



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

DISSERTAÇÃO DE MESTRADO

**Uso da Técnica de Análise de Componentes Principais
na Redefinição do Parâmetro BLA**

Carlos Alberto Moreira de Melo Neto

Orientador: Prof. Dr. Demétrio Antônio da Silva Filho

Brasília-Distrito Federal

2016

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Uso da Técnica de Análise de Componentes Principais na Redefinição do Parâmetro BLA

Carlos Alberto Moreira de Melo Neto

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade de Brasília (PPGF-UnB) como parte dos requisitos necessários para obtenção do título de Mestre em Ciências (Física).

Orientador: Prof. Dr. Demétrio Antônio da Silva Filho

Brasília-Distrito Federal

2016

*“Dedico este trabalho aos meus pais,
por todo o apoio crucial para
que eu chegasse até aqui.”*

Agradecimentos

Agradeço primeiramente aos meus pais que sempre me incentivaram e me deram apoio em todas as escolhas que fiz, sem eles não teria terminado esta dissertação. Ao meu irmão que sempre me apoiou e me deu apoio emocional mesmo a distância.

Um agradecimento especial também à minha namorada, Daísma, que também foi fundamental em muitos momentos difíceis e me deu muito apoio durante esses anos em Brasília.

Ao professor Demétrio pela orientação que, me deu boa condição para a conclusão desta dissertação e, sempre paciente, me guiou de forma fundamental para a finalização deste trabalho.

Ao professor Kent Mann, da Universidade de Illinois por sugerir o projeto para esta dissertação.

Aos colegas do grupo, Ana, Akira, Leonardo e Patrícia, por contribuir com minha formação através de discussões feitas.

Aos colegas da pós do instituto de física pela companhia e por várias discussões produtivas, principalmente os colegas do conjunto um, com quem tenho contato todos os dias.

Ao amigo Valdeir pela amizade e por várias discussões sobre física e sobre vários outros assuntos.

Ao amigo Thiago pelo apoio quando cheguei a Brasília e pela amizade e boa convivência.

Aos funcionários da secretaria da pós-graduação do Instituto de Física da UnB e a todos os professores com os quais tive contato durante este período e, desde já, agradeço aos professores Ricardo Gargano e Heibbe de Oliveira por aceitarem o convite para participar de minha banca.

Carlos.

*Uma experiência nunca é um fracasso,
pois sempre vem demonstrar algo.
Thomas Alva Edison*

Resumo

A alternância dos comprimentos de ligação (*Bond Length Alternation, BLA*) em cadeias moleculares conjugadas longas tem sido tópico de discussões por muitas décadas, tanto experimental quanto teoricamente. O *BLA* é um parâmetro estrutural que vem da diferença entre o comprimento de ligações duplas e simples ao longo da cadeia conjugada.

Neste trabalho nós utilizamos um método estatístico muito utilizado em várias áreas do conhecimento, a Análise dos Componentes Principais (do inglês *Principal Components Analysis - PCA*), criado por *Karl Pearson* em 1901, para redefinir o parâmetro *BLA*. Com esta técnica calculamos o *BLA* de um grupo de 17 moléculas derivadas de tertiofeno e, a partir disto, fizemos comparações com a forma mais comum de calcular o *BLA*, o que nos forneceu uma forma alternativa de se calcular este parâmetro. Com a *PCA* podemos ir além ao cálculo do parâmetro estrutural e, por exemplo, podemos identificar quais ligações têm maior relevância para o valor do *BLA*. Outro ponto desta análise que merece destaque, foi mostrar a evolução dos coeficientes gerados por esta técnica, chamados de PC_1 , assim como, a evolução dos valores do *BLA* com o aumento do número de moléculas presentes no grupo inicial para o cálculo e a diferença em relação a aromaticidade de tais moléculas, mostrando qual o número mínimo para que esta técnica possa ser aplicada.

Palavras chave: PCA, BLA, Componentes principais, Comprimento de ligação.

Abstract

The bond length alternation (BLA) in long conjugated molecular chains have been the topic of discussion for many decades, both experimentally and theoretically. The BLA is a structural parameter that comes from the difference between the length of double and single bonds along the conjugated chain.

In this work we have used a statistical method widely used in many fields of knowledge, the Principal Component Analysis, created by Karl Pearson in 1901, to re-define the parameter BLA. With this technique we calculated the BLA of a group of 17 molecules derived from terthiophene and made comparisons with the most common way to calculate the BLA, which provided us with an alternative way to calculate the parameter. With the PCA we can go further the BLA and we can also identify which bonds are more relevant to its value. Another point that is worth mentioning, was showing the evolution of the coefficients generated by this technique, as well as the evolution of BLA values, along with the increasing number of molecules present in the initial group for the calculation and the differences between the aromaticity of such molecules, resulting in the minimum number to apply this technique.

Keywords: PCA, BLA, Principal components, bond lengths.

Lista de Figuras

1	Exemplo de tertiofeno.	10
1.1	Gráfico com os pontos aleatórios de duas dimensões utilizados	14
1.2	Gráfico dos pontos centralizados na média x' e y'	16
1.3	Gráfico dos pontos ajustados x' e y' com a direção dos autovetores	17
1.4	Gráfico dos dados finais para o vetor característico com os dois autovetores .	20
1.5	Gráfico dos dados finais para o vetor característico associado ao maior autovalor	21
2.1	Exemplo de polieno com BLA 0,08 Å.	25
4.1	Moléculas: (1) - $C_{20}H_{10}S_3N_4$ e (2) - $C_{20}H_{10}S_3N_4O_2$	35
4.2	Moléculas: (3) - $C_{30}H_{26}S_3O_2$, (4) - $C_{24}H_{26}S_3O_4$, (5) - $C_{20}H_{14}S_3N_4$, (6) - $C_{16}H_{14}S_3$ e (7) - $C_{14}H_{12}S_3$	36
4.3	Moléculas: (8) - $C_{20}H_8S_3N_4O_2$, (9) - $C_{14}H_{12}S_3O_2$, (10) - $C_{40}H_{30}S_3$, (11) - $C_{40}H_{26}S_3$ e (12) - $C_{24}H_{14}S_3O_6N_2$	37
4.4	Moléculas: (13) - $C_{24}H_{18}S_3O_2N_2$, (14) - $C_{20}H_{18}S_3O_6$, (15) - $C_{30}H_{28}S_3O_2$, (16) - $C_{14}H_{10}S_3O_4N_2$ e (17) - $C_{24}H_{10}S_3N_6$	38
4.5	Gráfico do BLA e do BLA_{PCA}	44
4.6	Gráfico do BLA e do BLA_{PCA} dividido pelo fator de escala	44
4.7	Padrão de ligações de uma molécula (1) aromáticas e (2) quinóides. . .	45
4.8	BLA_{PCA} calculado através do método da PCA	46
4.9	Importância das ligações das moléculas aromáticas.	47
4.10	Importância das ligações das moléculas quinóides.	47
4.11	PC_1 para o grupo gradual	48
4.12	PC_1 para o grupo de quinóides para aromáticas.	49
4.13	PC_1 para o grupo de aromáticas para quinóides.	49
4.14	PC_1 para o grupo alternado.	50

4.15 <i>BLA</i> da molécula 1 ordem aleatória inicial	51
4.16 <i>BLA</i> da molécula 1 crescendo de quinóides para aromáticas	51
4.17 <i>BLA</i> da molécula 1 crescendo de aromáticas para quinóides	53
4.18 <i>BLA</i> da molécula 1 crescendo de forma alternada	53
4.19 Exemplo de ftalocianina	55

Lista de Tabelas

1.1	Conjunto de dados aleatórios de duas dimensões	14
1.2	Dados centralizados na média	15
1.3	Dados finais para o vetor característico com os dois autovetores	19
1.4	Conjunto de dados finais com redução de dimensão (apenas um autovetor)	21
4.1	Comprimentos de ligação (CL) carbono-carbono (em Å) via <i>B3LYP/6-31G**</i>	39
4.2	Comprimentos de ligação (CL) carbono-enxofre (em Å) via <i>B3LYP/6-31G**</i>	40
4.3	Coeficientes do BLA_{PCA} (PC_1 e primeira linha da matriz C	41
4.4	Autovalores e seus pesos proporcionais	42
4.5	Coeficientes da PC_1	42
4.6	Tabela comparativa entre BLA e BLA_{PCA} em Å	43

Sumário

Introdução	10
1 Análise de Componentes Principais (PCA)	12
1.1 Exemplo Prático de Implementação da <i>PCA</i>	13
1.1.1 A Matriz Covariância e Seus Autovalores e Autovetores	15
1.1.2 O Vetor Característico e a Formação do Novo Conjunto de Dados	17
1.2 Aplicações	22
1.2.1 Exemplos de aplicação	23
2 Alternância dos comprimentos de ligação (BLA)	25
3 Metodologia	29
3.1 Teoria do Funcional da Densidade (<i>DFT</i>)	29
3.1.1 Aproximações para o Termo de Troca e Correlação	30
3.2 Encontrando a <i>PCA</i>	32
3.2.1 Calculando o BLA_{PCA}	34
4 Resultados	35
4.1 Obtenção do <i>BLA</i>	39
4.2 Aromaticidade	45
4.3 Número mínimo para um <i>PCA</i> funcional	46
Conclusão e Perspectivas	52
A Código do <i>Mathematica</i> da <i>BLA-PCA</i>	57
Referências Bibliográficas	60

Introdução

A *PCA* é um método estatístico criado por *Karl Pearson* em 1901 [1] muito utilizado para filtrar ruídos e diminuir a dimensão de um conjunto de dados. Esta técnica utiliza combinações lineares de um grande conjunto de dados, tornando-os linearmente descorrelacionados.

A *PCA* tem sido amplamente utilizada para analisar dados experimentais, como por exemplo, investigar experimentos de cristalização [2], analisar narcóticos em misturas sólidas [3,4], reduzir erro de espectro *Raman* [5], investigar atividade do conjunto neuronal [6], modelar explicitamente as diferenças na ascendência entre casos e controles [7], entre outros.

Neste trabalho temos como objetivo realizar uma *PCA* em um conjunto de moléculas derivadas de tertiofenos, que são três tiofenos ligados entre si (ver Figura 1), a fim de encontrar um parâmetro estrutural que depende apenas dos comprimentos de ligação destas moléculas.

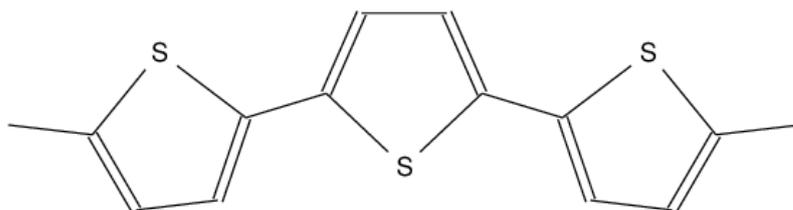


Figura 1: Exemplo de tertiofeno.

Os tiofenos e seus derivados têm sido estudados extensivamente desde quando as propriedades do poliacetileno dopado foram descobertas e foram previstas as primeiras aplicações para polímeros conjugados. Sistemas de tiofenos são atualmente responsáveis por muitos destaques tecnológicos em uma variedade de tecnologias eletrônicas de “plástico”. Exemplos destes sistemas incluem sensores, células sola-

res orgânicas, diodos emissores de luz orgânicos (*OLED's*) e transistores de efeito de campo orgânicos (*OFET's*). Devido a facilidade de funcionalização e grande versatilidade, os derivados de tiofeno devem ser objeto de estudo de muitas pesquisas por muitos anos.

Com o parâmetro calculado com a *PCA*, utilizando os comprimentos de ligação deste grupo de moléculas, faremos uma comparação com um parâmetro bastante conhecido, o *BLA*, já que ambos os parâmetros dependem apenas de tais comprimentos.

O *BLA* é um parâmetro que vem da diferença no comprimento das ligações duplas e simples em uma cadeia carbônica. Foi estabelecida uma relação entre o *BLA* e a hiper polarizabilidade eletrônica em moléculas orgânicas [8] e, com isso, pode ser feita uma avaliação da quantidade de polarização de uma molécula através deste parâmetro.

No primeiro capítulo, temos uma explanação do método de análise dos dados, a *PCA*, um método de redução de variáveis bastante conhecido e utilizado em várias áreas do conhecimento que tem como forte característica a redução de redundância nos dados, onde poderemos entender como funciona e como utilizar este método de forma genérica.

No segundo capítulo, veremos uma apresentação do *BLA*, a cronologia de seu estudo e sua importância. A seguir, no Capítulo 3, temos uma breve introdução a tópicos de estrutura eletrônica para ilustrar melhor as técnicas utilizadas para os cálculos dos comprimentos de ligação. Também neste capítulo, mostraremos como o método foi empregado para o caso proposto nesta dissertação.

Por fim, no Capítulo 4 veremos os resultados da análise aqui apresentada, assim como uma comparação dos resultados obtidos com o *BLA*, tal qual iremos explorar outras características interessantes deste método quando aplicado a encontrar um parâmetro estrutural e também discutiremos suas capacidades e limitações para este tipo de cálculo.

Capítulo 1

Análise de Componentes Principais (PCA)

A PCA é um procedimento matemático que utiliza uma transformação ortogonal para converter um conjunto de variáveis possivelmente correlacionadas em um conjunto de vetores de variáveis linearmente descorrelacionadas, chamadas componentes principais (*principal components - PC's*).

A PCA é um método estatístico de redução de variáveis que é útil quando obtém-se dados de um grande número de variáveis e acredita-se ter alguma redundância nessas variáveis. Neste caso, redundância significa que algumas das variáveis estão correlacionadas com outras, possivelmente por elas estarem medindo o mesmo objeto.

Por causa desta redundância, acredita-se que deve ser possível reduzir as variáveis observadas em um número menor de componentes principais (variáveis artificiais) que irão representar a maior parte da variância nas variáveis observadas. As componentes principais podem então ser usadas como variáveis de previsão ou de critério nas análises seguintes.

Entre as muitas técnicas analíticas, a PCA é a mais comumente utilizada pois, é o ponto inicial no processo de mineração de dados que visa minimizar a dimensão dos dados. Entretanto, é comum lidar com muitos dados onde um grupo de n objetos é descrito por um número de p variáveis. Os dados são agrupados em uma matriz X com n linhas e p (ver Equação 1.1), com elementos x_{ij} referindo-se a um elemento de

X na i -ésima linha da j -ésima coluna.

Comumente, as linhas de X correspondem à uma observação, pode ser um espectro ou, de forma mais geral, uma curva analítica, obtida da análise de uma amostra real, medida com instrumento produzindo curvas analíticas como dados de saída e uma coluna de X é geralmente uma variável dessas medidas.

$$X = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}}_{\text{Variáveis}} \left. \vphantom{\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix}} \right\} \text{Objetos} \quad (1.1)$$

No que diz respeito ao tipo de análise que nos interessa, somos comumente confrontados com dados multidimensionais $n \times p$, onde n e p são da ordem de algumas centenas ou até milhares. Nessas situações fica difícil identificar, em um grupo de dados tão grande, qualquer informação relevante sem a ajuda de uma técnica matemática como a *PCA*.

1.1 Exemplo Prático de Implementação da *PCA*

Tecnicamente, componentes principais podem ser definidas como uma combinação linear de variáveis observadas otimamente ponderadas. Para entendermos o significado desta definição, descreveremos como os valores dos indivíduos em componentes principais são calculados.

Primeiramente, temos que selecionar um grupo de dados (para uma melhor visualização utilizaremos usar um grupo de dados aleatórios de apenas duas dimensões, ver Tabela 1.1 e Figura 1.1).

Em seguida, devemos tirar a média e subtraí-la de cada item do grupo, ou seja, calcular a média de todos os valores de x e de y , obtendo um \bar{x} e um \bar{y} , que serão subtraídos de cada valor de x e de y respectivamente, resultando em $x' = x - \bar{x}$ e $y' = y - \bar{y}$ como mostrado na Tabela 1.2.

Tabela 1.1: Conjunto de dados aleatórios de duas dimensões

x	y
4,1	4,7
1,6	2,5
3,2	3,4
2,2	2,4
1,2	1,7
5,7	5,6
5,2	4,6
3,3	4,5
2,8	2,9
1,6	2,0

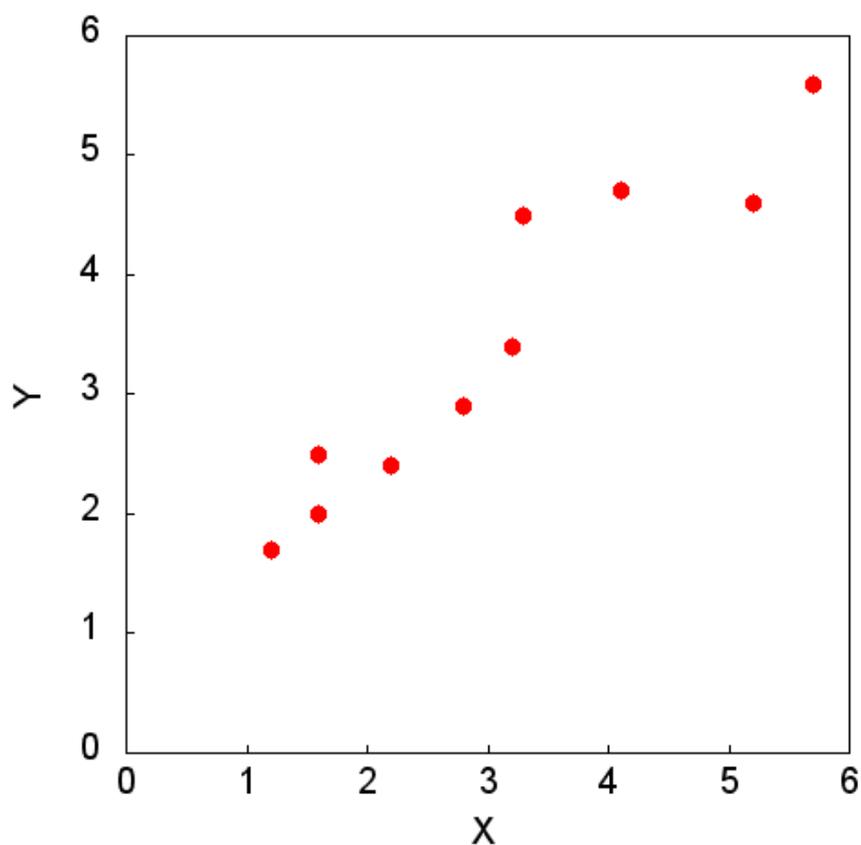


Figura 1.1: Gráfico com os pontos aleatórios de duas dimensões utilizados

Tabela 1.2: Dados centralizados na média

x'	y'
1,01	1,27
-1,49	-0,93
0,11	-0,03
-0,89	-1,03
-1,89	-1,73
2,61	2,17
2,11	1,17
0,21	1,07
-0,29	-0,53
-1,49	-1,43

Este procedimento produz um grupo de dados cuja média é zero e estão dispostos no gráfico da Figura 1.2 onde, podemos ver que em comparação com a Figura 1.1, este procedimento centraliza os dados na origem.

1.1.1 A Matriz Covariância e Seus Autovalores e Autovetores

O próximo passo é calcular a matriz covariância, que tem a seguinte forma:

$$C = \begin{pmatrix} cov(x, x) & cov(x, y) \\ cov(y, x) & cov(y, y) \end{pmatrix}, \quad (1.2)$$

onde, $cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$, com isso, para este exemplo temos a seguinte matriz covariância:

$$C = \begin{pmatrix} 2,36 & 1,94 \\ 1,94 & 1,79 \end{pmatrix}. \quad (1.3)$$

Então, como os elementos não diagonais nesta matriz covariância são positivos, podemos esperar que x e y cresçam juntos.

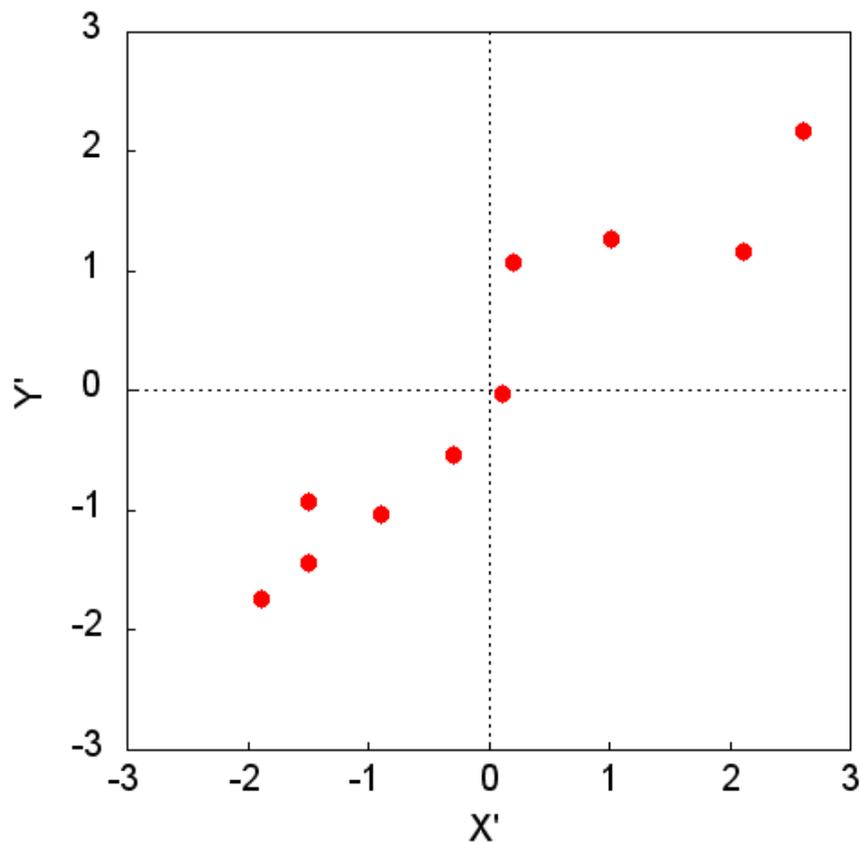


Figura 1.2: Gráfico dos pontos centralizados na média x' e y'

Calculando os autovalores e autovetores de C , temos:

$$\text{Autovalores} = \begin{pmatrix} 4,04 \\ 0,11 \end{pmatrix}; \quad (1.4)$$

$$\text{Autovetores} = \begin{pmatrix} 0,76 & 0,65 \\ -0,65 & 0,76 \end{pmatrix} \quad (1.5)$$

Como esperado da matriz covariância, podemos verificar que as duas variáveis crescem realmente juntas e nos fornecem informações sobre o padrão dos dados através de seus autovetores. No gráfico da Figura 1.3, vemos que o primeiro autovetor nos mostra como os dois conjuntos de dados estão relacionados ao longo da linha e, ela se encaixa nos pontos como um ajuste da curva, o último autovetor nos mostra outro padrão dos dados, que eles seguem a linha principal, porém, estão deslocados da mesma por uma dada quantia.

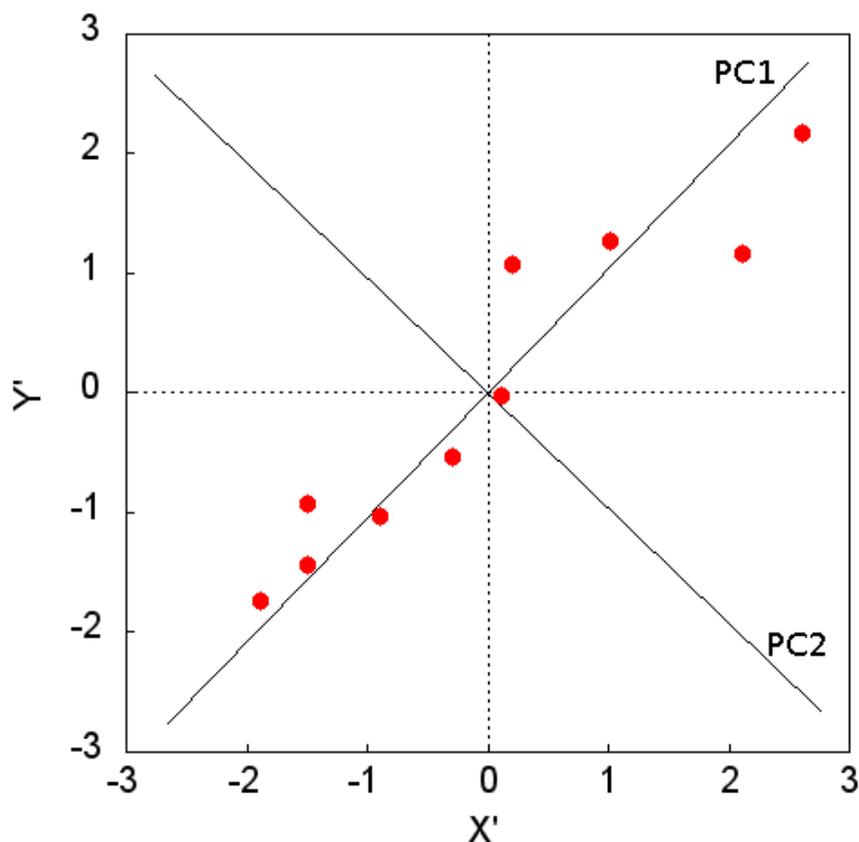


Figura 1.3: Gráfico dos pontos ajustados x' e y' com a direção dos autovetores

Então, através deste processo de calcular os autovetores da matriz covariância, fomos capazes de extrair linhas que caracterizam os dados, nos passos seguintes iremos transformar os dados de forma a expressá-los nos termos dos autovetores.

1.1.2 O Vetor Característico e a Formação do Novo Conjunto de Dados

Agora veremos onde a noção de compressão de dados e redução de dimensionalidade aparecem. Olhando para os autovetores, verificamos que seus respectivos autovalores são bem diferentes, o autovetor com o maior autovalor é o primeiro componente principal (PC_1) do nosso conjunto de dados, este é o vetor que passa pelo meio dos pontos no gráfico da Figura 1.3 e é a mais significativa relação entre as dimensões dos dados.

Ao encontrarmos os autovetores da matriz covariância devemos ordená-los por autovalor, do maior para o menor, assim teremos as *PC's* por ordem de significância. Com isto, podemos ignorar ou não as componentes de menor variância. Escolhendo descartar algum autovetor, perderemos um pouco de informação, entretanto, se os autovalores forem pequenos a informação não é muito significativa. Esta é das características da *PCA*, a técnica sempre coloca a maior parte da variância possível nas primeiras *PC's*

Deixando de lado algumas componentes, o conjunto final de dados será menor que o original, à rigor, se originalmente temos um conjunto de dados com n dimensões, iremos encontrar n autovalores e n autovetores. Escolhendo apenas os autovetores com a maior parte da variância, ou seja, escolhendo apenas os autovetores associados aos maiores autovalores, estaremos diminuindo a dimensão dos dados.

Feito isso, devemos montar um *vetor característico* (VC), que é uma matriz de vetores. Ele é composto pelos autovetores que escolhermos não ignorar do total calculado a partir da matriz covariância, onde cada um deles será uma coluna desta matriz.

$$VC = (avec_1 \quad avec_2 \quad avec_3 \quad \cdots \quad avec_n) \quad (1.6)$$

onde cada $avec_i$ é uma coluna da matriz com um autovetor.

Para o nosso exemplo temos apenas dois autovetores, então, poderemos formar um vetor característico com apenas um ou com os dois vetores.

Com os dois:

$$VC_1 = \begin{pmatrix} 0,76 & 0,65 \\ 0,65 & -0,76 \end{pmatrix} \quad (1.7)$$

ou com um:

$$VC_2 = \begin{pmatrix} 0,76 \\ 0,65 \end{pmatrix} \quad (1.8)$$

Uma vez que escolhermos os componentes principais (autovetores) mais significantes e montarmos o vetor característico, iremos multiplicar os dados originais ajustados (DA) por este vetor característico:

$$DF = DA \times VC \quad (1.9)$$

onde DF são os dados finais.

Com isso, teremos os dados originais em função apenas dos vetores escolhidos. Nos nossos dados originais temos dois eixos, x e y , então, os dados finais estarão em função de no máximo dois autovetores. É possível expressá-los em termos de quaisquer dois eixos que escolhermos. Se os eixos forem perpendiculares a expressão é mais eficiente, por este motivo é importante os autovetores serem ortogonais entre si.

Dessa forma, podemos expressar os dados em função de quantos autovetores decidirmos utilizar. No nosso exemplo temos duas opções, podemos utilizar os dois autovetores para o vetor característico, ou podemos escolher apenas o autovetor de maior variância.

Veremos agora os dois casos possíveis para o nosso exemplo, com o vetor característico formado com ambos os autovetores e formado apenas pelo de maior relevância. Fazendo a transposta do resultado teremos os dados finais na forma de uma Tabela x e y , como podemos verificar na Tabela 1.3.

Tabela 1.3: Dados finais para o vetor característico com os dois autovetores

x	y
1,59	0,30
-1,74	0,27
0,06	-0,09
-1,35	-0,20
-2,56	-0,07
3,39	-0,06
2,36	-0,49
0,86	0,67
-0,57	-0,21
-2,06	-0,11

No caso de usar os dois autovetores para o vetor característico podemos ver na Figura 1.4, que este processo nos dá basicamente os dados originais apenas com

uma rotação, que neste caso foi em torno de 40° , fazendo com que os autovetores agora, sejam os eixos do nosso sistema. O que é compreensível já que não perdemos nenhuma informação com esta escolha.

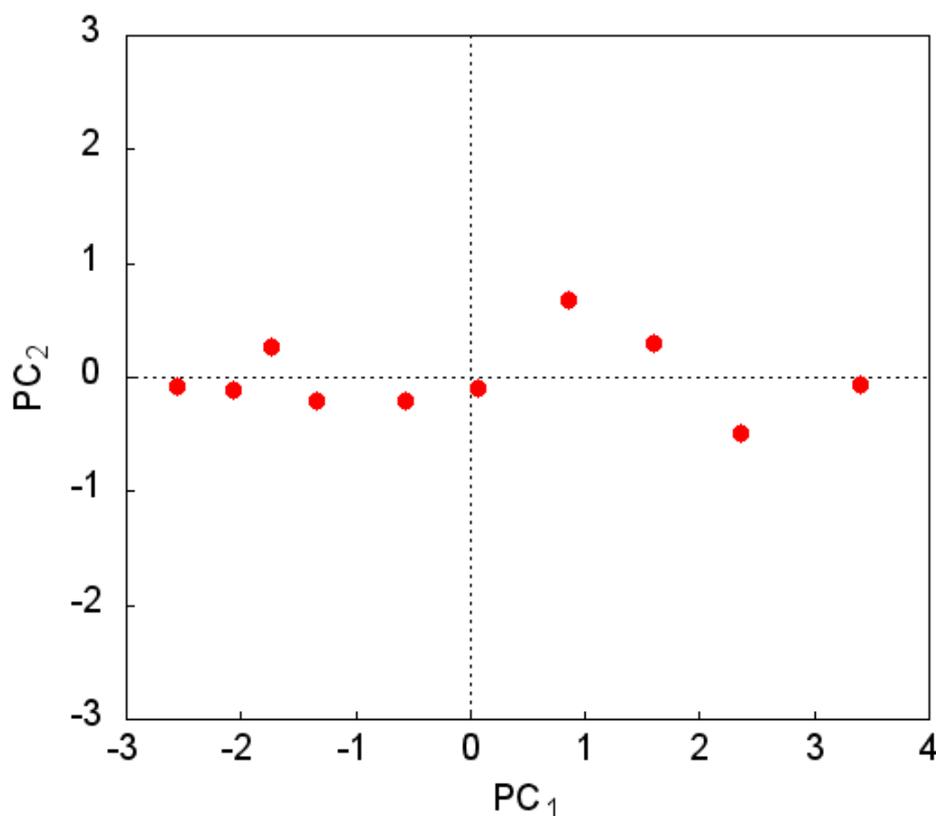


Figura 1.4: Gráfico dos dados finais para o vetor característico com os dois autovetores

Outra maneira de obter este conjunto de dados é escolher apenas o autovetor com o maior autovalor associado, por este motivo, temos apenas uma dimensão como mostrado na Tabela 1.4

Como nosso resultado tem apenas uma dimensão, todos os pontos estarão na mesma linha e, comparando com os dados obtidos para a transformação com os dois autovetores, podemos verificar que este grupo de pontos é exatamente a primeira coluna do primeiro grupo obtido, excluindo o outro eixo, ou seja, o outro autovetor. Assim, teremos um gráfico apenas em função do autovetor escolhido como vemos na Figura 1.5.

Após todo este processo temos uma transformação dos dados iniciais, resultando em um novo grupo de dados em função dos seus padrões, estes padrões são as

Tabela 1.4: Conjunto de dados finais com redução de dimensão (apenas um autovetor)

x
1,59
-1,74
0,06
-1,35
-2,56
3,39
2,36
0,86
-0,57
-2,06

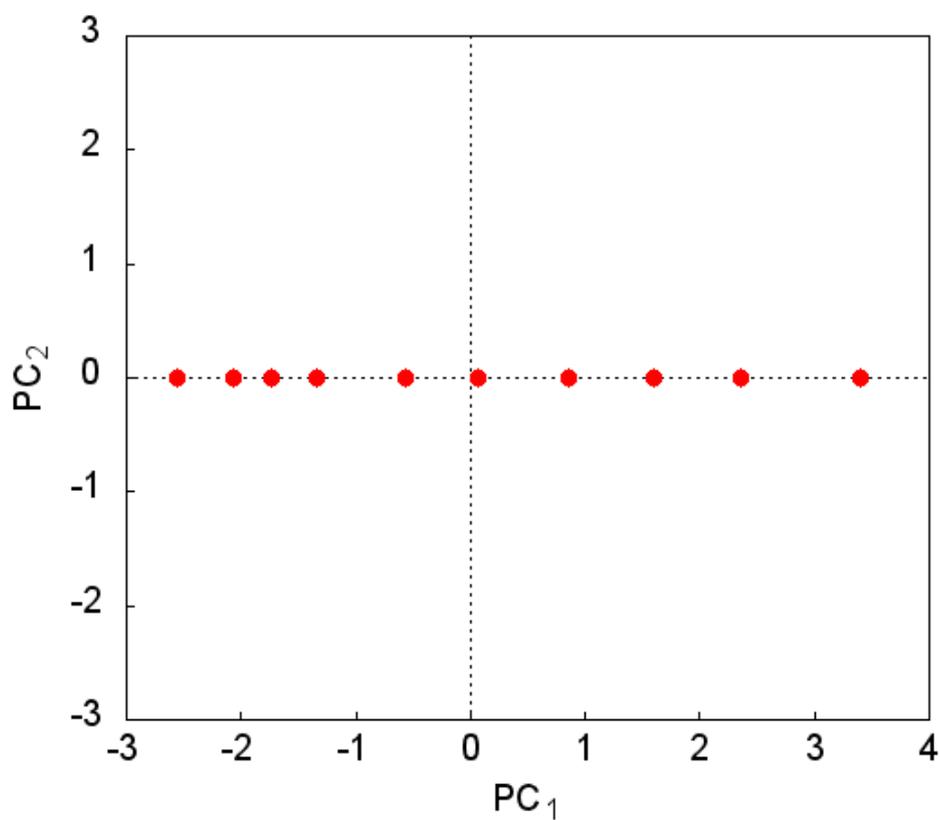


Figura 1.5: Gráfico dos dados finais para o vetor característico associado ao maior autovalor

linhas que descrevem a relação entre eles (os autovetores), com isso, os dados iniciais que, anteriormente não eram correlacionados, são agora uma combinação linear das contribuições de cada uma dessas linhas da matriz inicial.

No caso da transformação com os dois autovetores, apenas alteramos os dados para que eles sejam expressos em termos dos autovetores e não dos eixos iniciais. Este processo como foi apresentado, foi na verdade uma rotação, porém, com apenas um autovetor, removemos a influência do autovetor de menor significância, o que nos deixou com dados em função apenas do outro que está relacionado ao maior autovalor.

No caso de duas dimensões, a análise de componentes principais parece simples e pode ser visualizada sem dificuldades. No entanto, os problemas em que esta análise se aplica, são raramente bidimensionais. Na verdade, esta técnica pode ser aplicada por quantas dimensões forem possíveis de se diagonalizar a matriz correspondente e este número hoje chega à centenas de milhares. Esta técnica pode reduzir várias destas dimensões a um pequeno conjunto delas, facilitando a análise dos dados.

1.2 Aplicações

O principal objetivo da *PCA* é revelar informações escondidas num grupo de dados, com isso, podemos usá-lo para vários fins, como:

- Identificar como variáveis diferentes trabalham juntas para criar a dinâmica do sistema;
- Reduzir a dimensionalidade dos dados;
- Diminuir a redundância nos dados;
- Filtrar ruídos nos dados;
- Simplificar os dados para análises posteriores.

É tarefa difícil dar uma descrição completa de quando a *PCA* é utilizável, pois a mesma tem sido usada em incontáveis aplicações estatísticas. Para se ter uma ideia, seguem alguns exemplos de aplicações deste método.

1.2.1 Exemplos de aplicação

Usando este tipo de técnica matricial em visão computacional, precisamos considerar representação de imagens na forma de uma matriz. Uma imagem quadrada $N \times N$ pode ser representada como um vetor de dimensão N^2 :

$$X = \left(x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_{N^2} \right) \quad (1.10)$$

onde as linhas de *pixels* na imagem são posicionadas uma depois da outra para formar uma imagem unidimensional, ou seja, os primeiros N elementos $x_1 \rightarrow x_N$ serão a primeira linha da imagem, os próximos N elementos são a próxima coluna e assim por diante. Os valores no vetor são os valores de intensidade da imagem, possivelmente, um único valor de escala de cinzas.

Com isso, podemos usar a *PCA* para encontrar padrões, digamos que temos vinte imagens. Cada imagem tem N *pixels* de altura e N *pixels* de largura. Para cada imagem podemos criar um vetor imagem como descrito anteriormente, assim, podemos montar com estes vetores uma matriz-imagem grande da seguinte forma:

$$MatrizdeImagens = \begin{pmatrix} VetorImagem_1 \\ VetorImagem_2 \\ \vdots \\ VetorImagem_{20} \end{pmatrix} \quad (1.11)$$

o que nos dá o ponto inicial da nossa análise de componentes principais.

Resolvendo a *PCA*, teremos nossos dados originais em termos dos autovetores da matriz covariância. A utilidade deste procedimento é, por exemplo, reconhecimento facial, onde nossas imagens originais são de rostos de pessoas. Então, o problema é, dado uma nova imagem saber qual rosto do grupo inicial ela seria. Note que a nova imagem não é uma das vinte que tínhamos inicialmente. A forma de proceder é medir a diferença entre os *pixels* da nova imagem e das imagens do grupo inicial (do banco de dados por exemplo), porém, não ao longo do eixo original e sim dos eixos derivados da *PCA*.

Temos que estes eixos funcionam bem melhor para reconhecer faces pois, a *PCA* nos deu as imagens originais em termos das diferenças e das similaridades entre elas, o método identificou os padrões estatístico nos dados.

Como todos os vetores têm dimensão N^2 , iremos ter N^2 autovetores, na prática, podemos descartar alguns autovetores menos significantes e o reconhecimento ainda funciona bem.

A *PCA* também pode ser usada pra compressão de imagens, também conhecida como transformada de *Hotelling* [9] ou transformada de *Karhunen e Leove* (KL) [10–12]. Se temos vinte imagens, cada uma com N^2 *pixels*, formaremos N^2 vetores com vinte dimensões cada. Cada vector é formado por todos os valores de intensidade do mesmo *pixel* de cada imagem. Isto é diferente do exemplo anterior porque nós tínhamos um vetor para cada imagem e, cada item neste vetor era um *pixel* diferente, já agora, temos um vetor para cada *pixel* e, cada item do vetor é de uma imagem diferente.

Feito isso, podemos fazer a *PCA* nestes dados. Teremos vinte autovetores, já que cada vetor tem vinte dimensões e, para comprimir os dados podemos escolher transformar os dados apenas com os quinze mais relevantes, o que nos dará um grupo final de dados com quinze dimensões, o que irá economizar espaço, porém, quando quisermos reproduzir as imagens originais, elas terão perdido parte da informação, a imagem restaurada não é exatamente a mesma, em geral, ela perde qualidade.

Outro exemplo que podemos citar utiliza a capacidade de redução de dimensionalidade para de filtrar ruídos em um grupo de dados experimentais muito grande, como por exemplo um espectro, já que é comum que a soma das variâncias das primeira *PCs* excedam 80% da variância total.

R.Y. Sato-Berr e colaboradores por exemplo, utilizaram a *PCA* para determinar a pressão de transição de fase de espectro *Raman*.

No próximo capítulo iremos apresentar o *BLA*, discutiremos sua importância e veremos a cronologia de seu estudo.

Capítulo 2

Alternância dos comprimentos de ligação (*BLA*)

A alternância dos comprimentos de ligação (*Bond Length Alternation, BLA*) em cadeias moleculares conjugadas longas tem sido tópico de discussões por muitas décadas, tanto experimental quanto teoricamente [13–17].

Exemplos típicos de sistemas onde a relação íntima entre *BLA* e as propriedades eletrônicas, que tem sido amplamente demonstrada, são os polienos e os poliacetilenos. Na aproximação de um elétron e na ausência do *BLA*, o *bandgap* de polienos lineares desapareceriam no limite de cadeias longas, isto é, o poliacetileno seria um metal. Porém, um efeito pseudo-*Jahn-Teller* conhecido como distorção *Peierls* [18] acontece, produzindo um *BLA* significativa de aproximadamente 0,08 Å e um *bandgap* não nulo da ordem de 1,5 - 1,8 eV [19, 20]. O polieno é uma série de unidades de CH conectadas por ligações duplas e simples como mostra a Figura 2.1.



Figura 2.1: Exemplo de polieno com *BLA* 0,08 Å.

Por outro lado, no caso da cianina, que são íons de camada fechada com um número ímpar de carbonos conjugados ao longo da cadeia principal, espera-se que

o *BLA* desapareça como resultado de formas de ressonância no qual dois apresentam uma carga formal em extremidades opostas da cadeia de polimetina.

As maiores diferenças nas propriedades químicas e óticas de polienos e cianinas mostram a importância que o *BLA* tem nas propriedades eletrônicas de cadeias moleculares conjugadas[21], como *bandgap*[22], polarizabilidades lineares e não lineares[23–25], absorção de dois fótons [26–28] e espectro *Raman* [29,30]. Logo, a habilidade de determinar geometrias moleculares precisas, em particular, *BLAs* precisos, devem ser o primeiro e mais importante passo no cálculo de propriedades eletrônicas de cadeias conjugadas com primeiros princípios.

O *BLA* pode ser usado para verificar a quantidade de polarização em uma molécula, ele pode ser definido como a média da diferença entre comprimentos de ligação carbono-carbono adjacentes em uma cadeia de polieno (CH_n). O *BLA* é definido da seguinte forma:

$$BLA = Med.BL_p - Med.BL_i \quad (2.1)$$

onde $Med.BL_p$ é a média dos comprimentos de ligação de número par e $Med.BL_i$ a média dos comprimentos de número ímpar.

Porém, obter *BLAs* confiáveis de teorias de estrutura eletrônica não é uma tarefa trivial. Enquanto a teoria de Hartree-Fock (HF) superestima o *BLA* significativamente, as aproximações da teoria de perturbação de segunda ordem (MP2) e as muito utilizadas aproximações para o funcional de correlação e troca, da teoria do funcional da densidade (*DFT*), a subestimam [19,20,31–34]. Um grande número de trabalhos têm investigado o papel da correlação eletrônica no problema do *BLA* e sua inclusão por correções pós-HF [15, 16, 20, 32–34].

Mesmo quando aproximações de alto nível são utilizadas para determinar as propriedades eletrônicas, as geometrias ainda são obtidas via *DFT* ou *MP2*. Porém, ainda existe a necessidade por aproximações analíticas que permitam um cálculo mais preciso do *BLA* por um custo computacional razoável para sistemas com dezenas à centenas de átomos.

As primeiras pesquisas teóricas deste problema foram feitas por *Lennard-Jones* em 1937, onde ele estudou os polienos lineares de fórmula $C_{2n}H_{2n+2}$ através da te-

oria molecular orbital *LCAO* (*Linear Combination of Atomic Orbitals*) e descobriu que quando n aumenta, ligações alternadas ficam mais próximas em comprimento.

Em 1938, Coulson estudou o radical de polienos lineares, $C_{2n+1}H_{2n+3}$, e encontrou que para n menor não teria nenhuma diferença significativa entre comprimentos de ligação. Além disso, Coulson também afirmou que para o limite $n \rightarrow \infty$, a configuração na qual ligações alternadas eram de dois comprimentos diferentes seriam menos estáveis do que aquelas em que todos os comprimentos eram iguais.

Em outro artigo em 1939, Coulson introduziu a definição de orbital molecular e de ordem de ligação. Ele mostrou que se todos os comprimentos em um polieno são considerados iguais, as ordens de ligação tendem a igualdade à medida que se afastam das extremidades da cadeia. Ele considerava esta conclusão como suporte da hipótese de que no meio de polienos longos os comprimentos de ligação de fato tendem a igualdade.

Por quase dez anos, estas conclusões teóricas permaneceram incontestadas. Em 1949, numa pesquisa sobre espectro eletrônico de polienos, Kuhn encontrou que as frequências de absorção observadas só poderiam ser interpretadas pela teoria do elétron livre [35], se os comprimentos de ligação alternassem ao longo da cadeia, mesmo em cadeias muito longas.

Dewar, em 1952 [36], interpretou os mesmos dados pela teoria orbital molecular *LCAO* e apontou que na ausência de uma alternância persistente as frequências de absorção devem tender a zero a medida que n tende ao infinito, enquanto que as frequências observadas parecem tender a um limite finito. Concordância entre teoria e experimento pode ser restaurada supondo que todas as ligações duplas e ligações simples num polieno são equivalentes entre si.

Foi Platt, em 1956, quem apresentou a primeira explicação da discrepância entre a teoria de *Lennard-Jones* e o experimento. Ele sugeriu que a interação configuracional poderia dar uma resposta ao problema. Com isso ele quis dizer que, com uma forte interação entre o estado fundamental eletrônico e alguns estados pouco excitados, produzidos não por repulsão eletrônica, mas por uma distorção antissimétrica da estrutura nuclear.

No ano seguinte, uma ideia equivalente foi proposta por Labhart, baseada nos orbitais moleculares de elétron livre. Labhart mostrou que uma alternância dos compri-

mentos de ligação, representado como um potencial periódico, necessariamente leva a um aumento na estabilidade se o sistema conjugado for grande o suficiente. Ele citou dados cristalográficos para suportar a sua conclusão. De forma independente, *Ooshika* propôs uma ideia baseada na teoria *LCAO* auto-consistente e encontrou que a configuração de equilíbrio com todos os comprimentos de ligação iguais a 1,389 Å é menos estável do que a configuração cuja a alternância dos comprimentos de ligação varia entre 1,346 e 1,463 Å.

No próximo capítulo veremos como o método foi utilizado com o propósito de obtermos um parâmetro estrutural que pode ser comparado ao *BLA*.

Capítulo 3

Metodologia

Neste capítulo iremos descrever o método utilizado para fazer as análises descritas no próximo capítulo.

Para otimizar a geometria das moléculas presentes neste trabalho, utilizamos o *Gaussian 09*, através do método *DFT* [37, 38] e do funcional híbrido *B3LYP* [39–42] juntamente com a base *6-31G(d,p)* [43, 44]. Este método é conhecido por gerar geometrias confiáveis para sistemas orgânicos [45].

Para o cálculo do parâmetro estrutural com o *PCA* usamos um código escrito no programa *Mathematica* a ser descrito mais adiante neste capítulo.

3.1 Teoria do Funcional da Densidade (*DFT*)

Formulada por *Hohenberg, Kohn e Sham* [37, 38], a *DFT* reduz o problema de muitos corpos a um problema de um corpo submetido a um potencial efetivo. O ponto essencial é que o cálculo da função de onda de muitos corpos é contornado e o conhecimento da densidade eletrônica do estado fundamental é suficiente para calcular todas as propriedades físicas de interesse.

O *DFT* é baseado nos dois teoremas a seguir:

- **Teorema 1:** O potencial externo $v(\mathbf{r})$ é univocamente determinado pela densidade eletrônica $\rho(\mathbf{r})$, a menos de uma constante trivial e positiva.

- **Teorema 2:** Seja $\rho(\mathbf{r})$ uma densidade eletrônica não negativa normalizada a N . Então $E_0 < E_v[\rho]$ para:

$$E_v(\rho(\mathbf{r})) = F(\rho(\mathbf{r})) + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r}, \quad (3.1)$$

onde:

$$F(\rho(\mathbf{r})) = \langle \Psi(\rho(\mathbf{r})) | \hat{\mathbf{T}} + \hat{\mathbf{U}} | \Psi(\rho(\mathbf{r})) \rangle \quad (3.2)$$

em que $F(\rho(\mathbf{r}))$ é um funcional universal representado pela energia cinética $\hat{\mathbf{T}}$ e pela energia potencial $\hat{\mathbf{U}}$.

Neste contexto, podemos considerar que o problema de n elétrons interagentes pode ser transformado em n problemas de um elétron.

$$[-\nabla^2 + V_{KS}(\rho(\mathbf{r}))]\psi_j(\mathbf{r}) = E_j\psi_j(\mathbf{r}), \quad (3.3)$$

ou

$$\hat{H}_{KS}\psi_j(\mathbf{r}) = E_j\psi_j(\mathbf{r}) \quad (3.4)$$

Esta é uma equação tipo Schrödinger de uma partícula, onde o primeiro termo representa a energia cinética e o segundo termo é a energia potencial e ψ_j representa as autofunções para uma dada partícula.

Observe que V_{KS} depende de ρ que depende de $\psi_j(\mathbf{r})$. Desta forma, para resolver esta equação precisamos de um cálculo iterativo auto consistente.

3.1.1 Aproximações para o Termo de Troca e Correlação

Na equação de *Kohn-Sham*, apresentada na seção anterior, temos que o potencial $V_{KS}(\rho(\mathbf{r}))$ é dado pela soma da contribuição nuclear (V_N), do potencial eletrostático (V_E) e do termo de correlação e troca (v_{xc}). Este termo de correlação e troca, para

um sistema real, não pode ser obtido de forma exata, sendo necessário o uso de aproximações.

Uma aproximação bastante utilizada é a *LDA* (“Local Density Aproximation”), onde a energia de correlação e troca para um sistema de gás de elétrons não homogêneo de densidade $\rho(\mathbf{r})$ no ponto r , é assumida igual a energia de correlação e troca de um gás de elétrons homogêneo com a mesma densidade. Supõe-se ainda que $\rho(\mathbf{r})$ varia suavemente nas proximidades do ponto r , desta forma, o funcional para a energia de correlação e troca ($E_{xc}(\rho(\mathbf{r}))$) é dado por:

$$E_{xc}(\rho(\mathbf{r})) = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho(\mathbf{r})) d^3\mathbf{r} \quad (3.5)$$

onde $\epsilon_{xc}(\rho(\mathbf{r}))$ é a energia de correlação e troca por elétron de um gás de elétrons homogêneo de densidade $\rho(\mathbf{r})$.

Desta forma, temos que o potencial é dado por:

$$v_{xc}(\rho(\mathbf{r})) \equiv \frac{d}{d\rho(\mathbf{r})} [\rho(\mathbf{r}) \epsilon(\rho(\mathbf{r}))]. \quad (3.6)$$

Esta é uma aproximação eficiente para tratar sistemas em que a densidade eletrônica varia lentamente.

Uma generalização da *LDA* foi formulada para incluir o spin eletrônico, conhecida como *Local Spin-Density Aproximation (LSDA)* e o termo de troca e correlação é dado da seguinte forma:

$$E_{xc}[\rho_{\uparrow}, \rho_{\downarrow}] = \int \epsilon_{xc}(\rho_{\uparrow}, \rho_{\downarrow}) d^3\mathbf{r} \quad (3.7)$$

Já a aproximação do gradiente generalizado (*GGA - Generalized Gradiente Aproximation*), que também é uma aproximação local, leva em conta o gradiente da densidade na mesma coordenada:

$$E_{xc}(\rho(\mathbf{r})) = \int \epsilon_{xc}(\rho(\mathbf{r}), \vec{\nabla}\rho(\mathbf{r})) d^3\mathbf{r} \quad (3.8)$$

Ao contrário da aproximação *LDA*, onde existe um único $\epsilon_{xc}(\rho(\mathbf{r}))$ correto, na aproximação *GGA* existem diferentes parametrizações para $\epsilon_{xc}(\rho(\mathbf{r}), \vec{\nabla}\rho(\mathbf{r}))$ que originam funcionais diferentes.

Também existe uma classe de funcionais híbridos para o funcional da energia de troca e correlação que combinam uma parte exata, da teoria de *Hartree-Fock*, para o termo de troca e funcionais *DFT* puros [46].

Neste trabalho utilizamos um funcional híbrido bastante conhecido, o *B3LYP*. Este funcional vem de três parâmetros de *Becke* [39] e e do funcional de correlação de *Lee, Yang e Parr* (*LYP*) [42] e é definido da seguinte forma:

$$E_{xc}^{B3LYP} = E_x^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + E_c^{LDA} + a_c(E_c^{GGA} - E_c^{LDA}) \quad (3.9)$$

onde $a_0 = 0,20$, $a_x = 0,72$, $a_c = 0,81$ e E_x^{HF} é o termo de troca da teoria de *Hartree-Fock*. E_x^{GGA} é o funcional de troca de *Becke B88*, E_c^{GGA} é o funcional de correlação de *Lee, Yang e Parr* para o *B3LYP* e E_c^{LDA} é a aproximação da densidade local *VWN* [47] para o funcional de correlação.

Na próxima seção veremos como a técnica da *PCA* foi utilizada para alcançarmos o objetivo proposto para este trabalho.

3.2 Encontrando a *PCA*

Iniciamos definindo o conjunto de dados a ser analisados, este é um *array* que forma uma matriz, onde cada linha é uma molécula e cada elemento é um comprimento de ligação da molécula (como foi descrito no Capítulo 1):

$$data = \{ \{ Mol_1BL_1, Mol_1BL_2, Mol_1BL_3, \dots \}, \{ Mol_2BL_1, Mol_2BL_2, Mol_2BL_3, \dots \}, \\ \{ Mol_3BL_1, Mol_3BL_2, Mol_3BL_3, \dots \}, \dots \} \quad (3.10)$$

o que nos dá a seguinte matriz:

$$\overbrace{\begin{pmatrix} Mol_1BL_1 & Mol_1BL_2 & Mol_1BL_3 & \cdots \\ Mol_2BL_1 & Mol_2BL_2 & Mol_2BL_3 & \cdots \\ Mol_3BL_1 & Mol_3BL_2 & Mol_3BL_3 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}}^{\text{Moléculas}} \left. \vphantom{\begin{pmatrix} Mol_1BL_1 & Mol_1BL_2 & Mol_1BL_3 & \cdots \\ Mol_2BL_1 & Mol_2BL_2 & Mol_2BL_3 & \cdots \\ Mol_3BL_1 & Mol_3BL_2 & Mol_3BL_3 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}} \right\} \text{Ligações de cada molécula.} \quad (3.11)$$

Em seguida encontramos a matriz com média zero nas colunas, $zdata$:

$$zdata = ZeroMean[data] \quad (3.12)$$

este comando calcula a média de cada coluna e subtrai de cada elemento delas, nos dando assim, um conjunto com média nula em cada coluna.

Feito isso, temos que encontrar a matriz covariância c :

$$c = CovarianceMatrix[zdata] \quad (3.13)$$

Tendo a matriz c e, encontrando os seus autovetores, teremos as PC 's que precisamos para o cálculo do BLA_{PCA} .

$$PCA = -Eigenvectors[c] \quad (3.14)$$

A PCA será uma matrix composta pelos autovetores da matrix covariância que são chamados aqui de PC_i onde, $i = 1, 2, 3, \dots$

$$PCA = \begin{pmatrix} PC_1 \\ PC_2 \\ PC_3 \\ \vdots \end{pmatrix} \quad (3.15)$$

3.2.1 Calculando o BLA_{PCA}

Como visto no capítulo anterior, o BLA é calculado da seguinte forma:

$$BLA = Med.BL_p - Med.BL_i. \quad (3.16)$$

A nossa proposta é redefinir o BLA utilizando a PCA apenas com a PC_1 :

$$PC_1 = (a_1, a_2, a_3, \dots) \quad (3.17)$$

Com estes coeficientes poderemos encontrar um parâmetro que depende dos comprimentos de ligação da seguinte forma:

$$BLA_{PCA} = a_1BL_1 + a_2BL_2 + a_3BL_3 + \dots \quad (3.18)$$

Note que, enquanto o BLA está associado a um conjunto definido de ligações, ao longo da conjugação, no caso do BLA_{PCA} , a própria técnica indica quais são as ligações que devem compor o BLA uma vez que, os coeficientes (a_1, a_2, \dots, a_n) são variáveis e definidos pela própria técnica e, com esse coeficientes, conseguimos saber o impacto de cada ligação ao valor do BLA , conseguindo assim, definir quais ligações serão usadas para calcular o parâmetro, já que, se o coeficiente for muito pequeno, a ligação relacionada a ele não é relevante, podendo ser descartada.

Com esta característica, a técnica da PCA pode ser útil para cadeias mais complexas, onde não fica claro quais ligações devem ser utilizadas para calcular o BLA .

No próximo capítulo veremos os resultados obtidos ao utilizar a técnica PCA descrita até aqui, para encontrar o BLA_{PCA} de um conjunto de 17 moléculas, assim como, faremos uma comparação com a forma usual de calcular o BLA e, veremos como esta técnica pode ser usada para ir além do BLA e fazer outras análises como por exemplo, discutir a importância de cada ligação da cadeia carbônica e comparar com a previsão experimental.

Capítulo 4

Resultados

Neste capítulo apresentaremos os resultados deste trabalho, onde calculamos o *BLA* de dezessete moléculas derivadas de tertiofenos $R-C_{12}H_nS_3-R$, onde R representa os radicais que variam em cada molécula. Usando a análise dos componentes principais fomos capazes de encontrar um novo *BLA* e estabelecer uma relação com a definição padrão. Também pudemos mostrar que o *BLA* muda devido à aromaticidade que, varia de acordo com a forma das ligações em cada molécula, assim como a importância de cada ligação na cadeia. A seguir listamos as moléculas utilizadas para as análises deste trabalho.

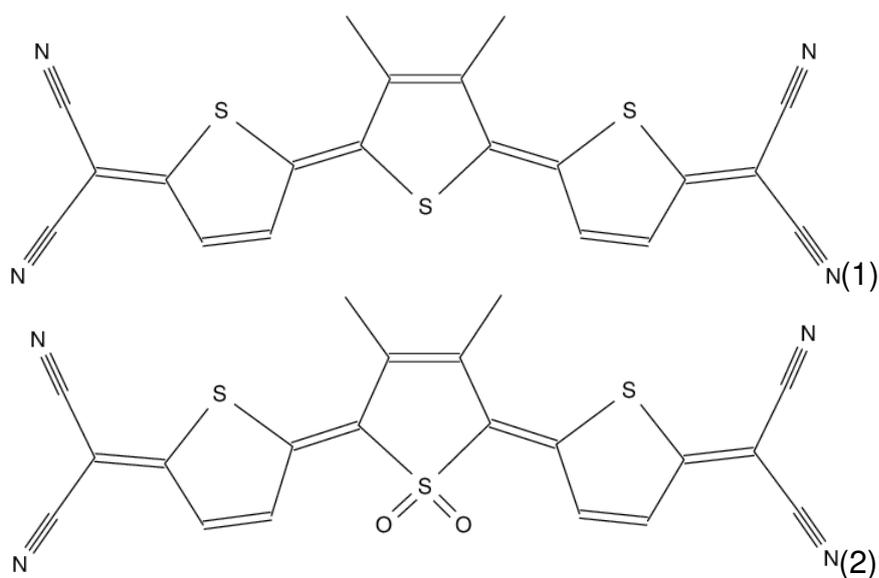


Figura 4.1: Moléculas: (1) - $C_{20}H_{10}S_3N_4$ e (2) - $C_{20}H_{10}S_3N_4O_2$

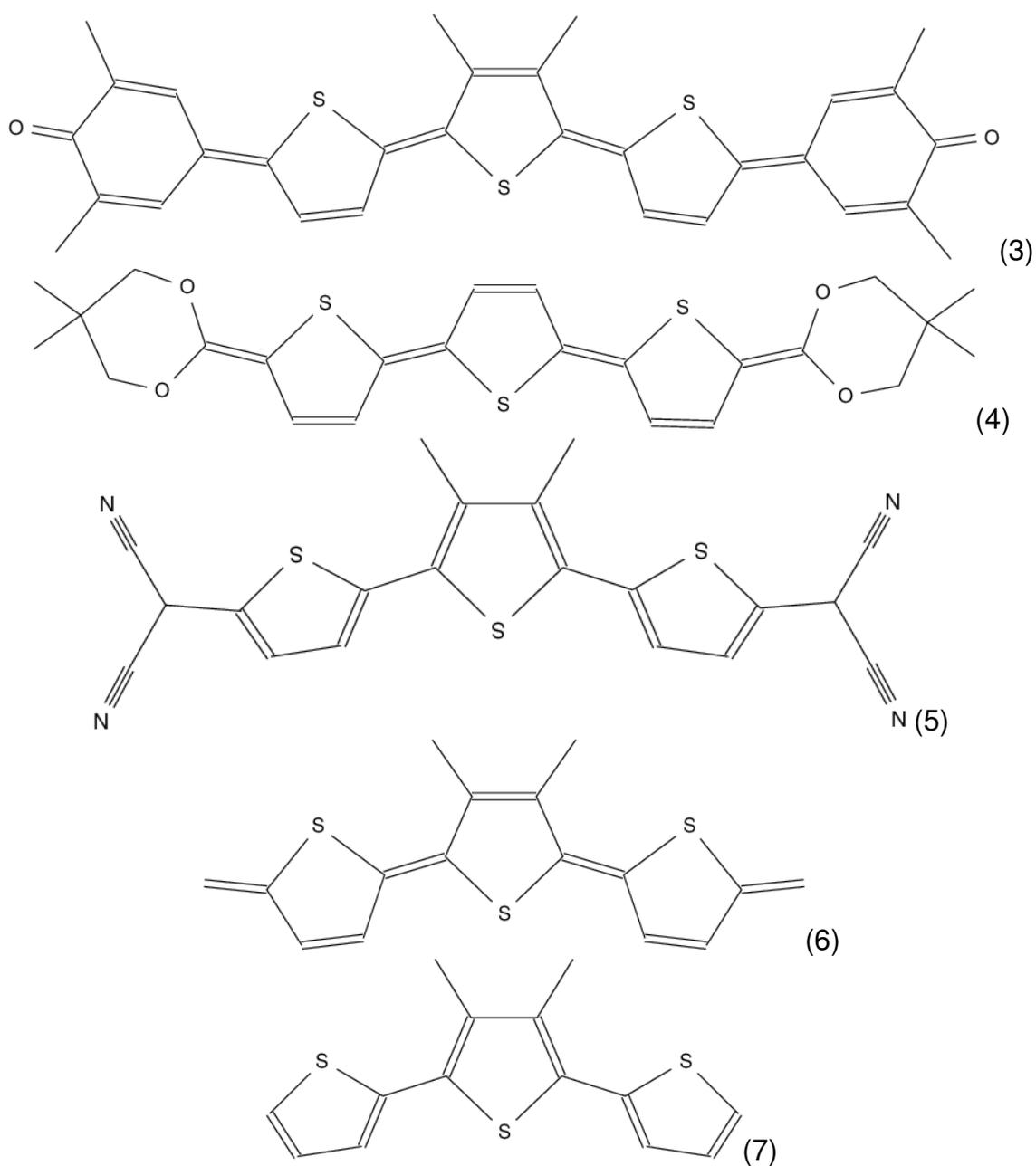


Figura 4.2: Moléculas: (3) - $C_{30}H_{26}S_3O_2$, (4) - $C_{24}H_{26}S_3O_4$, (5) - $C_{20}H_{14}S_3N_4$, (6) - $C_{16}H_{14}S_3$ e (7) - $C_{14}H_{12}S_3$

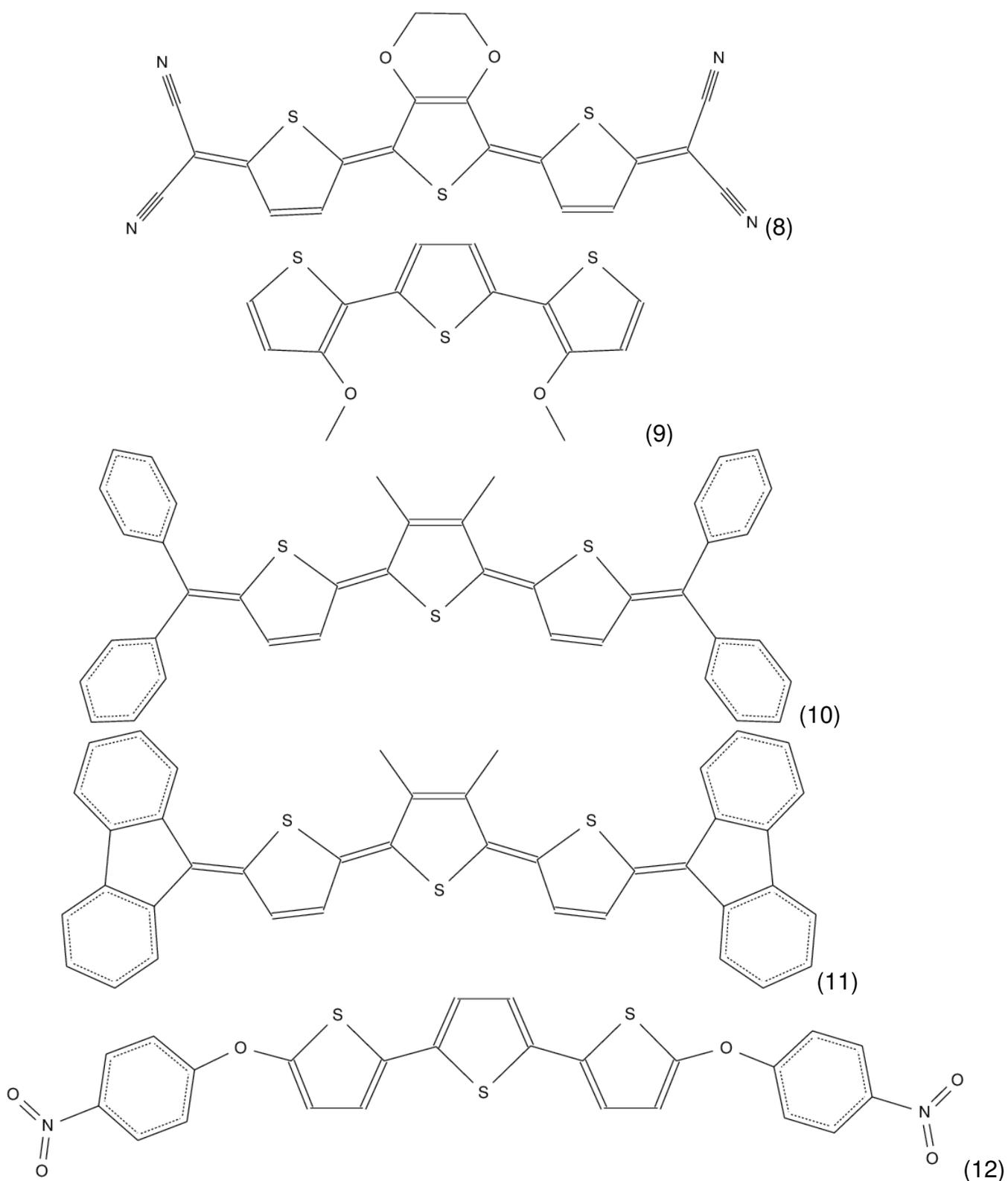


Figura 4.3: Moléculas: (8) - $C_{20}H_8S_3N_4O_2$, (9) - $C_{14}H_{12}S_3O_2$, (10) - $C_{40}H_{30}S_3$, (11) - $C_{40}H_{26}S_3$ e (12) - $C_{24}H_{14}S_3O_6N_2$

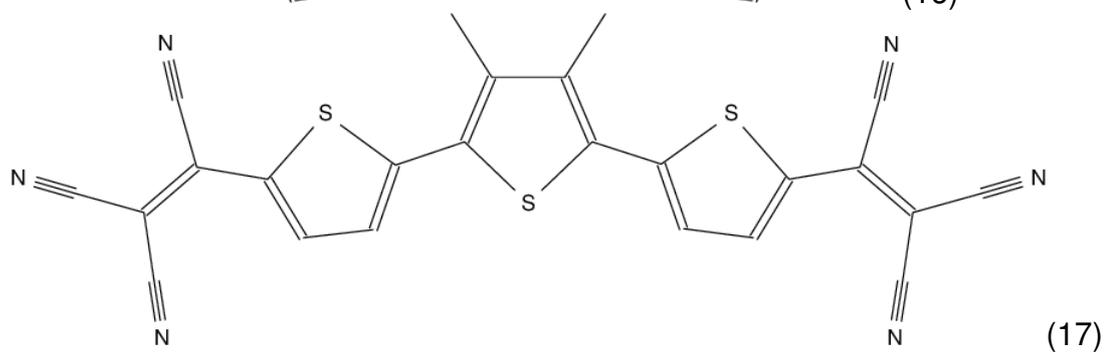
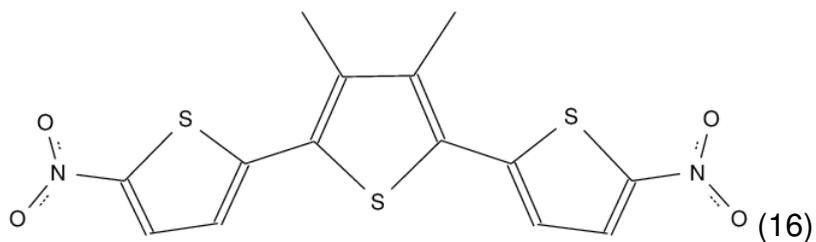
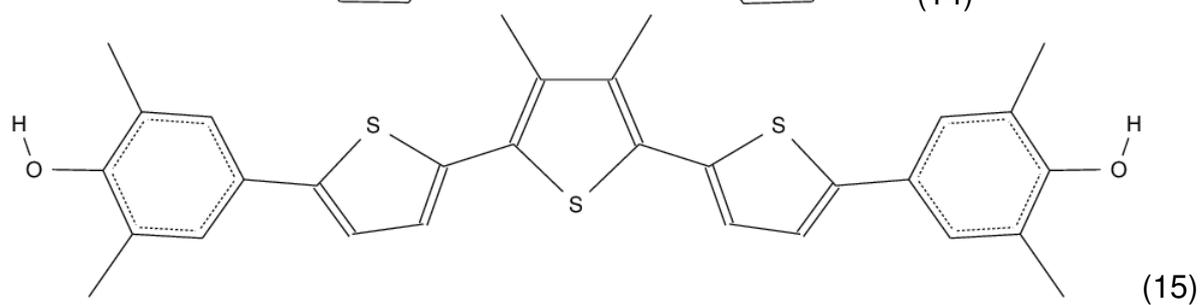
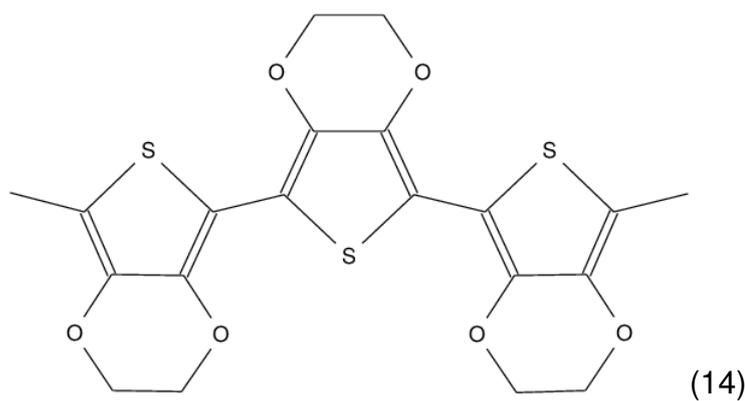
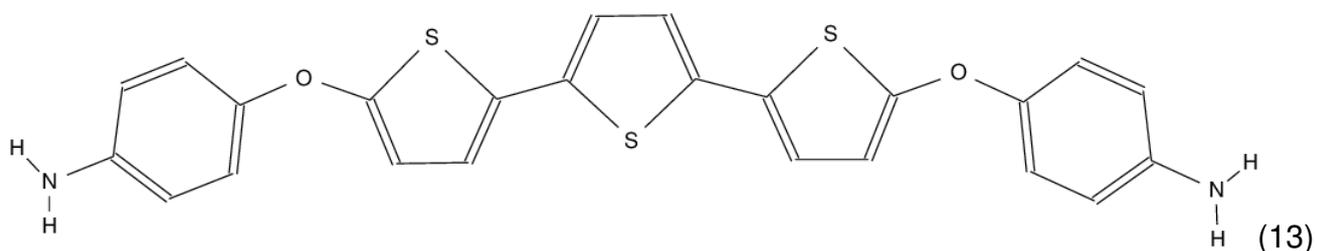


Figura 4.4: Moléculas: (13) - $C_{24}H_{18}S_3O_2N_2$, (14) - $C_{20}H_{18}S_3O_6$, (15) - $C_{30}H_{28}S_3O_2$, (16) - $C_{14}H_{10}S_3O_4N_2$ e (17) - $C_{24}H_{10}S_3N_6$

4.1 Obtenção do *BLA*

No procedimento realizado para fazer a *PCA*, a matriz inicial de dados é composta pelos comprimentos de ligação carbono-carbono (ver Tabela 4.1) entre doze átomos de carbono da cadeia principal e entre os três átomos de enxofre que completam os três anéis de tiofeno (ver Tabela 4.2), o que nos dá dezessete ligações. Cada linha desta matriz corresponde à uma molécula e cada coluna à um comprimento de ligação formando uma matriz $M_{17 \times 17}$.

Tabela 4.1: Comprimentos de ligação (CL) carbono-carbono (em Å) via *B3LYP/6-31G***

Molécula	CL_1	CL_2	CL_3	CL_4	CL_5	CL_6	CL_7	CL_8	CL_9	CL_{10}	CL_{11}
$C_{20}H_{10}S_3N_4$	1,43	1,36	1,43	1,39	1,44	1,39	1,44	1,39	1,43	1,36	1,43
$C_{20}H_{10}S_3N_4O_2$	1,44	1,36	1,45	1,37	1,44	1,39	1,44	1,37	1,45	1,36	1,44
$C_{30}H_{26}S_3O_2$	1,42	1,37	1,43	1,39	1,43	1,39	1,43	1,39	1,43	1,37	1,42
$C_{24}H_{26}S_3O_4$	1,44	1,36	1,44	1,38	1,43	1,37	1,43	1,38	1,44	1,36	1,44
$C_{20}H_{14}S_3N_4$	1,37	1,42	1,38	1,46	1,38	1,44	1,38	1,46	1,38	1,42	1,37
$C_{16}H_{14}S_3$	1,45	1,36	1,44	1,38	1,45	1,38	1,45	1,38	1,44	1,36	1,45
$C_{14}H_{12}S_3$	1,37	1,42	1,38	1,46	1,38	1,44	1,38	1,46	1,38	1,42	1,36
$C_{20}H_8S_3N_4O_2$	1,43	1,36	1,43	1,38	1,42	1,38	1,42	1,38	1,43	1,36	1,43
$C_{14}H_{12}S_3O_2$	1,37	1,43	1,39	1,44	1,38	1,41	1,38	1,44	1,39	1,43	1,37
$C_{40}H_{30}S_3$	1,44	1,36	1,43	1,38	1,44	1,39	1,44	1,38	1,43	1,36	1,44
$C_{40}H_{26}S_3$	1,43	1,37	1,43	1,39	1,44	1,39	1,44	1,39	1,43	1,37	1,43
$C_{24}H_{14}S_3O_6N_2$	1,36	1,42	1,38	1,45	1,38	1,42	1,38	1,45	1,38	1,42	1,36
$C_{24}H_{18}S_3O_2N_2$	1,37	1,43	1,38	1,44	1,38	1,42	1,38	1,44	1,38	1,43	1,37
$C_{20}H_{18}S_3O_6$	1,37	1,43	1,39	1,44	1,38	1,42	1,38	1,44	1,38	1,43	1,37
$C_{30}H_{28}S_3O_2$	1,38	1,42	1,38	1,45	1,38	1,44	1,38	1,45	1,38	1,42	1,38
$C_{14}H_{10}S_3O_4N_2$	1,37	1,41	1,39	1,45	1,39	1,43	1,39	1,45	1,39	1,41	1,37
$C_{24}H_{10}S_3N_6$	1,39	1,40	1,40	1,44	1,39	1,43	1,39	1,44	1,40	1,40	1,39

Tabela 4.2: Comprimentos de ligação (CL) carbono-enxofre (em Å) via *B3LYP/6-31G***

Molécula	CL_{12}	CL_{13}	CL_{14}	CL_{15}	CL_{16}	CL_{17}
$C_{20}H_{10}S_3N_4$	1,76	1,78	1,77	1,77	1,78	1,76
$C_{20}H_{10}S_3N_4O_2$	1,76	1,78	1,81	1,81	1,78	1,76
$C_{30}H_{26}S_3O_2$	1,77	1,78	1,77	1,77	1,78	1,77
$C_{24}H_{26}S_3O_4$	1,78	1,78	1,78	1,78	1,78	1,78
$C_{20}H_{14}S_3N_4$	1,74	1,76	1,75	1,75	1,76	1,74
$C_{16}H_{14}S_3$	1,79	1,79	1,78	1,78	1,79	1,79
$C_{14}H_{12}S_3$	1,73	1,76	1,75	1,75	1,76	1,73
$C_{20}H_8S_3N_4O_2$	1,76	1,78	1,78	1,78	1,78	1,76
$C_{14}H_{12}S_3O_2$	1,73	1,76	1,76	1,76	1,76	1,73
$C_{40}H_{30}S_3$	1,79	1,78	1,78	1,78	1,78	1,79
$C_{40}H_{26}S_3$	1,78	1,78	1,77	1,77	1,78	1,78
$C_{24}H_{14}S_3O_6N_2$	1,75	1,76	1,76	1,76	1,76	1,76
$C_{24}H_{18}S_3O_2N_2$	1,75	1,77	1,76	1,76	1,77	1,75
$C_{20}H_{18}S_3O_6$	1,75	1,77	1,77	1,77	1,77	1,75
$C_{30}H_{28}S_3O_2$	1,75	1,76	1,75	1,75	1,76	1,75
$C_{14}H_{10}S_3O_4N_2$	1,74	1,75	1,75	1,75	1,75	1,74
$C_{24}H_{10}S_3N_6$	1,76	1,74	1,75	1,75	1,74	1,76

No Apêndice A temos os comprimentos de ligação com mais algarismos significativos.

Iniciando a análise para esta matriz, teremos uma matriz covariância $C_{17 \times 17}$, dezessete autovetores e dezessete autovalores onde o primeiro autovetor é a PC_1 , o segundo é a PC_2 e assim por diante. Na Tabela 4.3 temos os coeficientes da PC_1 e a primeira linha da matriz covariância e suas ligações correspondentes. Este primeiro autovetor, que chamamos de PC_1 , é o vetor que contém os coeficientes para o cálculo do BLA_{PCA} .

Nota-se que as ligações dos carbonos com o enxofre são menos relevantes e não apresentarão diferenças muito significantes no BLA . Isto se dá devido ao fato de que a variação nestas ligações é pequena como podemos ver, de forma mais evidente, na

Tabela 4.3: Coeficientes do BLA_{PCA} (PC_1 e primeira linha da matriz C)

Ligação	PC_1	C_{1n}
C-C	-0,32	-0,0012
C-C	0,29	0,0010
C-C	-0,26	-0,0009
C-C	0,33	0,0012
C-C	-0,26	-0,0009
C-C	0,21	0,0007
C-C	-0,26	-0,0009
C-C	0,33	0,0012
C-C	-0,26	-0,0009
C-C	0,29	0,0010
C-C	-0,32	-0,0012
C-S	-0,15	-0,0005
C-S	-0,11	-0,0004
C-S	-0,12	-0,0004
C-S	-0,12	-0,0004
C-S	-0,11	-0,0004
C-S	-0,15	-0,0005

primeira linha da matriz covariância C na Tabela 4.3.

Caso os comprimentos de ligação com o enxofre fossem todos iguais, estas ligações não dariam nenhuma contribuição ao BLA , já que sem a variação, seus coeficientes correspondentes seriam nulos. Esta é uma das propriedades interessantes do cálculo da PCA . Esta técnica nos mostra quais ligações são mais ou menos significantes.

Sabendo disso, pode-se remover tais ligações da análise com a certeza de não estar perdendo contribuições consideráveis ao valor do BLA_{PCA} . Logo, a matriz inicial pode ser formada apenas pelas comprimentos de ligação entre os doze átomos de carbono da cadeia principal que estão na Tabela 4.1.

Refazendo a PCA apenas com as ligações carbono-carbono, formamos uma matriz inicial 11×11 e, conseqüentemente, teremos uma matriz covariância $C_{11 \times 11}$ e onze

autovetores e autovalores. Na Tabela 4.4 temos os autovalores da matriz covariância.

Tabela 4.4: Autovalores e seus pesos proporcionais

Autovalores	Proporção
0,01	97,3%
$1,70 \times 10^{-4}$	1,6%
$4,00 \times 10^{-5}$	0,4%
$3,00 \times 10^{-5}$	0,3%
$2,00 \times 10^{-5}$	0,2%
$8,03 \times 10^{-6}$	≈ 0
$1,09 \times 10^{-8}$	≈ 0
$1,06 \times 10^{-9}$	≈ 0
$3,50 \times 10^{-11}$	≈ 0
$8,54 \times 10^{-12}$	≈ 0
$3,72 \times 10^{-13}$	≈ 0

Analisando estes autovalores percebemos que o primeiro autovetor compreende aproximadamente 97,3% da covariância. Com isso, podemos então, utilizar apenas o primeiro autovetor associado ao maior autovalor, que podemos chamar de PC_1 (primeira componente principal). Apenas com este vetor iremos montar o vetor característico, com a segurança de não estar perdendo dados relevantes ao descartar o restante. Na Tabela 4.5 temos a PC_1 .

Tabela 4.5: Coeficientes da PC_1

PC_1	-0,34	0,30	-0,28	0,34	-0,28	0,22	-0,28	0,34	-0,28	0,30	-0,34

Comparando estes coeficientes com a fórmula para o BLA ($-\frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \dots$) percebe-se uma das funcionalidades da PCA , enquanto o BLA trata todas as ligações como iguais em relação a sua importância para o resultado, a PCA define coeficientes distintos, mostrando assim, a importância de cada ligação para o valor do BLA_{PCA} .

Com o vetor característico podemos calcular o BLA_{PCA} multiplicando o mesmo pela

matriz com as médias dos comprimentos de ligação, ou seja, ao realizar uma PCA na matriz inicial utilizando apenas um componente principal, teremos o BLA_{PCA} .

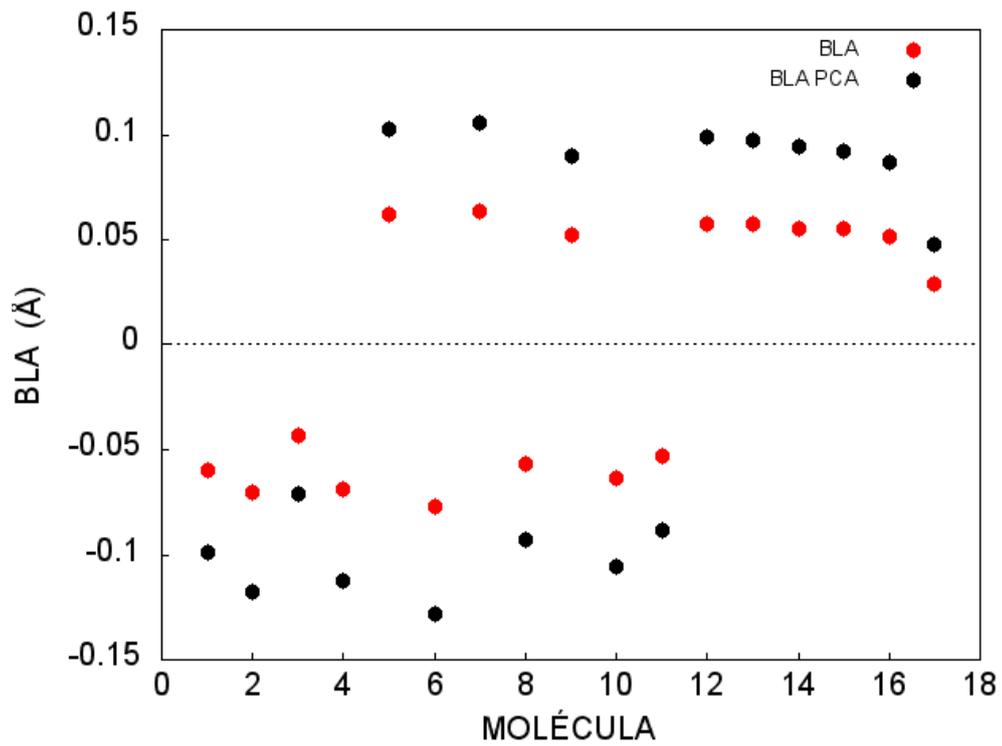
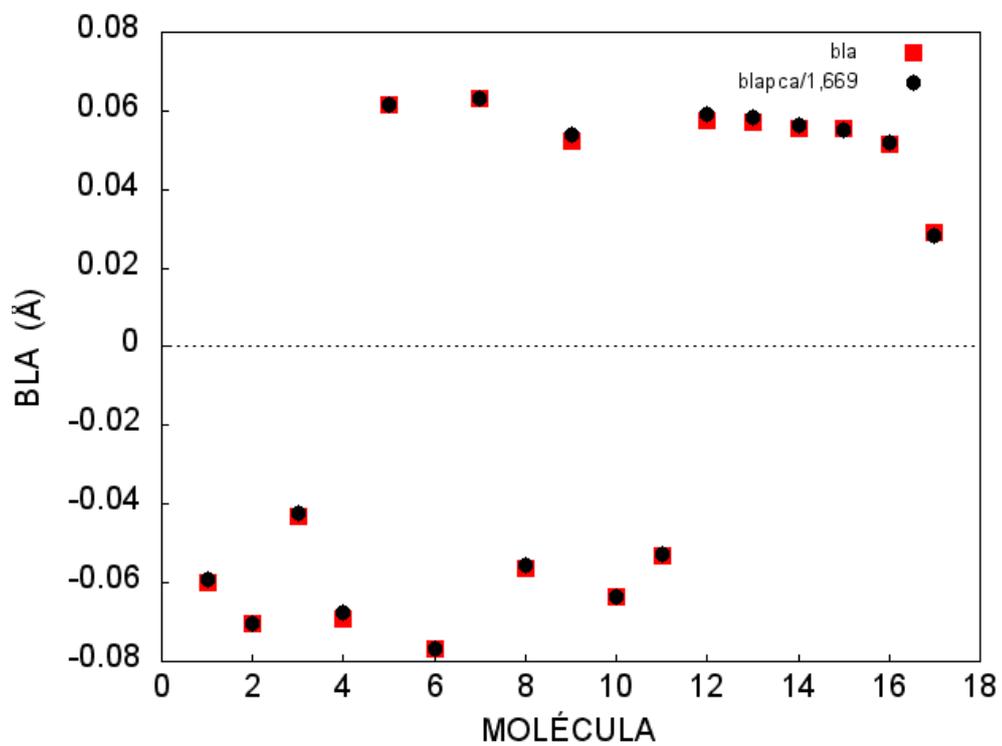
Feito isso, podemos fazer uma comparação com a forma usual do BLA (ver Tabela 4.6).

Tabela 4.6: Tabela comparativa entre BLA e BLA_{PCA} em Å

Molécula	BLA	BLA_{PCA}
$C_{20}H_{10}S_3N_4$	-0,06	-0,10
$C_{20}H_{10}S_3N_4O_2$	-0,07	-0,12
$C_{30}H_{26}S_3O_2$	-0,04	-0,07
$C_{24}H_{26}S_3O_4$	-0,07	-0,11
$C_{20}H_{14}S_3N_4$	0,06	0,10
$C_{16}H_{14}S_3$	-0,08	-0,13
$C_{14}H_{12}S_3$	0,06	0,11
$C_{20}H_8S_3N_4O_2$	-0,06	-0,09
$C_{14}H_{12}S_3O_2$	0,05	0,09
$C_{40}H_{30}S_3$	-0,06	-0,11
$C_{40}H_{26}S_3$	-0,05	-0,09
$C_{24}H_{14}S_3O_6N_2$	0,06	0,10
$C_{24}H_{18}S_3O_2N_2$	0,06	0,10
$C_{20}H_{18}S_3O_6$	0,06	0,09
$C_{30}H_{28}S_3O_2$	0,06	0,09
$C_{14}H_{10}S_3O_4N_2$	0,05	0,09
$C_{24}H_{10}S_3N_6$	0,03	0,05

Fazendo um gráfico com estes valores, podemos ver mais claramente as diferenças entre a BLA_{PCA} e a BLA (ver Figura 4.5).

Como o padrão do BLA_{PCA} parece ser o mesmo da BLA , a diferença entre eles deve ser apenas um fator de escala, assim, encontrando este fator e, fazendo um gráfico, podemos na Figura 4.6 ver como os dois se encontram quase perfeitamente.

Figura 4.5: Gráfico do BLA e do BLA_{PCA} Figura 4.6: Gráfico do BLA e do BLA_{PCA} dividido pelo fator de escala

Sendo assim, vemos que a *PCA* se faz útil para este tipo de cálculo para quando temos grupos muito grandes de moléculas e por outras características da técnica que veremos mais adiante.

4.2 Aromaticidade

Analisando as moléculas presentes neste trabalho, quanto ao padrão das ligações, percebe-se que há dois tipos de moléculas. Um grupo com tiofenos aromáticos e um com tiofenos quinóides, esta diferença se dá pela ordem das ligações duplas e simples. Na Figura 4.7, por exemplo, a molécula 1 é aromática, já a molécula 2 é quinóide. Note que o padrão de ligações duplas e simples é diferente, analisando uma mesma ligação, em um caso temos uma simples e no outro uma ligação dupla. Nesta seção iremos discutir algumas implicações desta diferença.

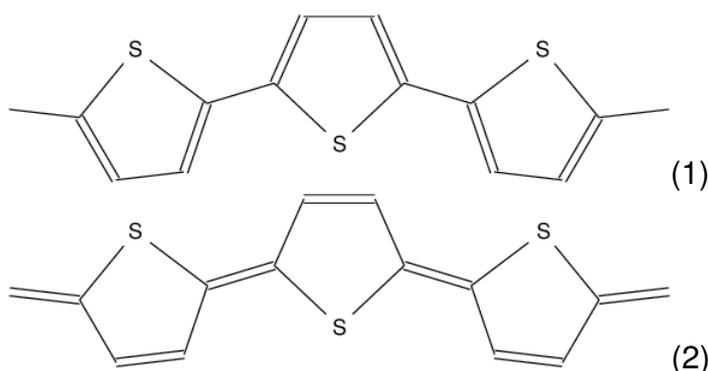


Figura 4.7: Padrão de ligações de uma molécula (1) aromáticas e (2) quinóides.

A diferença no padrão das ligações tem influência no seu comprimento, logo, o *BLA* também deve ser diferente entre elas. Como podemos ver no gráfico da Figura 4.8, as moléculas acima da linha são aromáticas com um valor de *BLA* maior que zero e abaixo da linha temos as quinóides com *BLA* negativo.

Ainda neste contexto, podemos analisar a importância de cada ligação na cadeia, já que este é um dos trunfos da *PCA* e, com esta diferença, a ordem das ligações duplas e simples é diferente. Para isto, faremos o mesmo cálculo, porém, desta vez vamos efetuar o procedimento nos dois grupos separadamente. Nas Figuras 4.9 e 4.10, vemos que a importância da cada ligação no *BLA* é um pouco diferente, mas, podemos perceber que as ligações mais externas têm maior significância em ambos

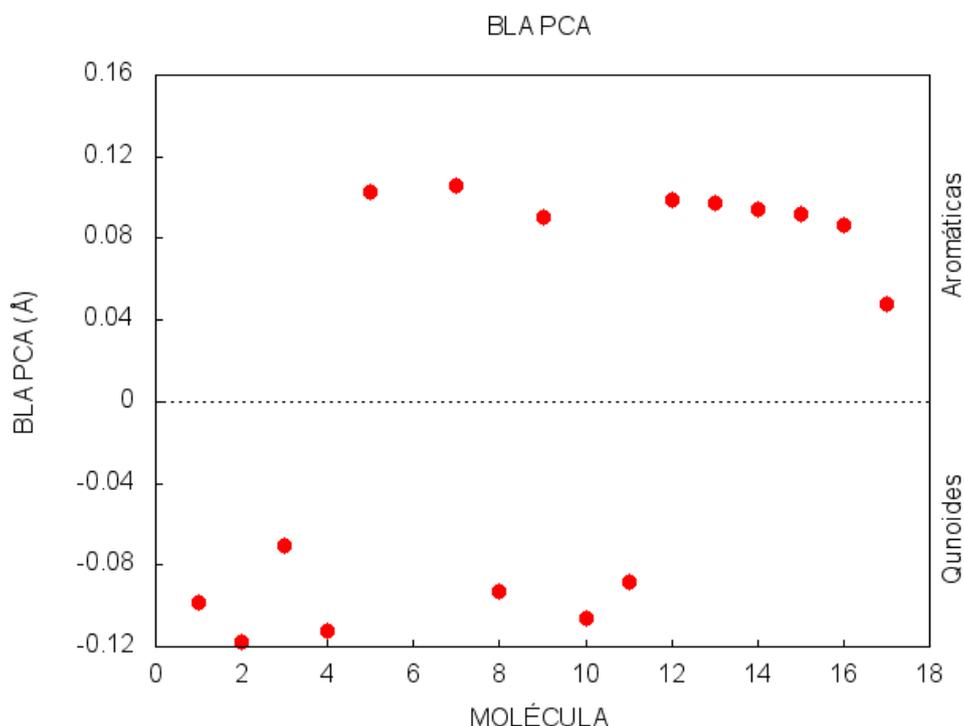


Figura 4.8: BLA_{PCA} calculado através do método da PCA

os grupos. A diferença entre elas é maior nas ligações que estão entre os anéis de tiofeno (ligações 4 e 8), onde nas moléculas com anéis quinóides, tais ligações têm significância bem maior que nas moléculas com os anéis aromáticos.

4.3 Número mínimo para um PCA funcional

Uma pergunta interessante que podemos fazer sobre a PCA é, se há um número mínimo de moléculas para realizar o procedimento? Qual seria este número? A partir desta dúvida podemos fazer algumas análises.

Uma análise que pode ser feita é sobre a variação da PC_1 de acordo com a quantidade de moléculas presentes no grupo. Então, pensando nisso, fizemos uma PCA gradual, começando com duas moléculas e acrescentando uma a uma para vermos como varia a PC_1 . No gráfico da Figura 4.11 vemos como se comportou a PC_1 .

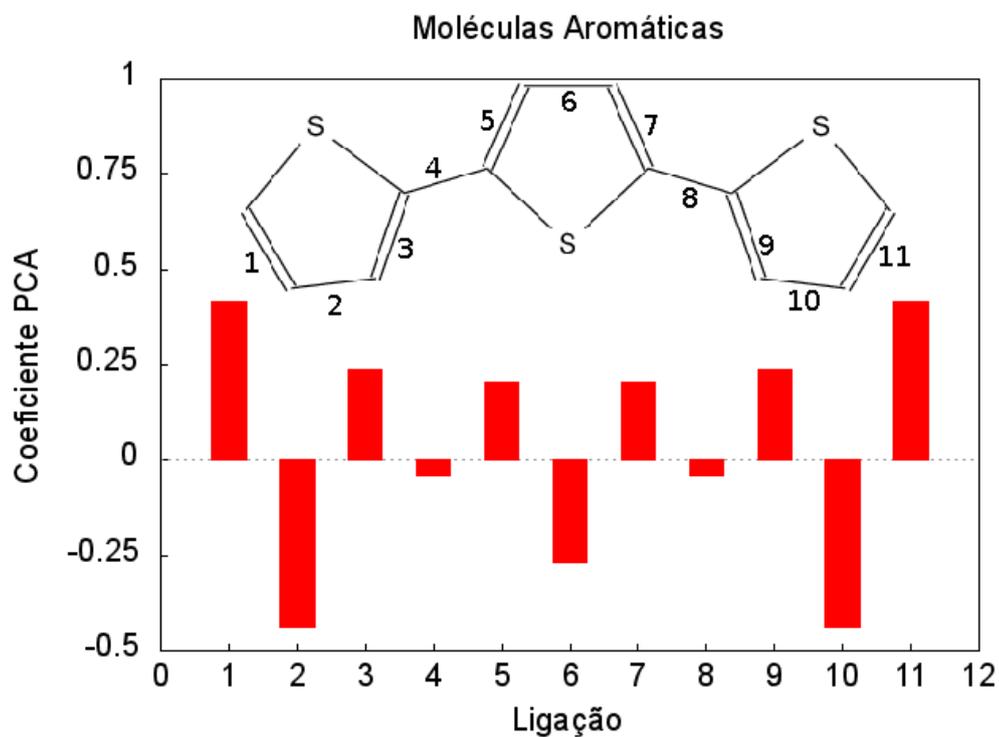


Figura 4.9: Importância das ligações das moléculas aromáticas.

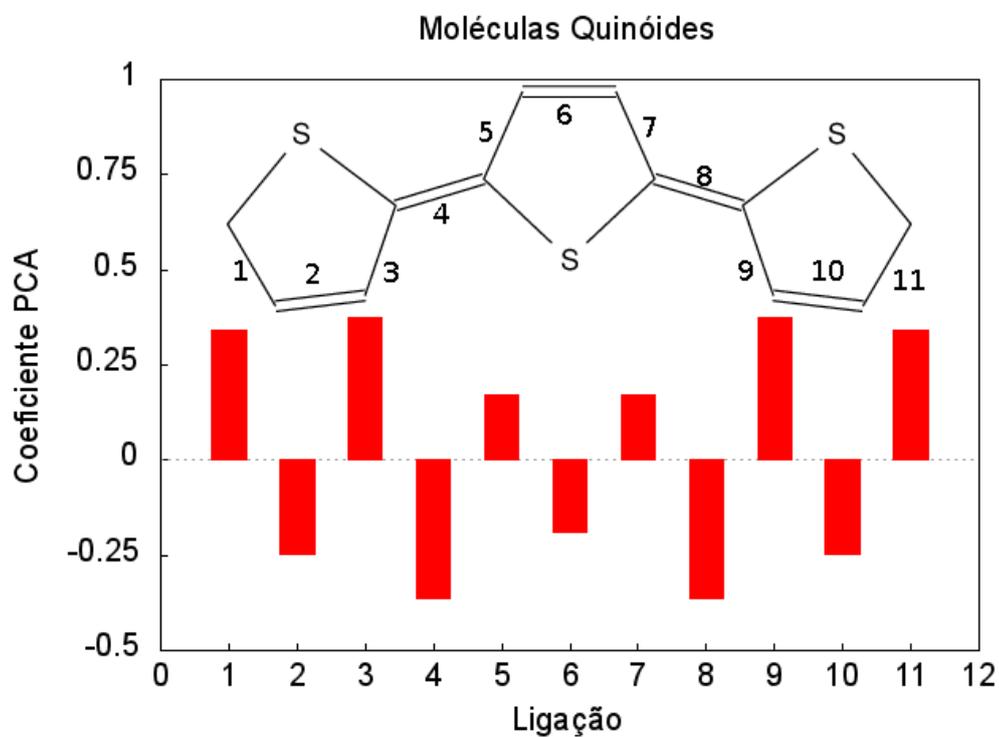


Figura 4.10: Importância das ligações das moléculas quinóides.

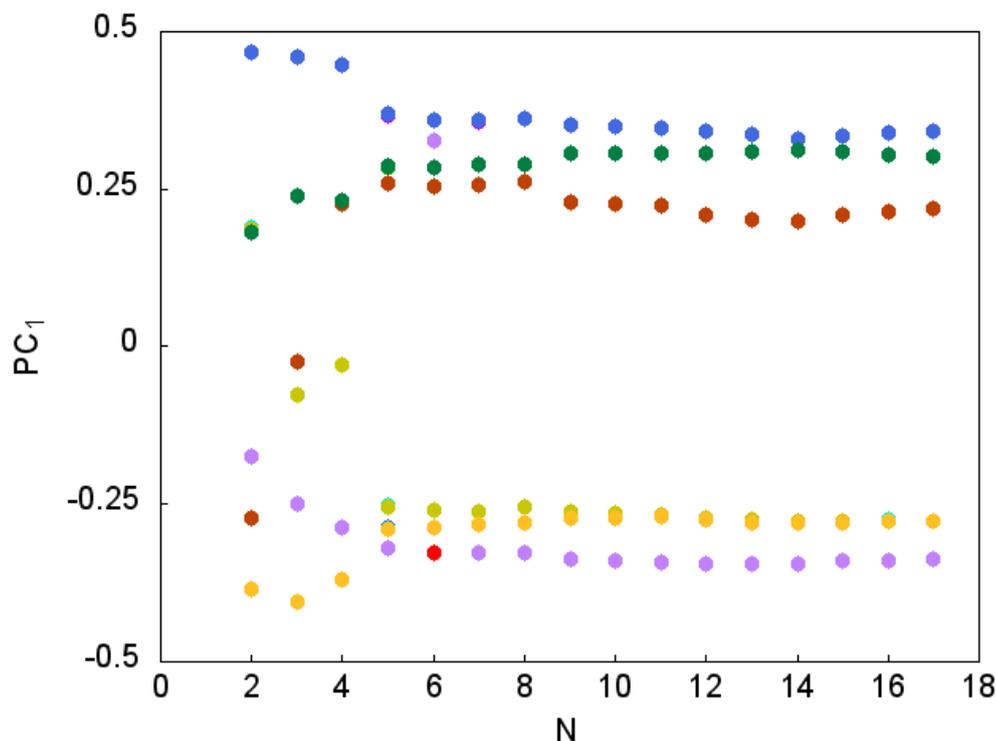


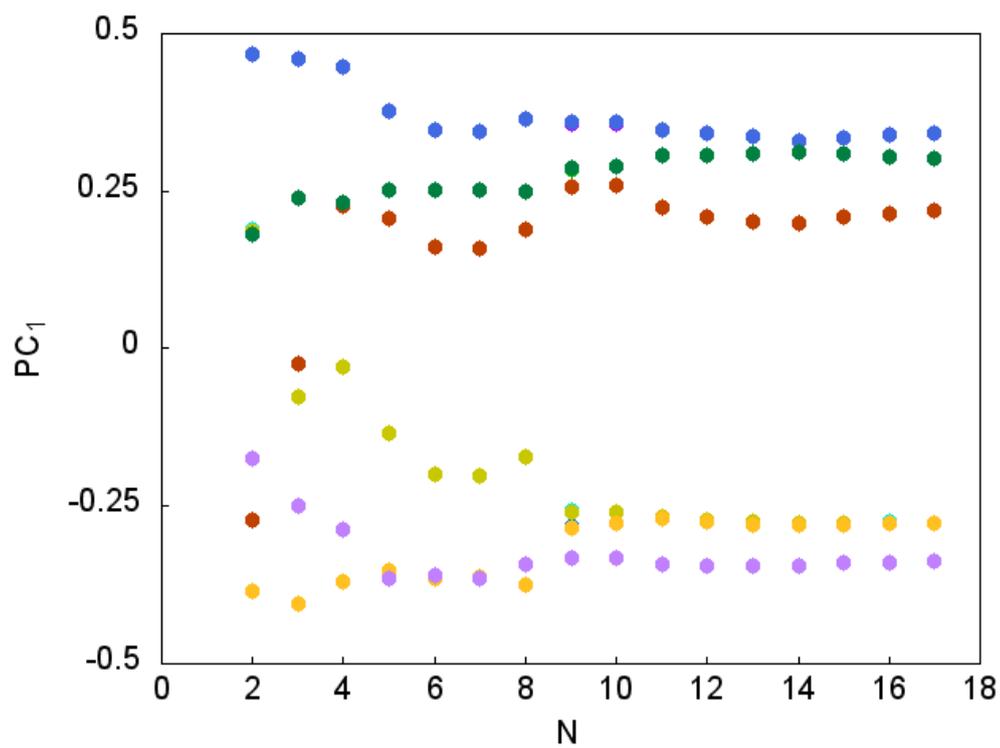
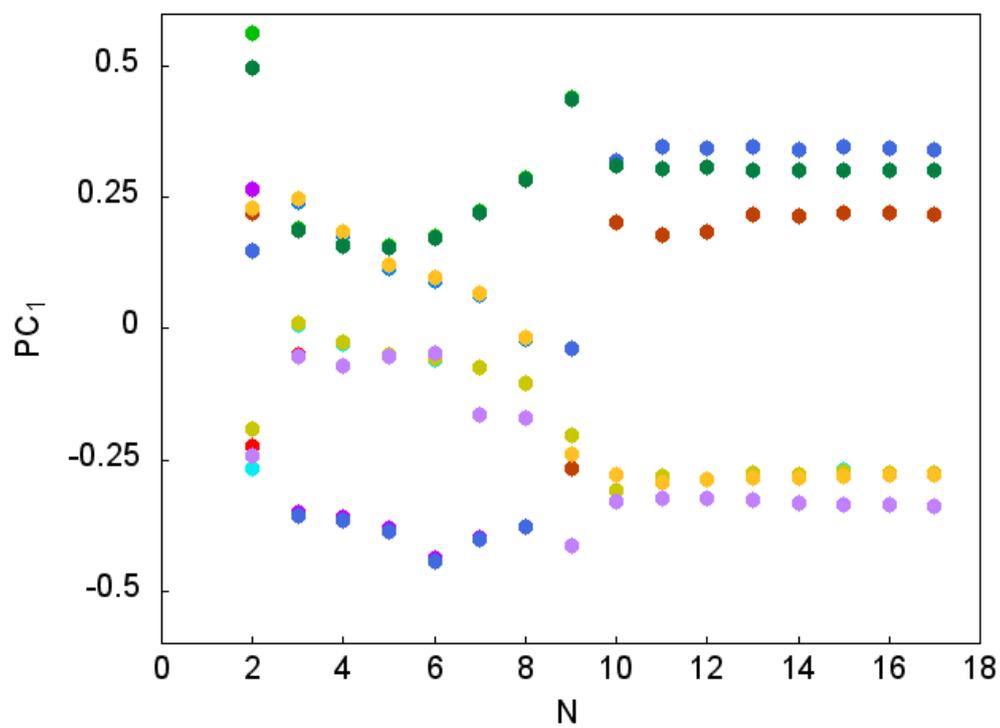
Figura 4.11: PC_1 para o grupo gradual

Como podemos ver, a PC_1 não sofre grande variação mesmo com poucas moléculas no grupo. Vemos que, com menos de cinco moléculas temos uma variação mais acentuada, porém, olhando para a organização das moléculas, as quatro primeiras são quinóides e a quinta é uma aromática. Por este motivo, repetimos a análise para o grupo com diferentes organizações. Primeiro começamos adicionando todas as quinóides, como podemos ver no gráfico da Figura 4.12.

Nota-se que a variação da PC_1 foi mais acentuada enquanto tínhamos apenas moléculas quinóides no grupo, porém, ao acrescentarmos as moléculas aromáticas a variação é pequena.

Para sabermos se este é um problema com as moléculas quinóides, fizemos a análise com o grupo invertido, começando pelas aromáticas e depois acrescentando as quinóides.

No gráfico da Figura 4.13 percebe-se que este comportamento se repete com as moléculas aromáticas também. Com isso, podemos concluir que a PCA não apresenta bons resultados para quando temos apenas um tipo de molécula no grupo do cálculo, sejam elas quinóides ou aromáticas, onde em alguns casos temos até a in-

Figura 4.12: PC_1 para o grupo de quinóides para aromáticas.Figura 4.13: PC_1 para o grupo de aromáticas para quinóides.

versão de sinal de alguns coeficientes. Porém, para sabermos se teríamos uma quantidade mínima devemos refazer esta análise, entretanto, agora devemos acrescentar as moléculas de forma alternada.

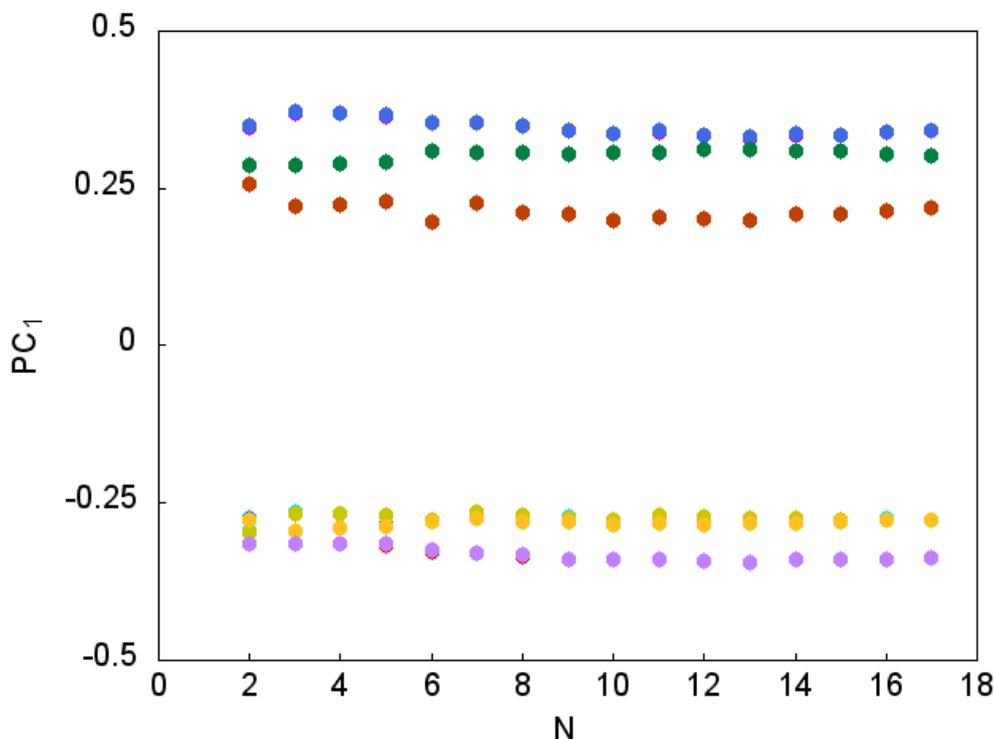


Figura 4.14: PC_1 para o grupo alternado.

Com o gráfico da Figura 4.14, pode-se ver que a PC_1 praticamente não variou mesmo quando o grupo continha apenas duas moléculas, com isso, podemos concluir que a PCA tem um problema para quando temos apenas um tipo de moléculas, porém, em relação a quantidade nos parece que não temos limitações mínimas, já que, a PC_1 apresenta pouca variação para qualquer quantidade desde que, tenhamos os dois tipos de moléculas presentes no grupo para realizarmos a análise.

Pensando nisso, fizemos uma análise similar, mas comparando o valor do BLA_{PCA} de uma mesma molécula, para ver como a variação da PC_1 influencia neste valor.

Vemos na Figura 4.15 um gráfico para a variação do valor do BLA da primeira molécula do grupo a medida que fomos aumentando o grupo.

Na Figura 4.16, vemos como se dá este comportamento para o grupo começando apenas com moléculas quinóides e acrescentando as moléculas aromáticas posteriormente.

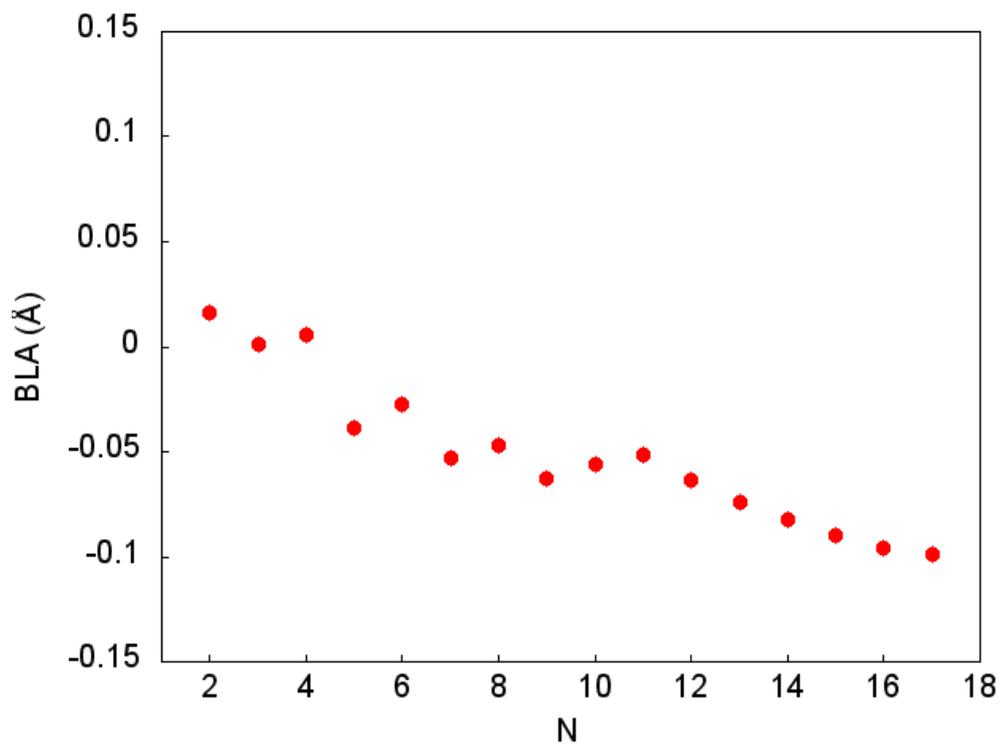


Figura 4.15: *BLA* da molécula 1 ordem aleatória inicial

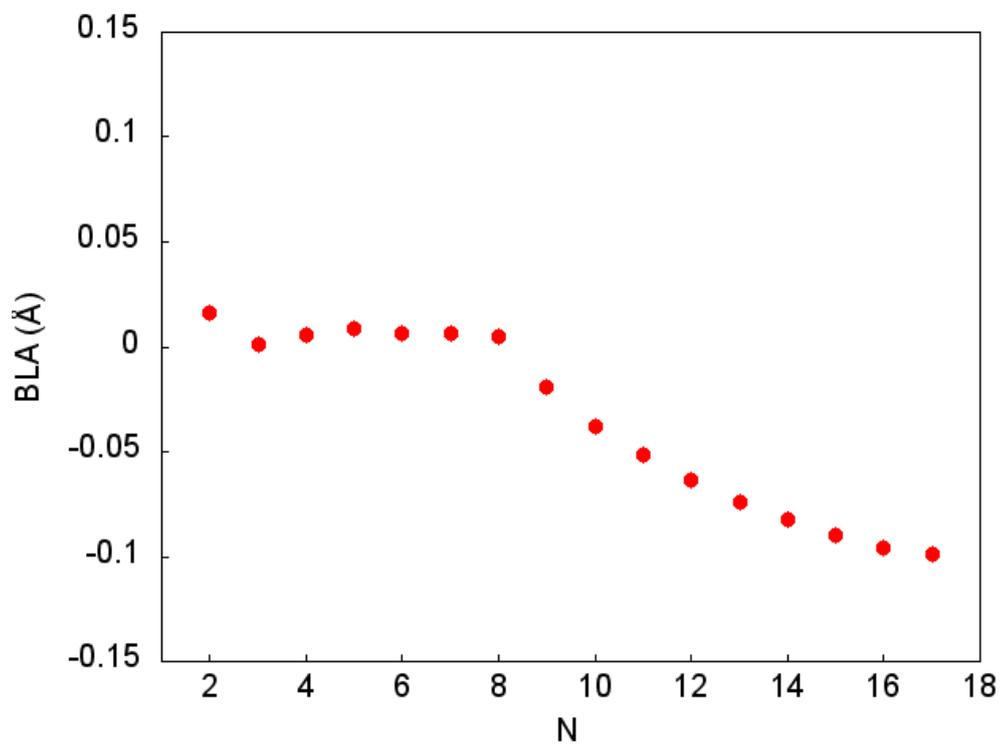


Figura 4.16: *BLA* da molécula 1 crescendo de quinóides para aromáticas

Vemos que a ordem não importa, sempre que adicionamos uma molécula aromática temos um decréscimo no *BLA* da primeira molécula, ao olharmos para as Figuras 4.16, 4.17 e 4.18, percebemos que este padrão se mantém. Podemos ver claramente nas Figuras 4.16 e 4.17 um padrão que começa a estabilizar a curva a partir de um certo número de moléculas e, na Figura 4.18, quando fomos crescendo o grupo de moléculas de forma alternada entre aromáticas e quinóides, percebemos um padrão linear no crescimento do valor do *BLA* da mesma molécula. Porém apesar deste comportamento, vemos que o valor em si não sofre grandes alterações já que a escala onde ele varia é bem pequena.

Com isto, podemos corroborar com os conhecimentos experimentais que já diziam que as moléculas quinóides são de fato diferentes das moléculas aromáticas no que se refere ao *BLA* [48].

Com a análise feita neste trabalho, conseguimos mostrar como a *PCA* pode ser um trunfo para o cálculo de *BLA*, que além de ficar próxima dos valores da forma já utilizada, ainda apresenta outras características interessantes e únicas, como a averiguação de quais ligações são mais importantes dentro da cadeia molecular. Vimos também que a quantidade de moléculas não importa, mas sim a variação do tipo de moléculas presentes no grupo, já que, para grupos com moléculas do mesmo tipo, mostramos que a *PCA* não é um método eficaz.

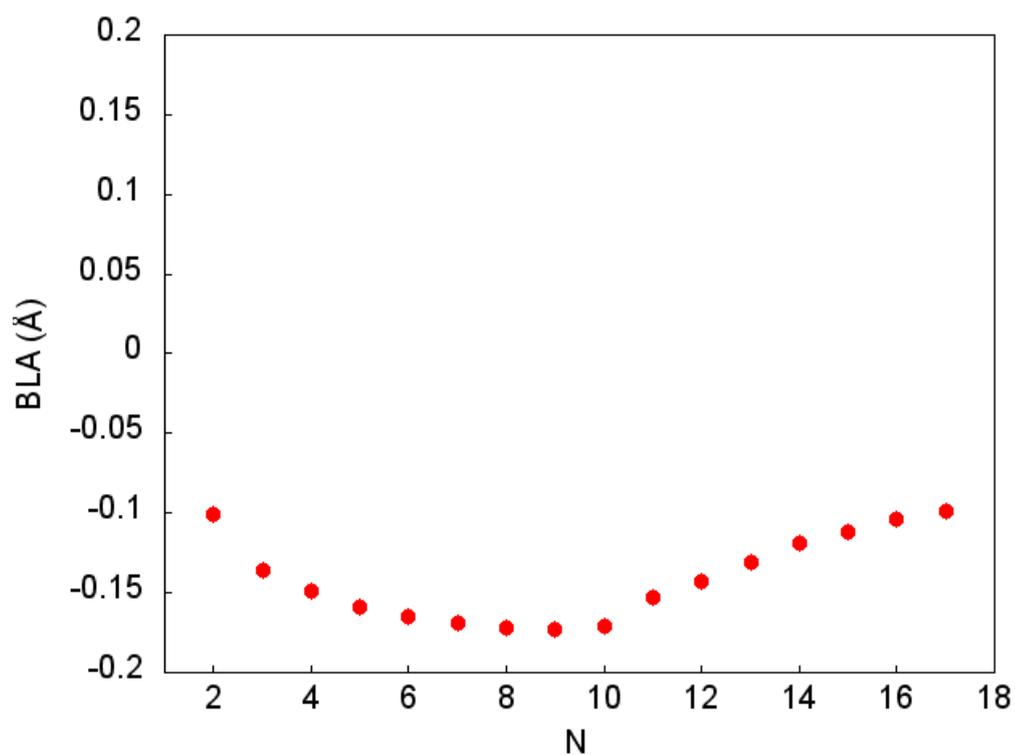


Figura 4.17: *BLA* da molécula 1 crescendo de aromáticas para quinóides

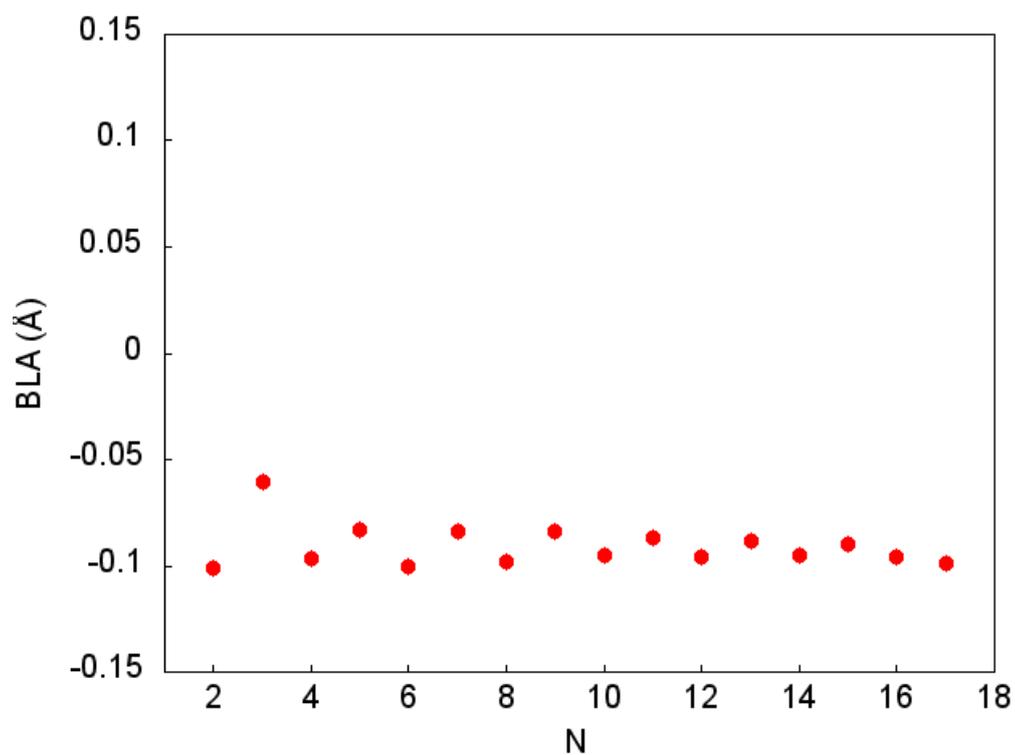


Figura 4.18: *BLA* da molécula 1 crescendo de forma alternada

Conclusão e Perspectivas

Neste trabalho utilizamos a *PCA* para analisar os comprimentos de ligação de um grupo de 17 moléculas de tertiofeno com o propósito de fazer comparações com o parâmetro estrutural *BLA*. Para o cálculo de otimização das geometrias utilizamos o software *Gaussian 09*, usando o método *DFT* de funcional híbrido *B3LYP* juntamente com a base *6-31G(d,p)*.

Para o comprimento das ligações duplas encontramos um valor de $\approx 1,437\text{\AA}$ e para as ligações simples de $\approx 1,377\text{\AA}$, o que está de acordo com outros resultados teóricos e experimentais.

Ao realizar o procedimento, em uma matriz composta pelos comprimentos das ligações C-C e C-S, mostramos como as ligações C-S são pouco relevantes para o valor da *BLA* pelo fato de apresentarem pouca variação em seus comprimentos, o que resulta numa variância desprezível. Com isso mostramos uma das características interessantes da *PCA*, que pode nos mostrar quais ligações são mais importantes dentro da cadeia.

Efetuada o procedimento novamente apenas com as ligações C-C, mostramos como a PC_1 , a primeira linha da matriz de autovetores, concentra quase toda a variância e pode ser usada para calcular o BLA_{PCA} das moléculas.

Com os valores obtidos com a técnica da *PCA*, fizemos comparações com a forma mais conhecida de calcular o *BLA*. Mostrando como os resultados são próximos e, com a *PCA*, temos certas características que nos fornecem informações que o *BLA* por si só não consegue nos mostrar, como quais ligações seriam mais relevantes para o valor do parâmetro onde, mostramos que as ligações dos anéis da extremidade da cadeia principal tem maior variância que as ligações do anel central.

No fim do capítulo ?? levantamos uma pergunta relacionada à quantidade mínima de moléculas necessárias para realizar a *PCA*. Partindo desta dúvida, analisamos a variação do coeficiente gerado pela técnica, a PC_1 , e vimos que não temos de fato um número mínimo para realizar a análise, porém, precisamos de uma variedade de moléculas, já que, para os casos onde tínhamos somente um tipo de molécula, sejam elas quinóides ou aromáticas, a *PCA* não se mostrou eficiente, mostrando ter um defeito para grupos com apenas um tipo de molécula.

Discutimos também, como a quantidade de moléculas afeta o valor do BLA_{PCA} e, mostramos que a presença de moléculas quinóides e aromáticas no grupo, contribuem de forma diferente. Mostramos como a adição de moléculas quinóides faz o *BLA* aumentar e o acréscimo de moléculas aromáticas ao grupo faz com que o valor do parâmetro, na molécula usada como referência, diminua.

Com isso, concluímos nosso trabalho atestando o sucesso do uso da *PCA* para encontrarmos um parâmetro comparado ao *BLA* porém, com características adicionais muito interessantes que nos dão informações importantes sobre um dado grupo de moléculas.

Neste trabalho utilizamos um grupo de moléculas simples, onde os resultados obtidos já eram esperados, com o propósito de atestar o *PCA* para este tipo de análise. Porém, existem sistemas em que não está claro quais ligações devem ser consideradas no cálculo da *BLA*. Um deles é o das ftalocianinas por exemplo (ver Figura 4.19).

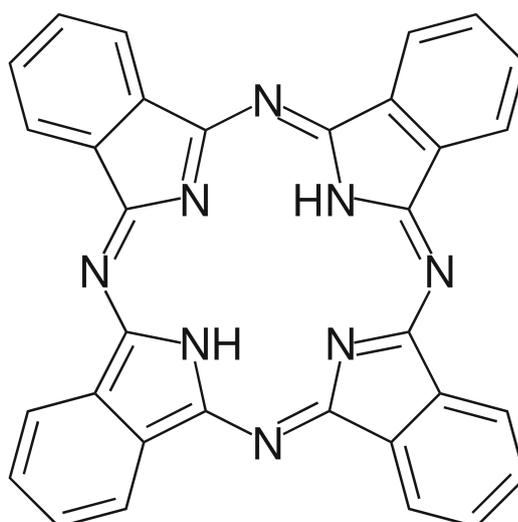


Figura 4.19: Exemplo de ftalocianina

Para casos como este, uma técnica como a *PCA*, pode nos ajudar a identificar quais ligações são mais relevantes para esta molécula.

Apêndice A

Código do *Mathematica* da *BLA-PCA*

Abaixo temos os códigos do *Mathematica 5* para os cálculos realizados neste trabalho com os comprimentos de ligação entre os doze átomos de carbono das dezesete moléculas e seus respectivos *outputs*.

```
data = {{1.43234, 1.36334, 1.43397, 1.38552, 1.44287, 1.38519, 1.44287, 1.38552, 1.43397, 1.36334, 1.43234},
        {1.43787, 1.35763, 1.44610, 1.37076, 1.43692, 1.39376, 1.43693, 1.37076, 1.44609, 1.35763, 1.43786},
        {1.42472, 1.37011, 1.42503, 1.39460, 1.43251, 1.39291, 1.43251, 1.39460, 1.42503, 1.37012, 1.42472},
        {1.43962, 1.36038, 1.43816, 1.37691, 1.43321, 1.36521, 1.43321, 1.37691, 1.43816, 1.36038, 1.43962},
        {1.36865, 1.42109, 1.37836, 1.45573, 1.38321, 1.43730, 1.38292, 1.45617, 1.37805, 1.42134, 1.36872},
        {1.44840, 1.35526, 1.44345, 1.37738, 1.44652, 1.38055, 1.44652, 1.37738, 1.44345, 1.35526, 1.44840},
        {1.36778, 1.42326, 1.37894, 1.45675, 1.38218, 1.43815, 1.38218, 1.45675, 1.37894, 1.42326, 1.36778},
        {1.43385, 1.36423, 1.43076, 1.37929, 1.42461, 1.37907, 1.42460, 1.37929, 1.43077, 1.36423, 1.43385},
        {1.36640, 1.42899, 1.38736, 1.44367, 1.38295, 1.41117, 1.38296, 1.44367, 1.38737, 1.42899, 1.36641},
        {1.44110, 1.36124, 1.43443, 1.38411, 1.44077, 1.38549, 1.44077, 1.38411, 1.43443, 1.36124, 1.44110},
        {1.43390, 1.36539, 1.42811, 1.38948, 1.43670, 1.38918, 1.43670, 1.38948, 1.42811, 1.36539, 1.43390},
        {1.36498, 1.42365, 1.37774, 1.44728, 1.37907, 1.41620, 1.37907, 1.44728, 1.37774, 1.42365, 1.36498},
        {1.36864, 1.42574, 1.37541, 1.44456, 1.37900, 1.41750, 1.37879, 1.44464, 1.37547, 1.42570, 1.36875},
        {1.36709, 1.42865, 1.37893, 1.43887, 1.38013, 1.41955, 1.38010, 1.43891, 1.37893, 1.42861, 1.36710},
        {1.37787, 1.41769, 1.37982, 1.45364, 1.38361, 1.43712, 1.38361, 1.45364, 1.37982, 1.41770, 1.37787},
        {1.37188, 1.41320, 1.38717, 1.45272, 1.38577, 1.43441, 1.38577, 1.45272, 1.38717, 1.41320, 1.37188},
        {1.39399, 1.39904, 1.39530, 1.44382, 1.39339, 1.43031, 1.39333, 1.44390, 1.39524, 1.39909, 1.39397}}
```

```
zdata = ZeroMean[data]
```

```
c = CovarianceMatrix[zdata]
```

```
PCA = Eigenvectors[c]
```

$$PC_1 = -PCA[[1]]$$

$$PCABLA = zdata.PC1$$

```

{{0.0300, -0.0295, 0.0269, -0.0318, 0.0344, -0.0214, 0.0344, -0.0318, 0.0269, -0.0295, 0.0300},
 {0.0355, -0.0352, 0.0390, -0.0465, 0.0284, -0.0128, 0.0285, -0.0466, 0.0391, -0.0352, 0.0355},
 {0.0224, -0.0227, 0.0180, -0.0227, 0.0240, -0.0137, 0.0241, -0.0227, 0.0180, -0.0227, 0.0224},
 {0.0373, -0.0324, 0.0311, -0.0404, 0.0247, -0.0414, 0.0248, -0.0404, 0.0311, -0.0325, 0.0373},
 {-0.0336, 0.0282, -0.0286, 0.0383, -0.0252, 0.0306, -0.0254, 0.0387, -0.0289, 0.0284, -0.0335},
 {0.0461, -0.0376, 0.0364, -0.0399, 0.0380, -0.0261, 0.0381, -0.0400, 0.0364, -0.0376, 0.0460},
 {-0.034, 0.0303, -0.0280, 0.0393, -0.0262, 0.0314, -0.0262, 0.0393, -0.0280, 0.0303, -0.0345},
 {0.0315, -0.0286, 0.0237, -0.0380, 0.0161, -0.0275, 0.0162, -0.0381, 0.0237, -0.0286, 0.0315},
 {-0.0359, 0.0361, -0.0196, 0.0263, -0.0254, 0.0045, -0.0254, 0.0262, -0.0196, 0.0361, -0.0358},
 {0.0387, -0.0316, 0.0274, -0.0332, 0.0323, -0.0211, 0.0323, -0.0332, 0.0274, -0.0316, 0.0387},
 {0.0316, -0.0274, 0.0211, -0.0278, 0.0282, -0.0174, 0.0282, -0.0279, 0.0211, -0.0274, 0.0315},
 {-0.0373, 0.0307, -0.0292, 0.0299, -0.0293, 0.0095, -0.0293, 0.0298, -0.0292, 0.0307, -0.0373},
 {-0.0336, 0.0328, -0.0315, 0.0272, -0.0294, 0.0108, -0.0296, 0.0272, -0.0315, 0.0328, -0.0335},
 {-0.0352, 0.0357, -0.0280, 0.0215, -0.0283, 0.0129, -0.0283, 0.0215, -0.0280, 0.0357, -0.0352},
 {-0.0244, 0.0248, -0.0271, 0.0362, -0.0248, 0.0304, -0.0247, 0.0362, -0.0271, 0.0248, -0.0244},
 {-0.0304, 0.0203, -0.0198, 0.0353, -0.0226, 0.0277, -0.0226, 0.0353, -0.0198, 0.0203, -0.0304},
 {-0.0083, 0.0061, -0.0117, 0.0264, -0.0150, 0.0236, -0.0150, 0.0265, -0.0117, 0.0062, -0.0083}}

```

```

{{0.0011, -0.0010, 0.0009, -0.0011, 0.0009, -0.0007, 0.0009, -0.0011, 0.0009, -0.0010, 0.0011},
 {-0.0017, 0.0009, -0.0008, 0.0009, -0.0008, 0.0006, -0.0008, 0.0010, -0.0008, 0.0009, -0.0010},
 {0.0009, -0.0008, 0.0007, -0.0009, 0.0007, -0.0005, 0.0007, -0.0009, 0.0007, -0.0008, 0.0009},
 {-0.0011, 0.0009, -0.0009, 0.0012, -0.0009, 0.0007, -0.0009, 0.0012, -0.0009, 0.0009, -0.0011},
 {0.0009, -0.0008, 0.0007, -0.0009, 0.0007, -0.0005, 0.0007, -0.0009, 0.0007, -0.0008, 0.0009},
 {-0.0007, 0.0006, -0.0005, 0.0007, -0.0005, 0.0005, -0.0005, 0.0007, -0.0005, 0.0006, -0.0007},
 {0.0009, -0.0008, 0.0007, -0.0009, 0.0007, -0.0005, 0.0007, -0.0009, 0.0007, -0.0008, 0.0009},
 {-0.0011, 0.0010, -0.0009, 0.0012, -0.0009, 0.0007, -0.0009, 0.0012, -0.0009, 0.0010, -0.0011},
 {0.0009, -0.0008, 0.0007, -0.0009, 0.0007, -0.0005, 0.0007, -0.0009, 0.0007, -0.0008, 0.0009},
 {-0.0010, 0.0009, -0.0008, 0.0009, -0.0008, 0.0006, -0.0008, 0.0010, -0.0008, 0.0009, -0.0010},
 {0.0011, -0.0010, 0.0009, -0.0011, 0.0009, -0.0007, 0.0009, -0.0011, 0.0009, -0.0010, 0.0011}}

```

{0.3379, -0.3027, 0.2771, -0.3417, 0.2761, -0.2191, 0.2765, -0.3421, 0.2772, -0.3029, 0.3377},
{0.1944, -0.2648, 0.0367, 0.3671, 0.1620, 0.6778, 0.1616, 0.3688, 0.0346, -0.2630, 0.1937},
{-0.1212, -0.0531, 0.2863, -0.3260, -0.3877, 0.5370, -0.3866, -0.3260, 0.2860, -0.0532, -0.1220},
{0.3559, -0.2456, -0.2002, 0.0978, -0.4810, -0.2502, -0.4825, 0.0979, -0.1997, -0.2456, 0.3559},
{-0.1642, -0.0974, 0.5057, 0.3585, -0.1178, -0.3662, -0.1124, 0.3546, 0.5055, -0.0965, -0.1668},
{-0.4237, -0.5167, -0.2092, -0.0690, 0.0473, -0.0855, 0.0445, -0.0661, -0.2116, -0.5148, -0.4231},
{-0.0727, -0.2682, 0.3305, -0.4662, 0.3095, -0.0021, -0.3109, 0.4671, -0.3252, 0.2721, 0.0738},
{-0.3011, 0.1830, -0.3621, 0.0138, 0.4928, 0.0001, -0.4930, -0.0112, 0.3669, -0.1837, 0.2998},
{-0.2893, -0.5579, -0.0554, 0.3178, -0.0176, 0.0004, 0.0175, -0.3188, 0.0545, 0.5583, 0.2896},
{-0.5651, 0.2320, 0.2044, -0.0864, -0.2779, -0.0001, 0.2774, 0.0871, -0.2042, -0.2326, 0.5654},
{0.0304, 0.1670, 0.4635, 0.4160, 0.2886, 0.0001, -0.2886, -0.4161, -0.4633, -0.1665, -0.0303}}

{{-0.33792}, {0.30276}, {-0.27712}, {0.34174}, {-0.27618}, {0.21916}, {-0.27652},
{0.34213}, {-0.27729}, {0.30291}, {-0.33779}}

{{-0.09867}, {-0.11752}, {-0.07084}, {-0.11264}, {0.10295}, {-0.12828}, {0.10561},
{-0.09288}, {0.09006}, {-0.10585}, {-0.08825}, {0.09883}, {0.09741}, {0.09419},
{0.09179}, {0.08662}, {0.04749}}

Referências Bibliográficas

- [1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2(6), pp. 559–572, 1901.
- [2] J. A. Falcon and K. A. Berglund, "In situ monitoring of antisolvent addition crystallization with principal components analysis of raman spectra," *Crystal Growth e Design*, vol. 4(3), pp. 457–463, 2004.
- [3] A. G. Ryder, G. M. O'Connor, and T. J. Glynn, "Quantitative analysis of cocaine in solid mixtures using raman spectroscopy and chemometric methods," *Journal of Raman Spectroscopy*, vol. 31(3), pp. 221–227, 2000.
- [4] A. G. Ryder, "Classification of narcotics in solid mixtures using principal component analysis and raman spectroscopy," *Journal of Forensic Sciences*, vol. 47(2), pp. 275–284, 2002.
- [5] A. Lorén, C. Eliasson, M. Josefson, K. V. G. K. Murty, M. Käll, J. Abrahamsson, and K. Abrahamsson, "Feasibility of quantitative determination of doxorubicin with surface-enhanced raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 32, pp. 971–974, 2001.
- [6] J. K. Chapina and M. A. L. Nicolelis, "Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations," *J Neurosci Methods*, vol. 94(1), pp. 121–140, 1999.
- [7] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, pp. 904–909, 2006.

- [8] S. R. Marder, C. B. Gorman, B. G. Tiemann, J. W. Perry, G. Bourhill, and K. Mansour, "Relation between bond-length alternation and second electronic hyperpolarizability of conjugated organic molecules," *Science*, vol. 261(5118), pp. 186–189, 1993.
- [9] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24(6 and 7), pp. 417–441 and 498–520, 1933.
- [10] K. Karhunen, "Über lineare methoden in der wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae, Series A 1, Math. Phys.*, vol. 37, pp. 3–79, 1946.
- [11] M. Loève, "Fonctions aléatoires de second ordre," *Rev. Sci*, vol. 84(4), pp. 195–206, 1946.
- [12] M. Loève, "Probability theory: Foundations, random sequencies," *D. Van Nostrand Company*, 1955.
- [13] H. C. Longuet-Higgins and L. Salem, "The alternation of bond lengths in long conjugated chain molecules," *The Royal Society*, vol. 251(1265), pp. 172–185, 1959.
- [14] W. P. Su, J. R. Schrieffer, and A. J. Heeger, "Solitons in polyacetylene," *Physical Review Letters*, vol. 42(25), pp. 1698–1701, 1979.
- [15] C.-Q. Wu, X. Sun, and K. Nasu, "Electron correlation and bond alternation in polymers," *Physical Review Letters*, vol. 59(7), pp. 831–834, 1987.
- [16] J. L. Brédas and A. J. Heeger, "Comment on electron correlation and bond alternation in polymers," *Physical Review Letters*, vol. 63(22), p. 2534, 1989.
- [17] S. Capponi, N. Guihery, J.-P. Malrieu, B. Miguel, and D. Poilblanc, "Bond alternation of polyacetylene as a spin-peierls distortion," *Chemical Physics Letters*, vol. 255(4), pp. 238–243, 1996.
- [18] R. E. Peierls, *Quantum Theory of Solids*. Clarendon, Oxford, 1955.
- [19] D. Jacquemin, A. Femenias, H. Chermette, I. Ciofini, C. Adamo, J.-M. André, , and E. A. Perpète, "Assessment of several hybrid dft functionals for the evaluation of

- bond length alternation of increasingly long oligomers," *J. Phys. Chem. A*, vol. 110, pp. 5952 – 5959, 2006.
- [20] C. H. Choi, M. Kertesz, and A. Karpfen, "The effects of electron correlation on the degree of bond alternation and electronic structure of oligomers of polyacetylene," *L. Chem. Phys.*, vol. 107, p. 6712, 1997.
- [21] W. P. Su, J. R. Schrieffer, and A. J. Heeger, "Soliton excitations in polyacetylene," *Phys. Rev. B*, vol. 22, p. 2099, 1983.
- [22] C. S. Yannoni and T. C. Clarke, "Molecular geometry of cis- and trans-polyacetylene by nutation nmr spectroscopy," *Phys. Rev. Lett.*, vol. 51, p. 1191, 1983.
- [23] J. L. Brédas, C. Adant, P. Tackx, A. Persoons, and B. M. Pierce, "Third-order nonlinear optical response in organic materials: Theoretical and experimental aspects," *Chem. Rev.*, vol. 94, p. 243, 1994.
- [24] J. L. Brédas, "Relationship between band gap and bond length alternation in organic conjugated polymers," *J. Chem. Phys.*, vol. 82, p. 3808, 1985.
- [25] G. Bourhill, J.-L. Brédas, L.-T. Cheng, S. R. Marder, F. Meyers, J. W. Perry, and B. G. Tiemann, "Experimental demonstration of the dependence of the first hyperpolarizability of donor-acceptor-substituted polyenes on the ground-state polarization and bond length alternation," *J. Am. Chem. Soc.*, vol. 116, p. 2619, 1994.
- [26] S. R. Marder, C. B. Gorman, F. Meyers, J. W. Perry, G. Bourhill, J.-L. Brédas, and B. M. Pierce, "A unified description of linear and nonlinear polarization in organic polymethine dyes," *Science*, vol. 265, pp. 632–635, 1994.
- [27] W. Bartkowiak, R. Zaleśny, and J. Leszczynski, "Relation between bond-length alternation and two-photon absorption of a push-pull conjugated molecules: a quantum-chemical study," *Chemical Physics*, vol. 287(1-2), pp. 103–112, 2003.
- [28] T. Kogej, D. Beljonne, F. Meyers, J. Perry, S. Marder, and J. Brédas, "Mechanisms for enhancement of two-photon absorption in donor-acceptor conjugated chromophores," *Chemical Physics Letters*, vol. 298(1-3), pp. 1–6, 1998.

- [29] M. Takahashi and J. Paldus, "Bond-length alternation and vibrational spectra of polyacetylene," *Canadian Journal of Physics*, vol. 62(12), pp. 1226–1231, 1984.
- [30] S. R. Marder, J. W. Perry, B. G. Tiemann, C. B. Gorman, S. Gilmour, S. L. Bid-
dle, and G. Bourhill, "Direct observation of reduced bond-length alternation in do-
nor/acceptor polyenes," *J. Am. Chem. Soc.*, vol. 115, p. 2524, 1993.
- [31] D. Jacquemin and C. Adamo, "Bond length alternation of conjugated oligomers:
Wave function and dft benchmarks," *J. Chem. Theory Comput.*, vol. 7, pp. 369 –
376, 2010.
- [32] P. Horsch, "Correlation effects on bond alternation in polyacetylene," *Phys. Rev.
B*, vol. 24, p. 7351, 1981.
- [33] J. Paldus and E. Chin, "Bond length alternation in cyclic polyenes. i. restricted
hartree-fock method," *Int. J. Quantum Chem.*, vol. 24, p. 373, 1983.
- [34] M. Takahashi and J. Paldus, "Bond length alternation in cyclic polyenes. vi. cou-
pled cluster approach with wannier orbital basis," *Int. J. Quantum Chem.*, vol. 28,
pp. 459–479, 1985.
- [35] H. Kuhn, "A quantum-mechanical theory of light absorption of organic dyes and
similar compounds," *J. Chem. Phys.*, vol. 17, p. 1198, 1949.
- [36] M. J. S. Dewar, "Colour and constitution. part iii. polyphenyls, polyenes, and
phenylpolynes; and the significance of cross-conjugation," *J. Chem. Soc.*,
pp. 3544–3550, 1952.
- [37] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136,
p. B864, 1964.
- [38] W. Kohn and L. Shan, "Self-consistent equations including exchange and correla-
tion effects," *Phys. Rev.*, vol. 140, p. A1133, 1965.
- [39] A. Becke, "A new mixing of hartree-fock and local density-functional theories," *J.
Chem. Phys.*, vol. 98, pp. 1372–1377, 1993.
- [40] A. Becke, "Density-functional exchange-energy approximation with correct
asymptotic behavior," *Phys. Rev. A*, vol. 38, p. 3098, 1988.

- [41] A. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *J. Chem. Phys.*, vol. 98, p. 5648, 1993.
- [42] C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, p. 785, 1988.
- [43] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, "Self-consistent molecular orbital methods. xxiii. a polarization-type basis set for second-row elements," *J. Chem. Phys.*, vol. 77, p. 3654, 1982.
- [44] M. J. Frisch, J. A. Pople, and J. S. Binkley, "Self-consistent molecular orbital methods. xxv. supplementary functions for gaussian basis sets," *J. Chem. Phys.*, vol. 80, p. 3265, 1984.
- [45] J. S. Márquez and M. F. Núñez, "On the determination of bond lengths by ab initio methods: estimation of errors and some improvements," *Journal of Molecular Structure*, vol. 624, no. 1-3, pp. 239–249, 2003.
- [46] W. Kohn, A. Becke, and R. G. Parr, "Density functional theory of electronic structure," *J. Phys. Chem.*, vol. 100(31), pp. 12974–12980, 1996.
- [47] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Canadian Journal of Physics*, vol. 58(8), pp. 1200–1211, 1980.
- [48] I. F. Perepichka and D. F. Perepichka, *Handbook of Thiophene-based Materials: Applications in Organic Electronics and Photonics*. Wiley, 2009.