 + 0, 10\gamma_4)(0, 4305 + 0, 0145\gamma_7) \\ 10(0, 8955 + 0, 0071\gamma_6) + (4, 92 + 0, 12\gamma_5)(0, 4305 + 0, 0145\gamma_7) \end{bmatrix} \quad (4.60)$$

$$[A][v]_{analitico} = PCA_{analitico} = \begin{bmatrix} [2, 2000; 2, 2992] \\ [4, 4086; 4, 5049] \\ [6, 6387; 6, 7729] \\ [8, 8559; 9, 0142] \\ [11, 0731; 11, 2688] \end{bmatrix}, \text{ com } \gamma_i = 0 \text{ e } \gamma_i = 1 \text{ para m\u00ednimo e m\u00e1ximo.} \quad (4.61)$$

Visando compararmos as quatro componentes calculadas anteriormente, podemos calcular a diferen\u00e7a entre os limites superiores e inferiores dos intervalos de cada componente e calcular a m\u00e9dia das diferen\u00e7as, conforme apresentado na Tabela 6:

Tabela 6 – Média das diferenças

PCA	Média Limites Superior e Inferior	Média das diferenças
Analítico	6,7976 - 6,6353	0,1623
Single-Level	6,8101 - 6,5978	0,2123
CIA	6,7305 - 6,6766	0,0539
Deif com Leverrier-Faddeev	6,7853 - 6,6226	0,1626

Fonte: próprio autor.

Assim, percebemos que a modelagem através do método CIA é a modelagem que captou a variabilidade e a incerteza do processo sem aumentar consideravelmente o intervalo.

Diante de tais resultados, iremos testar o método CIA para a construção de PCA Intervalar aplicado em situações reais.

5 APLICAÇÃO DE COMPONENTES PRINCIPAIS INTERVALARES

Neste capítulo, iremos abordar duas aplicações para a análise de componentes principais intervalares, utilizando o método CIA para avaliar os resultados.

5.1 Avaliação de qualidade da água

O artigo “Uso de estatística multivariada como ferramenta na avaliação da qualidade da lagoa da Maraponga, Fortaleza, Ceará, Brasil” (Santos *et al.*, 2011) avaliou 15 amostras de água da lagoa coletadas a cada dois meses. Através da técnica de análise de componentes principais, Santos *et al.* (2011) identificou oito indicadores relevantes para a análise: cor verdadeira (CV), turbidez (TURB), sólidos totais (ST), sólidos totais voláteis (STV), ortofosfato solúvel (OPS), sulfato (SO₄), teor de manganês (Mn) e ferro (Fe).

Os dados completos destas amostras não estão disponíveis, mas no trabalho é disponibilizada informação mostrada na Tabela 7.

Tabela 7 – Parâmetros apresentados no artigo

Parâmetro	Mínimo	Máximo	Média	Desvio-padrão
CV (uH)	18	61	34	10
TURB (uT)	10	28	16	3
ST (mg/L)	325	874	512	120
STV (mg/L)	18	553	157	105
OPS (mg/L)	0	0,225	0,066	0,052
SO ₄ (mg/L)	1,5	47,7	17	8,5
Mn (mg/L)	0,002	0,114	0,032	0,016
Fe (mg/L)	0,05	0,799	0,217	0,126

Fonte: Santos *et al.* (2011).

Através das informações da Tabela 7, foi possível simular 50 amostras intervalares para cada um dos oito indicadores. As amostras foram simuladas utilizando a linguagem de programação Python e considerando a Distribuição Normal, com as médias e desvio-padrão da Tabela 7. Abaixo segue o exemplo do código para uma das bases simuladas:

```

1 import numpy as np
2 import pandas as pd
3
4 def generate_normal(mean, std_dev, min_val, max_val, size=1):
5     samples = []
6     while len(samples) < size:

```

```

7         num = np.random.normal(mean, std_dev)
8         if min_val <= num <= max_val:
9             samples.append(num)
10        return samples
11
12 # Exemplo de uso
13 matrix = np.array([generate_normal(mean=0.217, std_dev=0.116,
14                               min_val=0.05, max_val=0.799, size=2) for _ in range(50)])
15
16 # Exporta para CSV
17 pd.DataFrame(matrix).to_csv("dados_fe.csv", sep=';', decimal=',',
18                             , index=False, header=['fe_Min', 'fe_Max'])
19
19 print(matrix)

```

Como os indicadores possuem escalas diferentes, as amostras geradas foram normalizadas pelo método de valor máximo e mínimo, apresentados no artigo, para que nenhum indicador influencie erroneamente o processo de análise.

Após esse processo, a base de dados pode ser remodelada utilizando a metodologia CIA, apresentada nos capítulos anteriores, e chegamos à seguinte matriz de covariância intervalar (dados arredondados até a terceira casa decimal):

$$[\Sigma] = \begin{bmatrix}
[0,446;2,977] & [0,066;0,11] & [0,014;0,231] & [-0,011;0,112] & [-0,023;-0,011] & [-0,18;0,162] & [-0,007;0,042] & [0,072;0,631] \\
[0,066;0,11] & [0,176;1,553] & [0,065;0,253] & [-0,33;-0,004] & [-0,205;0,038] & [-0,412;-0,029] & [0,017;0,019] & [-0,167;0,057] \\
[0,014;0,231] & [0,065;0,253] & [0,435;2,391] & [-0,732;-0,086] & [0,041;0,197] & [-0,437;-0,116] & [-0,01;0,117] & [0,039;0,162] \\
[-0,011;0,112] & [-0,33;-0,004] & [-0,732;-0,086] & [0,174;1,763] & [0,013;0,076] & [-0,007;0,066] & [-0,01;0,011] & [-0,304;-0,045] \\
[-0,023;-0,011] & [-0,205;0,038] & [0,041;0,197] & [0,013;0,076] & [0,335;2,024] & [-0,767;-0,04] & [-0,158;0,027] & [0,009;0,073] \\
[-0,18;0,162] & [-0,412;-0,029] & [-0,437;-0,116] & [-0,007;0,066] & [-0,767;-0,04] & [0,585;2,418] & [-0,098;0,035] & [-0,103;0,006] \\
[-0,007;0,042] & [0,017;0,019] & [-0,01;0,117] & [-0,01;0,011] & [-0,158;0,027] & [-0,098;0,035] & [0,126;1,132] & [0,016;0,108] \\
[0,072;0,631] & [-0,167;0,057] & [0,039;0,162] & [-0,304;-0,045] & [0,009;0,073] & [-0,103;0,006] & [0,016;0,108] & [0,252;2,278]
\end{bmatrix} \tag{5.1}$$

Nesta matriz aplicamos os procedimentos dos dois capítulos anteriores e chegamos a dois autovetores intervalares. Abaixo apresentamos cada autovetor intervalar e ao seu lado o autovetor clássico descrito por Santos *et al.* (2011), sendo v_1 e v_2 os autovetores apresentados no artigo e $[v_1]$ e $[v_2]$ os autovetores intervalares calculados:

$$\begin{aligned}
 [v_1] = \begin{bmatrix} [1,484;3,067] \\ [0,291;0,864] \\ [0,57;1,696] \\ [0,212;0,996] \\ [0,228;0,938] \\ [0,331;1,684] \\ [0,142;0,303] \\ [0,985;1,663] \end{bmatrix} & \text{ e } v_1 = \begin{bmatrix} 0,684 \\ 0,768 \\ 0,62 \\ 0,692 \\ 0,626 \\ 0,732 \\ 0,865 \\ 0,782 \end{bmatrix} ; [v_2] = \begin{bmatrix} [-0,907;-0,008] \\ [0,007;0,196] \\ [0,02;0,337] \\ [0,002;0,037] \\ [0,013;0,818] \\ [0;0,995] \\ [0,002;0,038] \\ [-0,415;-0,002] \end{bmatrix} & \text{ e } v_2 = \begin{bmatrix} -0,594 \\ 0,246 \\ 0,673 \\ 0,526 \\ -0,39 \\ -0,28 \\ 0,059 \\ -0,214 \end{bmatrix} \\
 & \hspace{15em} (5.2)
 \end{aligned}$$

O que pretendemos avaliar é se o resultado intervalar contém o resultado clássico do trabalho. Avaliaremos abaixo da seguinte forma: caso o intervalo contenha o resultado clássico, então o erro será igual a zero; senão, calcula-se o erro quadrático do resultado clássico para a extremidade mais próxima do intervalo. A Tabela 8 abaixo traz as seguintes informações: autovetor intervalar calculado neste trabalho, autovetor clássico apresentado no artigo e o erro quadrático, o quanto o valor do artigo está coberto pelo intervalo.

Tabela 8 – Comparativo de autovetor intervalar com clássico

Autovetor Intervalar	Autovetor Clássico	Erro quadrático
[1,484;3,067]	0,684	0,640
[0,291;0,864]	0,768	0
[0,570;1,696]	0,620	0
[0,212;0,996]	0,692	0
[0,228;0,938]	0,626	0
[0,331;1,684]	0,732	0
[0,142;0,303]	0,865	0,316
[0,985;1,663]	0,782	0,041
Média	–	0,125
[-0,907;-0,008]	-0,594	0
[0,007;0,196]	0,246	0,002
[0,020;0,337]	0,673	0,113
[0,002;0,037]	0,526	0,239
[0,013;0,818]	-0,390	0,162
[0;0,995]	-0,280	0,078
[0,002;0,038]	0,059	0
[-0,415;-0,002]	-0,214	0
Média	–	0,074

Fonte: próprio autor.

O erro apresentado foi relativamente baixo, o que indica que o desempenho do método foi satisfatório. É importante considerar que, como o resultado obtido depende de uma simulação, ele pode variar dependendo das condições e dos parâmetros envolvidos. Além disso, o método foi aplicado a uma base de dados que continha oito colunas, o que representa uma estrutura de dados um pouco mais complexa. Mesmo com essa complexidade adicional, pode-se concluir que o método CIA apresentou resultados bons, demonstrando sua eficiência e robustez ao lidar com bases de dados maiores.

5.2 Análise psicométrica

Como discutido na introdução deste trabalho, a principal meta é aprimorar a avaliação da escala Likert, um instrumento amplamente utilizado em pesquisas psicométricas. Para atingir esse objetivo, é fundamental desenvolver uma técnica de Análise de Componentes Principais (PCA) que seja capaz de lidar com o contexto difuso, levando em consideração as diversas nuances e variabilidades envolvidas. No entanto, antes de aplicar essa técnica no contexto difuso, é necessário primeiro compreender e desenvolver a técnica de PCA no contexto intervalar. Ao longo deste trabalho, buscamos construir essa técnica, a qual nos proporciona as ferramentas adequadas para realizar uma análise psicométrica robusta e confiável dentro de um contexto intervalar. Isso é especialmente relevante porque, ao trabalhar com dados obtidos de escalas Likert, precisamos garantir que as transformações e análises dos dados respeitem as propriedades dessa escala.

O trabalho de Monte (2020), por sua vez, faz aplicação de avaliações da escala Likert com propostas difusas no processo seletivo do Programa de Aprendizagem Cooperativa em Células Estudantis (PACCE). Durante esse processo seletivo, os candidatos participam de uma formação em Aprendizagem Cooperativa que ocorre ao longo de cinco dias. Em cada um desses dias, os candidatos são avaliados por três avaliadores distintos, utilizando a escala Likert para medir diversos aspectos do seu desempenho. O grande desafio e a questão central da nossa investigação consistem em analisar se, ao longo desse período de formação, os candidatos conseguem, de fato, melhorar seus indicadores, ou seja, se existe uma evolução significativa em suas avaliações de um dia para o outro. É importante que esta avaliação seja realizada combinando os indicadores avaliados.

Para exemplificar essa abordagem, apresentamos abaixo uma tabela que ilustra o processo de avaliação de uma das salas participantes. A sala em questão é composta por 16

candidatos, e cada um deles é avaliado por três avaliadores diferentes. Caso haja consenso entre os avaliadores, a nota atribuída será um número clássico. No entanto, quando não há consenso entre os avaliadores, a avaliação assume a forma de um número intervalar, o que exige uma abordagem mais sofisticada para análise e interpretação dos dados:

Tabela 9 – Desempenho dos Candidatos por Indicadores

Dia	Candidato	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5
DIA 1	Candidato 1	4	4	3	3	2
	Candidato 2	[3; 4]	3	3	3	2
	Candidato 3	4	4	3	3	3
	Candidato 4	2	2	[2; 3]	[2; 3]	[2; 3]
	Candidato 5	3	3	3	3	3
	Candidato 6	3	3	3	3	4
	Candidato 7	[3; 4]	4	3	3	3
	Candidato 8	2	3	[2; 3]	3	3
	Candidato 9	3	3	3	3	4
	Candidato 10	3	3	3	4	3
	Candidato 11	[3; 4]	[3; 4]	[4; 5]	[3; 4]	[3; 4]
	Candidato 12	3	3	3	3	4
	Candidato 13	4	3	3	4	3
	Candidato 14	3	3	3	[3; 4]	[2; 3]
	Candidato 15	4	3	3	[3; 4]	2
	Candidato 16	3	3	3	3	2
DIA 2	Candidato 1	4	4	3	3	3
	Candidato 2	2	2	3	3	3
	Candidato 3	[4; 5]	[4; 5]	3	3	3
	Candidato 4	3	3	3	2	3
	Candidato 5	[3; 4]	[3; 4]	3	3	3
	Candidato 6	[3; 4]	3	3	3	[3; 4]
	Candidato 7	4	4	3	3	3
	Candidato 8	[3; 4]	4	3	3	[3; 4]
	Candidato 9	[3; 4]	[3; 4]	3	[3; 4]	[3; 4]
	Candidato 10	3	[3; 4]	3	3	3
	Candidato 11	3	[3; 4]	3	3	3
	Candidato 12	3	[3; 4]	[3; 4]	3	3
	Candidato 13	3	[3; 4]	3	3	3
	Candidato 14	[2; 3]	[2; 3]	[2; 3]	3	3
	Candidato 15	3	3	3	3	3
	Candidato 16	[2; 3]	2	3	3	3

Fonte: próprio autor.

Utilizando a modelagem do método CIA iremos encontrar as seguintes matrizes intervalares de covariância para os dias 1 e 2, respectivamente:

$$[\Sigma] = \begin{bmatrix} [10,0645;12,2581] & [10,0645;12,0968] & [9,3871;10,8065] & [9,5484;11,0645] & [9,1613;10,7742] \\ [10,0645;12,0968] & [10,2258;12,2581] & [9,4516;10,8387] & [9,6129;11] & [9,2903;10,8065] \\ [9,3871;10,8065] & [9,4516;10,8387] & [9,0323;10,0323] & [9,129;10,129] & [8,8387;9,9355] \\ [9,5484;11,0645] & [9,6129;11] & [9,129;10,129] & [9,4194;10,4839] & [9,0323;10,0968] \\ [9,1613;10,7742] & [9,2903;10,8065] & [8,8387;9,9355] & [9,0323;10,0968] & [9;10,2258] \end{bmatrix} \quad (5.3)$$

$$[\Sigma] = \begin{bmatrix} 10,0645 + 2,1935\gamma_1 & 10,0645 + 2,0323\gamma_2 & 9,3871 + 1,4194\gamma_3 & 9,5484 + 1,5161\gamma_4 & 9,1613 + 1,6129\gamma_5 \\ 10,0645 + 2,0323\gamma_2 & 10,2258 + 2,0323\gamma_2 & 9,4516 + 1,3871\gamma_3 & 9,6129 + 1,3871\gamma_4 & 9,2903 + 1,5161\gamma_5 \\ 9,3871 + 1,4194\gamma_3 & 9,4516 + 1,3871\gamma_3 & 9,0323 + 1\gamma_3 & 9,129 + 1\gamma_4 & 8,8387 + 1,0968\gamma_5 \\ 9,5484 + 1,5161\gamma_4 & 9,6129 + 1,3871\gamma_4 & 9,129 + 1\gamma_4 & 9,4194 + 1,0645\gamma_4 & 9,0323 + 1,0645\gamma_5 \\ 9,1613 + 1,6129\gamma_5 & 9,2903 + 1,5161\gamma_5 & 8,8387 + 1,0968\gamma_5 & 9,0323 + 1,0645\gamma_5 & 9 + 1,2258\gamma_5 \end{bmatrix} \quad (5.4)$$

Como, a princípio, estamos interessados apenas em uma métrica combinando os cinco indicadores, iremos nos atentar apenas ao maior autovalor. Logo, aplicamos o método numérico de *Leverrier-Faddeev* para determinar qual o autovetor associado ao maior autovalor intervalar. Abaixo segue o autovetor normalizado:

$$[v] = \begin{bmatrix} [0,4712;0,5438] \\ [0,4712;0,5183] \\ [0,3649;0,4274] \\ [0,3736;0,4359] \\ [0,4037;0,4281] \end{bmatrix} = \begin{bmatrix} 0,4712 + 0,0726\delta_1 \\ 0,4712 + 0,0471\delta_2 \\ 0,3649 + 0,0625\delta_3 \\ 0,3736 + 0,0623\delta_4 \\ 0,4037 + 0,0244\delta_5 \end{bmatrix}, \text{ com } 0 \leq \delta_i \leq 1 \quad (5.5)$$

Este autovetor será a componente principal utilizada na transformação linear. Agora podemos realizar o processo da transformação linear $[\Sigma][v]$ e calcular os valores máximos e mínimos para γ_j e δ_w , conforme realizado nos capítulos anteriores.

Abaixo temos a Tabela 10 com o procedimento realizado e comparando os candidatos no dia 1 e no dia 2:

Tabela 10 – Comparação dos candidatos entre o dia 1 e o dia 2

Candidato	DIA 1	DIA 2	Dif. Limite Inferior	Dif. Limite Superior	Soma
1	[6,7926; 7,6943]	[7,1962; 8,1224]	0,4036	0,4281	0,8317
2	[5,8502; 7,176]	[5,3114; 5,9983]	-0,5388	-1,1777	-1,7165
3	[7,1962; 8,1224]	[7,1962; 9,1844]	0	1,062	1,062
4	[4,1692; 5,9983]	[5,8802; 6,6244]	1,711	0,6261	2,3371
5	[6,2538; 7,0603]	[6,2538; 8,1224]	0	1,0621	1,0621
6	[6,6575; 7,4884]	[6,2538; 8,0322]	-0,4037	0,5438	0,1401
7	[6,725; 8,1224]	[7,1962; 8,1224]	0,4712	0	0,4712
8	[5,4177; 6,5165]	[6,725; 8,5504]	1,3073	2,0339	3,3412
9	[6,6575; 7,4884]	[6,2538; 8,9864]	-0,4037	1,498	1,0943
10	[6,6274; 7,4962]	[6,2538; 7,5786]	-0,3736	0,0824	-0,2912
11	[6,6187; 9,8411]	[6,2538; 7,5786]	-0,3649	-2,2625	-2,6274
12	[6,6575; 7,4884]	[6,2538; 8,006]	-0,4037	0,5176	0,1139
13	[7,0986; 8,04]	[6,2538; 7,5786]	-0,8448	-0,4614	-1,3062
14	[5,8502; 7,4962]	[4,9465; 7,0603]	-0,9037	-0,4359	-1,3396
15	[6,3214; 7,6119]	[6,2538; 7,0603]	-0,0676	-0,5516	-0,6192
16	[5,8502; 6,6322]	[5,3114; 6,542]	-0,5388	-0,0902	-0,629

Fonte: próprio autor.

Através da maior soma, podemos dizer que o candidato 8 foi o que teve maior evolução de um dia para o outro, conforme Tabela 11:

Tabela 11 – Desempenho do Candidato 8

Dia	Candidato	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5
DIA 1	Candidato 8	2	3	[2; 3]	3	3
DIA 2	Candidato 8	[3; 4]	4	3	3	[3; 4]

Fonte: próprio autor.

Veja que no indicador 1 a nota era 2 e agora está variando de 3 a 4, enquanto o indicador 2 aumentou a nota sem intervalos. No indicador 3 a nota variava de 2 a 3 e no segundo dia foi apenas 3 e, enquanto o indicador 4 não teve alteração, o indicador 5, no segundo dia, variou até a nota 4, sendo o limite inferior igual à nota obtida no primeiro dia.

Já o candidato que teve o pior resultado foi o candidato 11, conforme apresentado na Tabela 12:

Tabela 12 – Desempenho do Candidato 11

Dia	Candidato	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5
DIA 1	Candidato 11	[3; 4]	[3; 4]	[4; 5]	[3; 4]	[3; 4]
DIA 2	Candidato 11	3	[3; 4]	3	3	3

Fonte: próprio autor.

O candidato nos indicadores 1, 4 e 5 tinha a nota variando entre 3 e 4, mas no segundo dia a nota fixou no 3. O indicador 2 não houve alteração, porém o indicador 3, que variava de 4 a 5, reduziu para a nota 3, que era o limitante inferior em ambos os indicadores no primeiro dia.

Dessa forma, os resultados obtidos confirmam a eficácia do método CIA na avaliação psicométrica por meio da escala Likert, evidenciando sua aplicabilidade na identificação da evolução dos candidatos ao longo do processo seletivo. Além de validar a abordagem intervalar proposta, esses resultados abrem caminho para a expansão da técnica de Análise de Componentes Principais (PCA) em um contexto difuso, permitindo futuras investigações que aprofundem a modelagem da incerteza e ampliem as aplicações desse método em diferentes domínios de aplicação.

6 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho avaliou quatro propostas de modelagem para matrizes intervalares, visando desenvolver a técnica de Análise de Componentes Principais Intervalar, de modo que a incerteza representada pelo intervalo fosse considerada no processo e não houvesse aumento da dimensão da matriz, como acontece em algumas propostas da literatura.

As quatro propostas resultaram em componentes principais intervalares e não houve aumento da dimensão da matriz de dados ao longo do processo. Ou seja, há estratégias que podem ser aplicadas em casos mais específicos, mas quais das propostas de modelagem são mais interessantes? Vejamos individualmente cada proposta:

- a) **Método Deif**: é um método clássico da literatura e, ao ser avaliado na aplicação teórica, apresentou resultado de autovalor assertivo. Contudo, o método decompõe a matriz em duas outras matrizes e a escolha do resultado depende da matriz de sinais de autovetores de uma das duas matrizes, ou seja, o processo torna-se mais complexo. A necessidade de observar a matriz de sinais de autovetores pode levar a escolha errada e isso é uma problemática. Outro ponto de desvantagem é que a estratégia do método não permite apontar o autovetor associado, sendo necessário um processo analítico e que depende de outra estrutura de modelagem das informações.
- b) **Método Deif com Leverrier-Faddeev**: é a proposta de aprimoramento do método clássico buscando encontrar autovetores de forma mais eficiente. Ao comparar a PCA através deste método e a PCA analítico aplicados nos dados teóricos, percebemos que há pouca diferença, ou seja, essa combinação de métodos apresenta um autovetor relativamente consistente, em relação a versão analítica. Contudo, caso seja necessário avaliar, também, o autovalor, este método não demonstrou assertividade para os dois autovalores.
- c) **Single-level**: a proposta de reescrever os intervalos através da distância entre os limites e multiplicar por uma variável permitiu que métodos numéricos clássicos fossem utilizados para a identificação de autovalores e autovetores intervalares. Ao ser avaliado com dados teóricos, o método apresentou 100% de assertividade para o maior autovalor e não teve resultado esperado para o segundo autovalor. É provável que isso tenha ocorrido, porque o método utiliza uma única variável, enquanto na base teórica não havia nada que tornasse fixo os intervalos, que deveriam ser incrementados na mesma proporção. Então, é possível que este método seja melhor aproveitado em situações que todos os números intervalares devam variar na mesma proporção.

d) *Constraint Interval Arithmetic*: similar ao método anterior, possui a proposta de reescrever os intervalos através da distância entre os limites, mas a multiplicação é feita por uma variável para cada intervalo diferente. Assim como o método anterior, permitiu que métodos numéricos clássicos fossem utilizados para a identificação de autovalores e autovetores intervalares. Ao ser avaliado com dados teóricos, o método apresentou 100% de assertividade para os dois autovalores e foi o que apresentou menor diferença na componente principal. Dessa forma, o CIA mostrou todo o potencial de aplicabilidade, inclusive para matrizes maiores, como apresentado nos testes práticos.

O método CIA é mais do que a proposta para encontrar autovalores e autovetores, é uma metodologia de modelagem que se apresentou essencial para o desenvolvimento das técnicas, porque foi através desta metodologia que conseguimos remodelar as matrizes de informações para obter as matrizes intervalares de covariância. E foi nessas matrizes que os dados foram aplicados.

Sem o método CIA teríamos que abrir mão da incerteza do processo ou aumentar, consideravelmente, a matriz de informações para aplicar algum método clássico da literatura. Essas outras alternativas, em uma sociedade que cada vez usa mais dados, tornam-se inviáveis para os dias atuais, haja vista que há situações que podem chegar ao limite da capacidade tecnológica de processamento.

Assim, é interessante experimentarmos a combinação do método CIA com outras abordagens de incerteza como, por exemplo, a teoria de conjuntos difusos para desenvolver uma técnica de PCA Difusa.

Pensando em escalas psicométricas, pode-se averiguar a possibilidade do desenvolvimento de novas escalas ou métodos de avaliação de escalas que combinem os resultados apresentados neste trabalho com outras áreas da computação flexível como, por exemplo, a própria teoria de conjuntos difusos ou o aprendizado de máquina, visando trazer resultados mais rápidos, eficientes e que considerem a subjetividade e incerteza nas atitudes humanas.

REFERÊNCIAS

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, Hoboken, v. 2, n. 4, p. 433-459, 2010.
- ARIA, M.; CUCCURULLO, C. Bibliometrix: an r-tool for comprehensive science mapping analysis. **Journal of Informetrics**, [s. l.], v. 11, n. 4, p. 959-975, 2017.
Disponível em:
<https://doi.org/10.1016/j.joi.2017.08.007>. Acesso em: 7 maio 2024
- BASHA, N.; NOUNOU, M.; NOUNOU, H. Multivariate fault detection and classification using interval principal component analysis. **Journal of Computational Science**, [s. l.], v. 27, p. 1-9, 2018.
- CAZES, P.; DOUZAL, A.; DIDAY, E.; SCHEKTMAN, Y. Extensions de l'analyse en composantes principales à des données de type intervalle. **Revue de Statistique Appliquée**, [s. l.], v. 45, n. 3, p. 5-24, 1997.
- CHAPIN, J. K.; NICOLELIS, M. A. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. **Journal of Neuroscience Methods**, Amsterdam, v. 94, n. 1, p. 121-140, 1999.
- DEIF, A. The interval eigenvalue problem. **Journal of Applied Mathematics and Mechanics**, [s. l.], v. 71, n. 1, p. 61-64, 1991.
- FRANCO, N. B. Cálculo numérico. [S. l.]: Pearson Prentice Hall, 2006.
- GIORDANI, P.; KIERS, H. A comparison of three methods for principal component analysis of fuzzy interval data. **Computational Statistics & Data Analysis**, [s. l.], v. 51, p. 379-397, nov. 2006.
- IZEM, T. A.; BOUGHELOUM, W.; HARKAT, M. F.; DJEGHABA, M. Fault detection and isolation using interval principal component analysis methods. **IFAC-PapersOnLine**, [s. l.], v. 48, n. 21, p. 1402-1407, 2015.
- JOLLIFFE, I. Principal component analysis. 2nd ed. [S. l.]: Springer, 2013. (Springer Series in Statistics).
- KAPLAN, R.; SACCUZZO, D. Psychological testing: principles, applications, and issues. [S. l.]: Cengage Learning, 2012.
- LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, United States, v. 22, n. 140, p. 1-55, 1932.
- LODWICK, W. Special issue: interfaces between fuzzy set theory and interval analysis. [S. l.]: Elsevier, 2003. (Fuzzy Sets and Systems).
- MAITI, S.; CHAKRAVERTY, S. Evolution of interval eigenvalue problems and its applications to the uncertain dynamic problems. **Archives of Computational Methods in Engineering**, Dordrecht, v. 30, n. 2, p. 1167-1186, 2023.
- MIZUKOSHI, M. T.; LODWICK, W. A. The interval eigenvalue problem using constraint interval analysis with an application to linear differential equations. **Fuzzy Sets and Systems**, [s. l.], v. 419, p. 141-157, 2021.
- MONTE, L. G. Escala Likert difusa: um estudo sobre diferentes abordagens. 2020. 44 f. Trabalho de Conclusão de Curso (Graduação em Matemática Industrial) - Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2020.

NAKAGAWA, E.; SCANNAVINO, K.; FABBRI, S.; FERRARI, F. Revisão sistemática da literatura em engenharia de software: teoria e prática. [S. l.]: Elsevier, 2017.

PALUMBO, F.; LAURO, C. A PCA for interval valued data based on midpoints and RADII. *In: OKADA, A. et al. (org.). New Developments in Psychometrics. [S. l.]: Springer, 2003. p. 641-648.*

SANTOS, M. R. P.; ALCANTARA, L. O.; GOMES, R. B.; BUARQUE, H. L. B. Uso de estatística multivariada como ferramenta na avaliação da qualidade da água da lagoa da Maraponga, Fortaleza, Ceará, Brasil. *In: CONGRESSO BRASILEIRO DE ENGENHARIA SANITÁRIA E AMBIENTAL, 26., 2011, Porto Alegre. Anais [...]. Porto Alegre: ABES, 2011.*

SERRAO, R. G.; OLIVEIRA, M. R.; OLIVEIRA, L. Theoretical derivation of interval principal component analysis. **Information Sciences**, [s. l.], v. 621, p. 227-247, 2023.

SHEVKY, E.; WILLIAMS, M.; HAYNES, J. R.; FOUNDATION, D. H. The social areas of Los Angeles: analysis and typology. [S. l.]: John Randolph Haynes and Dora Haynes Foundation, 1949.

STEPHENS, M.; NOVEMBRE, J. Interpreting principal component analyses of spatial population genetic variation. **Nature Genetics**, London, v. 40, n. 5, p. 646-649, 2008.

ZHANG, S.; WANG, S. Spectral radius-based interval principal component analysis (sr-ipc) for fault detection in industrial processes with imprecise data. **Journal of Process Control**, [s. l.], v. 114, p. 105-119, 2022.