



**RECONSTRUÇÃO DE CAMPOS
ELETROMAGNÉTICOS UTILIZANDO
REGULARIZAÇÃO DE TIKHONOV-PHILLIPS
E INTELIGÊNCIA ARTIFICIAL**

HELTON ALVES DE AZEVEDO

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**RECONSTRUÇÃO DE CAMPOS
ELETROMAGNÉTICOS UTILIZANDO
REGULARIZAÇÃO DE TIKHONOV-PHILLIPS
E INTELIGÊNCIA ARTIFICIAL**

HELTON ALVES DE AZEVEDO

Orientador: PROF. DR. SÉBASTIEN R.M.J RONDINEAU, ENE/UNB

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

**PUBLICAÇÃO PPGENE.DM - PPGE 806/23
BRASÍLIA-DF, 28 DE NOVEMBRO DE 2023.**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**RECONSTRUÇÃO DE CAMPOS
ELETROMAGNÉTICOS UTILIZANDO
REGULARIZAÇÃO DE TIKHONOV-PHILLIPS
E INTELIGÊNCIA ARTIFICIAL**

HELTON ALVES DE AZEVEDO

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

APROVADA POR:

Prof. Dr. Sébastien R.M.J Rondineau, ENE/UnB
Orientador

Prof. Dr. Daniel Orquiza de Carvalho , ENE/UnB
Examinador interno

Prof. Dr. Gefeson Mendes Pacheco, ITA
Examinador Externo

BRASÍLIA, 28 DE MOVEMBRO DE 2023.

FICHA CATALOGRÁFICA

HELTON ALVES DE AZEVEDO

Reconstrução de Campos Eletromagnéticos Utilizando Regularização de Tikhonov-Phillips e Inteligência Artificial

2023xv, 147p., 201x297 mm

(ENE/FT/UnB, Mestre, Engenharia Elétrica, 2023)

Dissertação de Mestrado - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia Elétrica

REFERÊNCIA BIBLIOGRÁFICA

HELTON ALVES DE AZEVEDO (2023) Reconstrução de Campos Eletromagnéticos Utilizando Regularização de Tikhonov-Phillips e Inteligência Artificial. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGEE 806/23, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 147p.

CESSÃO DE DIREITOS

AUTOR: Helton Alves de Azevedo

TÍTULO: Reconstrução de Campos Eletromagnéticos Utilizando Regularização de Tikhonov-Phillips e Inteligência Artificial.

GRAU: Mestre ANO: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor se reserva a outros direitos de publicação e nenhuma parte desta dissertação de Mestrado pode ser reproduzida sem a autorização por escrito do autor.

Helton Alves de Azevedo

Rodovia DF 150, Km 12. Quadra 02, Conjunto B Casa 18. Fercal-DF

Agradecimentos

A execução deste trabalho foi algo ao mesmo tempo gratificante e desafiador. Gratificante por eu estar alcançando mais uma meta estabelecida em minha vida e por contribuir com a ciência brasileira. Já ao que compete ao desafio ele não é composto somente a execução deste trabalho, mas também ao contexto com o qual ele foi executado. Assim como toda a humanidade fui afetado pelos eventos mais recentes da pandemia. Felizmente não tive problemas graves diretos ou indiretos devido a tal situação, mas o meu ingresso ao programa de mestrado foi durante o pico da mesma, e devido a má gestão dos governos nas esferas distritais e federais em muitos aspectos, e falo isso com tristeza, tais atitudes afetaram a meu rendimento na execução desse trabalho, e por consequência na minha saúde mental como um todo.

Por esse motivo, gostaria agradecer aos meus pais, Magna Táthia Alves dos Santos e Honório Rosa de Azevedo, que desde o início me deram todo o suporte necessário para chegar onde eu cheguei, e principalmente por me ensinaram a persistir até o fim. Agradeço também ao meu professor orientador, Dr. Sébastien R.M.J Rondineau que por me dar a oportunidade de trabalhar com tema e de me ajudar da melhor maneira possível nas esferas acadêmica, profissional e pessoal. Agradeço também a Universidade de Brasília e ao Instituto de Eletrônica e Telecomunicações de Rennes por todo o apoio financeiro, estrutural e acadêmico.

Agradeço especialmente aos meus amigos Elpídio Cândido, Daniele Dias e Danovan Martins pelas conversas, apoio, revisões e dicas durante a execução deste trabalho. Agradeço ao meu amigo André Felipe Pereira de Araújo por me ouvir quando precisei desabafar. Agradeço também ao Dr. Luiz Frezzatto pelas dicas e por me acalmar nos momentos difíceis. Também estendo meus agradecimentos ao Gustavo Rinaldi, Gustavo Takaoka, Maurício Varella, Karoline Pache, Cassio Cioni e Mateus Alves por me tirarem de casa quando eu mais precisava, pelas conversas, risos e pela força e apoio. Agradeço especialmente também ao Valter Santiago por todo suporte e conselhos que ele me deu. Por fim, caso você tenha a oportunidade de ler este trabalho e seu nome não esteja aqui, mas que de alguma forma contribuiu para a execução do mesmo, fica aqui o meu sincero muito obrigado.

*"Eu acredito que às vezes são as pessoas que ninguém espera nada
que fazem as coisas que ninguém consegue imaginar."*

(Alan Turing)

Resumo

A caracterização experimental de antenas vem ao longo dos anos tornando-se um importante campo de estudo do eletromagnetismo aplicado. Isso fica mais evidente devido ao fato que o projeto de elementos irradiantes estão cada vez mais complexos devido ao significativo aumento das demandas eletrônicas, influência do ambiente circundante, especificações de operação e banda e funcionalidades adicionais aumentam a ainda mais a necessidade de investigações e estudos mais profundos dos elementos radiantes. No entanto, apesar do fato de tal área ter uma vasta quantidade de estudos conduzidos, ela ainda possui desafios, pois o problema de reconstrução do campo magnético por caracterização pode ser considerado como NP-difícil e não convexo, e a quantidade de e as estratégias de aquisição de dados possuem um grande impacto no sucesso final do processo.

Neste contexto, esse trabalho apresenta novas estratégias de reconstrução do campo eletromagnético. A primeira delas utiliza a técnica de regularização de Thikonov-Philips para determinar a melhor malha retangular de amostragem, utilizando o número de condicionamento e a quantidade de amostras como métrica. Tal estratégia permite a reconstrução do campo eletromagnético com dados faltantes e ruidosos, em que os efeitos de Gibbs nos limites da janela utilizada são praticamente desprezíveis, mitigando o efeito do ruído numérico presente nas amostras experimentais, obtendo assim estimadores ótimos, estáveis e com baixa carga computacional. O desenvolvimento de tal técnica foi um trabalho conjunto feito pela Universidade de Brasília (Brasil), Universidade de Rennes (França) e a Universidade de Cassino (Itália). O autor desta dissertação de mestrado foi o responsável pela implementação e análise de resultados. Já a segunda é uma nova abordagem de reconstrução do campo eletromagnético baseada na resolução do problema de recuperação de fase utilizando inteligência artificial: ela permite obter a distribuição de fase do campo eletromagnético em uma determinada esfera usando apenas os dados de amplitude.

Para a primeira técnica uma análise profunda composta de comparações entre os campos medido e reconstruído, propagação do ruído e independência garantem o erro de reconstrução mínimo. Já para o segundo item deste trabalho, os resultados obtidos mostram que os modelos de inteligência artificial conseguem recuperar a distribuição de fase do campo eletromagnético para reconstrução de maneira satisfatória. Os resultados também foram consolidados com análises estatísticas robustas.

Palavras-chave: Amostragem eletromagnética, modos esféricos, detecção comprimida, re-

gularização de Tikhonov, estimação, filtragem, recuperação de fase, aprendizagem de máquina, regressão.

Abstract

In recent years, experimental antenna characterization has become an essential field of study regarding applied electromagnetism. The current required performance for antenna design with significant antenna electric dimensions has added complexity regarding environmental influence, operational frequencies, electronic specifications, and additional features like multiple beam patterns and versatile operation modes for further studies and investigations about radiation sources. This characteristic increases the need for further studies about radiation sources. Nonetheless, although applied electromagnetism shows many studies about antenna testing, the field presents challenges to overcome. The field reconstruction problem is an NP-hard and non-convex problem, where the number of samples and the acquisition data strategies play an essential role in the electromagnetic field reconstruction.

Consequently, this work shows new approaches for electromagnetic field reconstruction. The first is an interpolation scheme based on the Tikhonov-Phillips regulation process with the optimization of field parallel-equatorial sampling positions, where the condition number is the metric for this goal. The presented approach allows field reconstruction with limited and noisy data, where the Gibbs effect in the window boundary is negligible with noise reduction and optimum estimators for a well-posed formulation. This technique is a project with contributions made by University of Brasília (Brazil), University of Rennes (France), and University of Cassino (Italy), where the author of this dissertation was responsible for the computational implementation and result analyses.

Furthermore, using artificial intelligence models, a new electromagnetic field reconstruction based on the phase retrieval problem uses the electromagnetic field amplitude data and its angular positions as input. The results have been successfully compared with real-world measurements for the first one. Also, robust noise propagation and independence analyses ensure that the technique reconstructs the field efficiently.

The second approach discussed in this work shows the feasibility of artificial intelligence models to retrieve the phase distribution from amplitude data to field reconstruction with satisfactory performance. This result was also grounded in robust statistical analyses.

Keywords: EM sampling, spherical modes, compressed sensing, Tikhonov regularization, estimation, filtering, phase retrieval, machine learning, regression.

SUMÁRIO

AGRADECIMENTOS	I
RESUMO	III
ABSTRACT	V
1 INTRODUÇÃO	1
1.1 MOTIVAÇÕES	1
1.2 OBJETIVOS.....	2
1.2.1 OBJETIVOS ESPECÍFICOS	2
1.3 ORGANIZAÇÃO	3
2 REFERENCIAL TEÓRICO	4
2.1 RECONSTRUÇÃO DO CAMPO ELETROMAGNÉTICO	4
2.1.1 ESTADO-DA-ARTE	4
2.1.2 TÉCNICAS DE MEDIÇÃO DO CAMPO ELETROMAGNÉTICO	12
2.1.3 MÉTRICAS DE AVALIAÇÃO	16
2.2 RECUPERAÇÃO DE FASE	18
2.2.1 ESTADO DA ARTE	18
2.3 APRENDIZADO DE MÁQUINA	32
2.3.1 APRENDIZAGEM NÃO SUPERVISIONADA	33
2.3.2 APREDIZAGEM POR REFORÇO	35
2.3.3 APRENDIZAGEM SUPERVISIONADA	38
2.3.4 VALIDAÇÃO CRUZADA	51
2.3.5 TESTES DE HIPÓTESES	54
3 RECONSTRUÇÃO DO CAMPO ELETROMAGNÉTICO PARA DADOS FALTANTES E RUIDOSOS	58
3.1 EXPANSÃO EM CAMPO DISTANTE	58
3.2 REGULARIZAÇÃO DE TIKHONOV-PHILLIPS	60
3.2.1 MATRIZ DE POSTO COMPLETO	60
3.2.2 MAPA ANGULAR PARCIAL: MATRIZ DE POSTO INCOMPLETO	61
3.3 AMOSTRAGEM DO CAMPO	61
3.3.1 AMOSTRAGEM EM ϕ	61

3.3.2	AMOSTRAGEM EM θ	64
3.3.3	NÚMERO TOTAL DE AMOSTRAS	68
3.4	ANÁLISE ESTATÍSTICA	69
3.4.1	PROPAGAÇÃO DO RUÍDO	69
3.4.2	DIAGNÓSTICO RESIDUAL	69
3.4.3	FILTRAGEM DE COEFICIENTES ESFÉRICOS	70
4	RECONSTRUÇÃO DE CAMPO COM INTELIGÊNCIA ARTIFICIAL	72
4.1	CONSTRUÇÃO DA BASE DE DADOS	72
4.1.1	TRANSLAÇÃO DA SUPERFÍCIE DE AQUISIÇÃO	72
4.1.2	SEQUÊNCIA DE APRENDIZAGEM	75
4.2	PROCESSO DE OTIMIZAÇÃO, VALIDAÇÃO E OBTENÇÃO DOS MODELOS	78
4.2.1	OTIMIZAÇÃO POR <i>Grid Search</i>	78
4.2.2	VALIDAÇÃO ESTATÍSTICA DOS ALGORITMOS	78
4.2.3	OBTENÇÃO DOS MODELOS	80
4.2.4	VALIDAÇÃO ELETROMAGNÉTICA	80
5	APLICAÇÕES E RESULTADOS	82
5.1	RECONSTRUÇÃO DO CAMPO ELETROMAGNÉTICO PARA DADOS FALTANTES E RUIDOSOS	82
5.1.1	LENTE PLANA DE LUNEBURG	84
5.1.2	ANTENA CORNETA	84
5.1.3	<i>Reflectarray</i>	84
5.1.4	ANÁLISE DO RUÍDO RESIDUAL	86
5.1.5	COMPARAÇÃO ENTRE AS TÉCNICAS PROPOSTA E A DE HANSEN	87
5.2	RECONSTRUÇÃO DE CAMPO COM INTELIGÊNCIA ARTIFICIAL	87
5.2.1	ANTENA GUIA DE ONDA	89
5.2.2	ANTENA CORNETA	93
5.2.3	INFLUÊNCIA DO RUÍDO	95
	CONCLUSÃO	97
5.3	CONTRIBUIÇÕES DESTE TRABALHO	98
5.4	TRABALHOS FUTUROS	99
	BIBLIOGRAFIA	100
	APÊNDICES	109
5.5	APÊNDICE A: ELETROMAGNETISMO PARA ELEMENTOS RADIANTES ...	109
5.5.1	EQUAÇÕES DE MAXWELL	109
5.5.2	CAMPOS RADIANTES ATRAVÉS DE POTENCIAIS	110
5.5.3	CAMPOS RADIANTES ATRAVÉS DE EXPANSÕES DE ONDAS ESFÉRICAS	114
5.5.4	POTÊNCIA RADIADA	119
5.5.5	DIRETIVIDADE E GANHO	119

5.5.6	POLARIZAÇÃO.....	120
5.5.7	PADRÃO DE RADIAÇÃO	121
5.6	APÊNDICE B: SISTEMA LINEAR PARA EXPANSÃO EM ONDAS ESFÉ- RICAS	122
5.7	APÊNDICE C: PROPRIEDADES DE CAMPO DISTANTE DAS FUNÇÕES DE BASE ESFÉRICAS	125

LISTA DE FIGURAS

2.1	Representação geométrica das normas. A definição é feita na equação 2.6. Tal definição estende-se para $p \rightarrow \infty$. Para $0 \leq p \leq 1$ tal definição para de ser identificada como norma. O uso de l_0 resume-se a identificar o número de elementos não nulos de \mathbf{x} , sendo a verdadeira medida do nível de esparsidade da solução. Apesar das regiões delimitadas pelas métricas apresentarem formas geométricas diversas, elas são comumente chamadas de bolas na literatura [1].....	7
2.2	Visualização da esparsidade induzida pelo processo de minimização por norma l_0 de sistemas lineares.	9
2.3	Diagrama de transição para sistemas esparsos. Quanto mais esparsos for o vetor \mathbf{x} menos pontos experimentais são necessários para recuperá-lo.....	10
2.4	Exemplo de caracterização experimental de campo radiado baseado no princípio da Equivalência.	11
2.5	Regiões de medições em torno da antena sob estudo de esfera mínima r_0 e λ o comprimento de onda de operação. $\delta(\lambda)$ denota a margem da região reativa [2], a sua determinação é extremamente difícil.	12
2.6	Ilustração de um caracterizador de antena de alcances compactos CATR utilizando um sistema refletor. Tal instrumento permite fazer caracterizações de campo distante em espaços reduzidos.	13
2.7	Ilustração da aquisição de dados com simetria planar. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.....	14
2.8	Ilustração da aquisição de dados com simetria cilíndrica. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.....	14
2.9	Ilustração da aquisição de dados com simetria esférica. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.....	15
2.10	Fluxo de atualização do algoritmo iterativo de entrada e saída. Caso na propagação o vetor \mathbf{x} satisfaça as restrições o campo receberá o resultado de tal operação, caso contrário ele será atualizado com o resultado da propagação adicionado de um fator de correção.	21

2.11 Fluxo de do algoritmo de entrada e saída. Caso na propagação o vetor x satisfaça as restrições o campo receberá o resultado de tal operação, caso contrário ele será atualizado com valor nulo.	23
2.12 Técnicas de aquisição de padrões de iluminação para resolução do problema de recuperação de fase via relaxação convexa. A figura (a) refere-se à técnica de ocultamento ótico, a figura (b) à técnica de grades óticas, figura (c) ao uso de pictografia e a figura (d) ilustra o uso de iluminações oblíquas.	24
2.13 Técnicas de aquisição da amplitude de ondas eletromagnéticas para resolução do problema de recuperação de fase via relaxação convexa. Tal técnica também é utilizada nos algoritmos iterativos do tipo Gerchberg-Saxton. A diferença reside na quantidade, enquanto na técnica interativas são utilizadas duas esferas, na relaxação convexa utiliza-se múltiplas medições.	25
2.14 Exemplificação do funcionamento dos algoritmos de agrupamento por aglomeração, figura (a) e divisão, figura (b). Tais técnicas são chamadas de algoritmos hierárquicos pois organizam os dados de maneira bem definida em agrupamentos específicos.	33
2.15 Exemplificação de uma estrutura em cascata gerada pelos os algoritmos de agrupamento hierárquico. Tal árvore permite a identificação do nível de interrelação entre os diferentes grupos e o grau de proximidade entre eles.	34
2.16 Princípio de funcionamento do algoritmo <i>k-Means</i> . Os dados são organizados de maneira aleatória em um número finito k de grupos com centro C e raio r . À medida que as iterações ocorrem o cálculo da distância entre os elementos e os centros dos grupos é calculada e a sua localização é alterada para uma mais adequada dependendo do resultado.	35
2.17 Fluxograma que ilustra o processo de aprendizagem por reforço. O agente ou inteligência artificial observa o estado do ambiente, interage com ele por meio de uma ação, recebendo uma recompensa positiva ou negativa dependendo se a mesmo foi ou não bem-sucedida, dessa maneira o agente consegue aprender a executar uma determinada tarefa para qual está sendo treinado.	36
2.18 Ilustração do processo de aprendizagem supervisionada para as aplicações de classificação, figura (a), e regressão, figura (b). Neste tipo de aprendizagem de máquina é necessário indicar quais são as variáveis independentes e dependente no processo de treinamento.	39
2.19 Ilustração da obtenção das predições tanto para classificação e regressão, figuras (a) e (b) respectivamente. Após a obtenção do modelo na etapa de treinamento é possível usar o mesmo para executar predições para novas entradas.	40
2.20 Ilustração do processo de aprendizagem de máquina para regressão linear. A linha preta simboliza o melhor resultado do processo de aprendizagem obtido a partir da resolução das equações 2.78 e 2.79 para os pontos experimentais em azul.	41

2.21	Esquemática de funcionamento do algoritmo de árvores de decisão tanto para classificação, figura (a), e regressão, figura (b). A partir do nó principal os dados são separados a partir da confirmação ou rejeição da regra seguindo a equação 2.80.....	43
2.22	Princípio de funcionamento do algoritmo de floresta randômica. A predição $hat{y}$ é obtida pela combinação das predições intermediárias h_n produzidas pelas T_n árvores de decisão.	44
2.23	Esquemática de funcionamento do algoritmo KNN tanto para classificação, figura (a), e regressão, figura (b). Para a classificação, os dados são agrupados em seções semelhantes dado a distância dos pontos ao centro C . Já para a regressão o algoritmo obtém o seu modelo a partir da otimização das distâncias entre os pontos experimentais.	45
2.24	Princípio de funcionamento do algoritmo de SVM. O algoritmo SVM busca pelo melhor hiperplano que separa os dados em dois subconjuntos maximizando as margens, em que os pontos experimentais próximos as margens são chamados de vetores de suporte.....	48
2.25	Ilustração de um neurônio biológico, figura (a), e um artificial, figura (b). O neurônio é a célula nervosa responsável pela transmissão de impulsos elétricos no sistema nervoso, sendo ativado ou não dependendo do estímulo aplicado. Já o neurônio artificial é um modelo matemático que busca reproduzir tal comportamento.....	49
2.26	Redes neurais. A figura (a) ilustra uma rede formada pelas células nervosas, enquanto a figura (b) a rede artificial. A camada de entrada é a responsável por receber as entradas x_n , enquanto as camadas ocultas são responsáveis pela adaptabilidade da rede aos dados. Finalmente a camada de saída é responsável por entregar a saída \hat{y} , ou seja, a predição.....	50
2.27	Procedimento comum para execução do treinamento dos modelos de inteligência artificial. A base de dados é separada uma parte para treinamento e outra para teste, em uma proporção de 70% e 30% respectivamente para a maioria das aplicações.....	51
2.28	Validação cruzada. Essa técnica de validação de modelos de inteligência artificial de aprendizagem supervisionada verifica a robustez e imunidade contra o problema de <i>overfitting</i> separando a base de dados em um número limitado de grupos, utilizando um deles para validação e os outros para treinamento. Tal procedimento é executado de maneira iterativa intercalando qual dos grupos criados será utilizado para validação para uma certa quantidade de testes.	52
3.1	Número de condicionamento $c_{\Theta, N}$ para o ângulo θ . Essa figura mostra que D_7 é a melhor distribuição com valor máximo de 6,15 e convergindo para 1,00 a medida que a ordem de truncamento N aumenta.	67

4.1	Esquemática do posicionamento entre os dois sistemas de coordenadas O e O'	73
4.2	Efeito de translação nas coordenadas de uma superfície S quando translada de O' para O . Nota-se que o raio é influenciado pelas coordenadas esféricas dos pontos de S em O' . A figura a mostra o efeito caso a translação for no eixo x , a figura b caso for feita no eixo y e a figura c mostra o caso da translação no eixo z	75
4.3	Ilustração do de aquisição de dados de amplitude para o treinamento os algoritmos de aprendizado para resolver o problema de recuperação de fase. A utilização de uma esfera permite uma visão completa do campo eletromagnético, além do fato de as estratégias de medição de campo em tal superfície é bem documentada e validada [3].	77
4.4	Visualização da base dados utilizada, em que $\Upsilon_{d_K} (\Lambda_{d_K}, \theta_{d_K}, \phi_{d_K})$ simboliza os dados de amplitude e posições angulares, e $\mathbf{f}_{d_K} (r_K, \theta_{d_K}, \phi_{d_K})$ os dados de fase. Os dados são adquiridos a partir do esquema de medição mostrada na figura 4.3.	78
4.5	Esquemática do processo de <i>grid search</i> . Uma certa configuração dos parâmetros de um algoritmo é selecionada e inserida no mesmo. Logo em seguida ele é treinado, testado e o resultado é armazenado para futuras comparações e análises.	79
4.6	Esquemática de como os algoritmos são selecionados para a geração de modelos. Somente aqueles que passam nos testes ANOVA e Tukey para uma métrica de adequação e uma de erro são selecionados para gerar modelos.....	80
5.1	Impacto do fator de regularização α na caracterização experimental executada para as antenas SGH, LL e RA. A figura (a) mostra tanto o quanto a variância do ruído residual σ^2 aumenta à medida que α aumenta, mantendo-se mais ou menos constante para $\alpha < -100$ dB. Já a relação sinal ruído decresce à medida que α aumenta. Já a figura (b) mostra como a potência de ruído só começa a aumentar para $\alpha > 0$ dB, enquanto potência possui um comportamento muito similar à SNR.	83
5.2	Mapas bidimensionais de amplitude e fase das componentes do campo distante E_θ e E_ϕ para comparação de campo elétrico mensurado e reconstruído para os parâmetros mostrados na tabela 5.1. A figura mostra os resultados para a antena corneta operando na banda Ka a 26 GHz no caracterizador CATR da IETR. Nota-se efeito de Gibbs no ponto de descontinuidade em $\theta = \pi/2$ onde o campo elétrico foi zerado devido a ausência de medições.	85

5.3	Mapas bidimensionais de amplitude e fase das componentes do campo distante E_θ e E_ϕ que compara o campo elétrico mensurado e o reconstruído para os parâmetros mostrados na tabela 5.1. Os resultados foram obtidos para o <i>reflectarray</i> de 193 células projetado pela Thalès Alenia Space no projeto R3MEMS. Nota-se efeito de Gibbs no ponto de descontinuidade em $\theta = \pi/2$ e isso ocorre que pois os dados campo elétrico são zerados devido a ausência de medições.....	85
5.4	Mapas bidirecionais das componentes E_θ and E_ϕ em campo distante comparando amplitude e fase reconstruída e medida utilizando os parâmetros de amostragem da tabela 5.1 para a antena de lente plana de Luneburg operando à 12 GHz. Essa antena foi projetada na estrutura MERLIN, um laboratório conjunto entre Thalès Alenia Space e o IETR.	86
5.5	Padrão de campo distante para ambas as planos componentes E_θ e E_ϕ do campo elétrico obtido baseado nos parâmetros mostrados na tabela 5.1, para (a) para antena de lente metálica de Luneburg projetada na estrutura de MERLIN para operar a 12 GHz, (b) mostra os resultados para uma antena corneta operando na banda Ka em 26 GHz projetada no IETR e (c) um <i>reflectarray</i> composto de 193 células, projetado para operar a 12,GHz pela Thalès Alenia Space na estrutura do projeto R3MEMS. Ambas as antenas atreladas às figuras (b) e (c) foram caracterizadas em um setor angular limitado.....	88
5.6	Densidade de probabilidade de erro para as três antenas caracterizas: a figura (a) para antena SGH, a figura (b) para a antena LL e a figura (c) para antena RA. Tais resultados mostram que a interpolação proposta é eficiente, e que o ruído associado já uma distribuição do tipo hiper gaussiana.	89
5.7	Estimação da magnitude de auto correlação, seguindo o cenário estabelecido na tabela 5.2 para as três antenas caracterizadas: a figura (a) para a SGH, a figura (b) para a antena LL e a figura (c) para antena RA. A matriz $\Gamma_{\epsilon\epsilon}$ é predominantemente diagonal para todas as antenas, justificando a independência estatística das amostras do ruído residual.	89
5.8	Campo total normalizado para: (a) reconstruído com a técnica proposta, (b) mensurado e (c) reconstruído com a técnica de Hansen. Tais resultados são para o <i>reflectarray</i> projetado na estrutura Thalès Alenia Space no projeto R3MEMS e caracterizado em um setor angular limitado. Devido ao fato da técnica de Hansen ser baseada na transformada de Fourier e necessitar que o campo fora de região interesse seja zerado para obter os coeficientes modais, ela sofre do efeito Gibbs nos limites da janela de amostragem e consequentemente tem um aumento do erro residual. A técnica proposta porém não necessita de tal manobra. Tal comportamento fica evidenciado pelas figuras (d) e (e) que correspondem ao erro residual das técnicas propostas e de Hansen respectivamente.....	90

5.9	Distribuição de fase para esfera centrada com raio de aquisição de 20% de λ para a antena guia de onda operando a 10 GHz. Ambas as componentes E_θ e E_ϕ são mostradas no setor circular em $\phi = 120^\circ$. O algoritmo DT obteve uma de ENL = -12,38 dB, RF de ENL = -13,35 dB e KNN de ENL = -13,83 dB.....	92
5.10	Fase recupera no setor circular $\phi = 120^\circ$ para ambas as componentes E_θ e E_ϕ para a antena corneta operando a 10 GHz com esfera de aquisição centrada com raio de caracterização de 20% da distância de campo distante. O modelo de árvore de decisão alcançou ENL = -13,76 dB, o algoritmo de floresta randômica alcançou ENL = -13,31 dB e o algoritmo KNN obteve ENL = -13,14 dB.....	94
5.11	Influência do ruído na recuperação de fase efetuada pelos algoritmos DT, RF e KNN para antena guia de onda, figura (a), e corneta figura (b) ambas operando a 10 GHz. Os algoritmos se perdem bastante para SNR abaixo de 0 dB, ao passo que para uma SNR acima de 5 dB a reconstrução torna-se satisfatória, aproximando-se dos resultados ideais à medida que ela aumenta. .	96
5.12	Representação do efeito não instantâneo da aplicação das fontes ρ e \mathbf{A} em uma determinada localização do espaço, demonstrando que a propagação das ondas eletromagnéticas é um efeito causal. Adaptado de [2].....	111
5.13	Aproximação em campo distante de \mathbf{E} e \mathbf{H} é descrita pelas funções potenciais V e \mathbf{A} . Geralmente somente o primeiro termo da expansão de R é suficiente para tal análise.	113
5.14	A aplicação da aproximação de campo distante depende do ponto de observação, uma convenção muito utilizada é a distância de Fraunhofer em $r = 2L^2/\lambda$, em que L é a maior dimensão da antena.	114
5.15	Os resultados da expansão das ondas esféricas implica que tanto campo elétrico \mathbf{E} quanto magnético \mathbf{H} são emanados a partir de uma esfera raio r_0 que engloba toda a antena independente de sua posição, tal esfera é chamada de esfera mínima [4].	115
5.16	Relação entre os índices m e n , em m pode ser considerado uma função n , do tipo $m = 2n + 1$. Nesta ilustração a ordem é truncada para $N = 3$. A quantidade de modos esféricos está intimamente ligada com o valor n , mais precisamente $Q_n = n(n + 2)$	118
5.17	Polarizações da onda eletromagnética. Figura (a) ilustra a polarização linear, a figura (b) a polarização circular, que pode ser do tipo LHCP e RHCP, e a figura (c) a polarização elíptica.....	121
5.18	Exemplos de diagramas de radiação obtidos via <i>software</i> Ansys Electronics para uma antena corneta operando a 10 GHz. A figura (a) mostra a disposição do campo no eixo θ enquanto (b) mostra sua disposição tridimensional.....	122

LISTA DE TABELAS

5.1	Resumo dos parâmetros de amostragem utilizados para as antenas caracterizadas em um mapa angular completo.....	82
5.2	Resumo dos parâmetros eletromagnéticos diretividade D , potência radiada em campo distante P_R^∞ , potência de ruído e relação sinal ruído SNR para as antenas SGH, LL e RA.....	83
5.3	Resultados da análise do ruído residual para as antenas SGH operando a 26 GHz, LL operando a 12 GHz e RA também a 12 GHz.....	83
5.4	Resumo dos resultados do processo CV para as métricas investigadas utilizando 10 grupos e 30 testes por iteração para a antena guia de onda.	91
5.5	Resultados de Shapiro-Wilk para as métricas investigadas para as componentes E_θ and E_ϕ . Nível de significância de 5%.....	91
5.6	Resumo dos resultados do teste de hipótese de Tukey para os melhores algoritmos para ambas as componentes E_θ e E_ϕ . Nível de significância de 5%.....	92
5.7	Resumo dos resultados para a validação cruzada para as métricas investigadas para ambas as componentes E_θ e E_ϕ . Foram utilizadas 10 divisões e 30 testes por iteração.....	93
5.8	Resultados para o teste de Shapiro-Wilk para as métricas investigadas para ambas as componentes E_θ and E_ϕ . Com nível de significância de 5%.....	93

LISTA DE TERMOS E SIGLAS

<i>AUT</i>	Antenna Under Test
<i>BIO</i>	Basic Input-Output Algorithm
<i>BP</i>	Basis Pursuit
<i>BPDN</i>	Basis Pursuit Denoising
<i>CATR</i>	Compact Antenna Test Range
<i>CV</i>	Cross Validation
<i>DT</i>	Decision Tree
<i>ENL</i>	Equivalent Noise Level
<i>ER</i>	Error Reduction Algorithm
<i>GB</i>	Gerchberg-Saxton
<i>HIO</i>	Hybrid Input-Output
<i>HSD</i>	Honesty Significant Difference
<i>HTP</i>	Hard Thresholding Pursuit
<i>IA</i>	Inteligência Artificial
<i>IoT</i>	Internet of Things
<i>KNN</i>	K-Nearest Neighbors
<i>LHCP</i>	Left Handed Polarization
<i>LHCP</i>	Linear Regression
<i>LL</i>	Luneburg Lens
<i>M2M</i>	Machine to Machine
<i>MAE</i>	Mean Absolute Error

<i>MDP</i>	Markov Decision Process
<i>ML</i>	Machine Learning
<i>MLP</i>	Multi-Layer Perceptron
<i>MoM</i>	Method of Moments
<i>MP</i>	Matching Pursuit
<i>MRS</i>	Minimal Redundant Sampling
<i>MSE</i>	Minimal Squared Error
<i>NFFT</i>	Near Field to Far Field Transformation
<i>OMP</i>	Ortogonal Matching Pursuit
<i>PC</i>	Phasecut Algorithm
<i>PL</i>	Phase Lift Algoritm
<i>PM</i>	Phasemax Algorithm
<i>PR</i>	Phase Retrieval
<i>PTD</i>	Phase Transition Diagrams
<i>RF</i>	Random Forest
<i>RHCP</i>	Right Handed Circular Polarization
<i>RL</i>	Reinforcement Learning
<i>SDP</i>	Semidefinte Programming
<i>SGH</i>	Standard Gain Horn Antenna
<i>SL</i>	Supervised Learnig
<i>SVM</i>	Support Vector Machine
<i>SWE</i>	Spherical Wave Expasions
<i>TE</i>	Transverse Electric
<i>TM</i>	Transverse Magnetic
<i>UL</i>	Unserpervised Learning

Capítulo 1

Introdução

1.1 Motivações

A caracterização experimental do campo eletromagnético é um ramo muito importante tanto para a abordagem teórica e prática do eletromagnetismo aplicado. A relevância desse tópico levou a muitos esforços para o desenvolvimento de técnicas otimizadas que vão desde a amostragem até aquelas focadas na recuperação das informações necessárias para reconstrução do campo eletromagnético.

O processo pode ser resumido como a identificação de um elemento dentro do grupo de todos os campos eletromagnéticos que a fonte pode irradiar. Mais especificamente, as estratégias de amostragem são baseadas em algum conhecimento *a-priori* sobre a configuração do campo do elemento sob investigação, sendo sempre algo que define parcialmente o conjunto de possíveis configurações. Dessa maneira, o problema de amostragem possui como objetivo selecionar o melhor conjunto de medições que define o elemento sob estudo de maneira única a partir do grupo de configurações de campo.

Outra consideração a ser feita é que quanto mais precisa é a descrição do antena a ser caracterizada menor o espaço de procura, e por consequência menor o número de amostradas [5, 6, 7, 8, 9]. Além disso, os dados obtidos experimentalmente estão sempre acompanhados de ruído e incertezas, o que leva o processo de recuperação do campo nem sempre admitir uma solução. Isso acontece pois o ruído modifica as amostras de tal maneira que elas não irão mais pertencer ao conjunto de campos que a fonte pode irradiar. Logo, a principal tarefa é encontrar uma *quasi*-solução, ou seja, uma solução que seja satisfatória o suficiente mesmo na presença do ruído e das incertezas. Um outro conjunto de técnicas para reconstrução do campo eletromagnético é a recuperação de fase. Tais técnicas possuem como objetivo recuperar a distribuição de fase do campo eletromagnético usando apenas medições em amplitude [10, 11, 12, 13, 14, 15].

Neste contexto, este trabalho apresenta um método de reconstrução de campo baseado na regularização de Tikhonov-Phillips para expansões modais esféricas, que usa o número

de condicionamento como a métrica principal para determinar a menor malha amostragem, além de garantir um processo de inversão rápido, robusto, estável e computacionalmente eficiente, permitindo a reconstrução do campo com dados incompletos e corrompidos por ruído. O desenvolvimento, implementação e validação de tal técnica é um trabalho conjunto internacional conduzido pela Universidade de Brasília (Brasil), Universidade de Rennes (França) e Universidade de Cassino (Itália), em que o autor deste trabalho ficou responsável pela implementação da técnica e pela análise dos resultados gerados.

Não obstante, uma nova abordagem de recuperação de fase é proposta usando algoritmos de aprendizagem de máquina supervisionada. Essa técnica usa um conjunto de superfícies esféricas nas quais as amplitudes de cada polarização estão amostradas para gerar uma base de dados robusta para treinamento desses algoritmos afim de recuperar a distribuição de fase a partir de medições em amplitude. Até onde o autor sabe, essa é a primeira vez que essa abordagem foi utilizada para resolver o problema de recuperação de fase.

1.2 Objetivos

Esse trabalho propõe duas técnicas de reconstrução: a primeira baseada na regularização de Tikhonov-Phillips, em que o número de condicionamento dos sistemas matriciais gerados é a métrica responsável por determinar a menor malha retangular de amostragem, além de garantir um sistema estável e computacionalmente eficiente para aplicações com dados faltantes e ruidosos. Já o segundo utiliza a capacidade de aprendizado dos algoritmos de aprendizagem supervisionada para resolver o problema de recuperação de fase de maneira rápida em condições de ruído controlado.

1.2.1 Objetivos Específicos

Os objetivos específicos são:

- Apresentar um nova técnica de caracterização de antenas baseada na regularização de Thikonov-Philips para determinar o melhor conjunto de pontos utilizando o número de condicionamento como métrica para determinar a melhor malha retangular de amostragem para reconstruir o campo de maneira computacionalmente estável, sendo aplicável para situações com dados faltantes e ruidosos;
- Utilizar algoritmos de inteligência artificial para resolver o problema de recuperação de fase, isso é reconstruir o campo eletromagnético usando apenas dados de amplitude e informações *a-priori*, recuperando a distribuição de fase de maneira rápida, em que a robustez mesmo em situações com dados com ruído e validada por análises estatísticas.

1.3 Organização

Este trabalho é organizado da seguinte maneira:

- Capítulo 1: o capítulo apresenta a introdução e as motivações deste trabalho;
- Capítulo 2: o capítulo apresenta todo o referencial teórico necessário para esse trabalho, apresentando conceitos sobre reconstrução de campo 2.1, problema de recuperação de fase 2.2 e aprendizagem de máquina 2.3;
- Capítulo 3: esse capítulo aborda toda a parte teórica da formulação da técnica de reconstrução de campo baseada na regularização de Thikonov-Philips para dados faltantes e ruidosos;
- Capítulo 4: esse capítulo apresenta o processo de construção da base de dados, bem como os processos de validação e seleção dos modelos de inteligência artificial para resolução do problema de recuperação de fase;
- Capítulo 5: apresenta todos resultados e discussões da implementação das técnicas apresentadas nesse trabalho. já as considerações finais são feitas no item 5.2.3.

Capítulo 2

Referencial Teórico

2.1 Reconstrução do Campo Eletromagnético

O procedimento de caracterização de antenas pode ser definido como processo de recuperar a informação necessária da antena sob teste, do inglês *Antenna Under Test* (AUT). Como discutido anteriormente em 5.5 um elemento irradiante pode ser descrito tanto pelas fontes ρ e \mathbf{J} tanto pelo coeficientes Q_{smn} ligados a expansão de ondas esféricas. Esta seção possui como objetivo mostrar as principais técnicas que foram desenvolvidas para a reconstrução de campo eletromagnético a partir de dados obtidos de maneira experimental, além de conceitos importantes para a realização de tal tarefa.

2.1.1 Estado-da-Arte

2.1.1.1 Reconstrução por Expansão de Ondas Esféricas

Como discutido em 5.5, os campos elétricos \mathbf{E} e magnético \mathbf{H} podem ser representados pela SWE. Sendo assim, alguns métodos de reconstrução de campo eletromagnético foram desenvolvidos ao longo das décadas identificar os coeficientes Q_{smn} . De acordo com [3] a escolha do método é baseada no número de pontos experimentais disponíveis com a quantidade de coeficientes a serem identificados $N_c = 2N(N + 2)$:

Método Analítico

Apresentado por [4], esse método identifica os coeficientes Q_{smn} da SWE usando projeções integrais de Fourier nos dados experimentais adquiridos. Considerando a hipótese de sinal de largura de banda limitada, os coeficientes podem ser recuperados a partir da transformada de Fourier e do teorema de amostragem de Nyquist para expansão de ondas esféricas

da seguinte maneira [4]:

$$Q_{smn}^{(3)} \tilde{R}_{sm}^{(4,\gamma)}(kr) + Q_{smn}^{(4)} \tilde{R}_{sm}^{(3,\gamma)}(kr) = \frac{\sqrt{\eta}}{k} (-1)^m \int_0^{2\pi} \int_0^\pi \mathbf{E}(r_e, \theta, \phi) \cdot \mathbf{F}_{s-mn}^{(\gamma)}(r_e, \theta, \phi) \dots \sin(\theta) d\theta d\phi, \quad (2.1)$$

em que \mathbf{F}_{s-mn} é uma função vetorial de onda, r_e o raio de aquisição e $\tilde{R}_{sm}^{c,\gamma}$ é definida como:

$$\tilde{R}_{sm}^{c,\gamma} = R_{sn}^{(c)}(kr) R_{sn}^{(\gamma)}(kr) + \delta_{s2} [n(n+1)] \frac{z_n^{(c)}(kr)}{kr} \frac{z_n^{(\gamma)}(kr)}{kr}, \quad (2.2)$$

em que ondas irradiadas e recebidas, $c = 3, 4$, são consideradas. O método analítico possui a característica de recuperar o campo radiado de maneira exata [3] sendo computacionalmente eficiente, mesmo para altas ordens N de truncamento. Contudo, devido ao fato de estar calcado no teorema de amostragem de Nyquist ele necessita de muitos dados experimentais.

Método Matricial

Como mostrado em 5.5, a expansão em ondas esféricas é uma combinação linear ponderada pelos coeficientes Q_{smn} . Essa característica permite escrever o campo eletromagnético em uma forma matricial como mostrado em [16], em que a prova detalha pode ser vista em 5.6. Nesta abordagem as direções esféricas são discretizadas, θ e ϕ , enquanto o processo de amostragem é realizado. Somente o campo irradiado, $c = 3$, é considerado e por esse motivo será omitido das funções vetoriais \mathbf{F}_{smn} . Além disso, o campo é amostrado em duas direções ortogonais ou tangenciais para P amostras em torno da superfície de medição com coordenadas (θ_p, ϕ_p) para $p = 1, 2, \dots, P$. Dessa maneira a SWE pode ser escrita como um sistema de equações do tipo [16]:

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \quad (2.3)$$

Para:

$$\mathbf{y} = \begin{bmatrix} E^\theta(r, \theta_1, \phi_1) \\ E^\phi(r, \theta_1, \phi_1) \\ \vdots \\ E^\theta(r, \theta_P, \phi_P) \\ E^\phi(r, \theta_P, \phi_P) \end{bmatrix}, \quad \mathbf{A} = \frac{k}{\sqrt{\eta}} \begin{bmatrix} \mathbf{F}_{1mn}^\theta & \mathbf{F}_{2mn}^\theta \\ \mathbf{F}_{1mn}^\phi & \mathbf{F}_{2mn}^\phi \end{bmatrix}, \quad \text{e} \quad \mathbf{x} = \begin{bmatrix} \mathbf{Q}_{1mn} \\ \mathbf{Q}_{2mn} \end{bmatrix}. \quad (2.4)$$

Onde $\mathbf{x} \in \mathbb{C}^{N_c \times 1}$, $\mathbf{A} \in \mathbb{C}^{P \times N_c}$ e $\mathbf{y} \in \mathbb{W}^{P \times 1}$, em que \mathbb{W} simboliza o fato de \mathbf{y} poder pertencer à \mathbb{R} ou \mathbb{C} dependendo de condições da aquisição dos pontos experimentais. Em \mathbf{A} , \mathbf{F}_{1mn}^α denota o bloco de colunas das componentes TE e \mathbf{F}_{2mn}^α o bloco das componentes TM, em que $\alpha = \theta, \phi$. Já em \mathbf{x} , \mathbf{Q}_{1mn} denota o bloco de linhas dos coeficientes Q_{smn} para as componentes TE e \mathbf{Q}_{2mn} para as componentes TM. A formulação matricial da expansão

das ondas esféricas permite o uso das bem conhecidas técnicas de resolução de sistema lineares por inversão [16, 3], como por exemplo o método de mínimos quadrados ou técnicas regularizadas por decomposição do valor singular. Os processos de inversão necessitam de uma quantidade de dados de pelo menos igual ou superior ao número de variáveis para produzir bons resultados, ou seja, $P \geq N_c$.

Pelas características do método matricial, a solução obtida para os coeficientes Q_{smn} é uma solução aproximada, mas possui a vantagem de utilizar metade de pontos experimentais quando comparado com o método analítico, ou seja:

$$P_M \approx \frac{P_A}{2}, \quad (2.5)$$

em que P_A simboliza a quantidade total de pontos experimentais necessários no método analítico. Adicionalmente, o método matricial é mais flexível no que diz respeito à amostragem dos dados, visto que [16] utilizou um método diferente da mostrada por [4] chamada de amostragem iglu, que será abordada com mais profundidade neste trabalho. Dessa maneira, a reconstrução SWE numérica é mais flexível quando se refere a amostragem [3], permitindo o uso de poucos pontos experimentais, mesmo que a solução ser aproximada. A estabilidade da solução depende da estratégia de inversão, podendo exigir maior procedimento experimental.

Método por Recuperação Esparsa

Apesar dos métodos matriciais performarem melhor para situações que $P \geq N_c$, em algumas situações isso não pode acontecer, sendo necessário trabalhar em cenários em que a quantidade de pontos experimentais é menor que o número de variáveis a serem recuperadas. Dessa maneira, o sistema linear mostrado na equação 2.3 é subdeterminado, apresentando um número infinito de soluções [17]. Nesse cenário é assumido que a solução verdadeira é esparsa, ou seja, \mathbf{x} é composto por um número grande zeros, o que permite utilizar um conjunto adequado de técnicas para recuperá-la [1, 18, 19]. Adicionalmente, o fato de o campo eletromagnético emitido pela AUT ser simétrico e regular, diminui de maneira considerável a complexidade do problema, permitindo o uso de abordagens baseadas em minimização de norma e regularização.

Minimização de Norma l_0

O método de minimização de normas é uma de maneiras de recuperação de soluções esparsas de sistemas subdeterminados. Primeiramente, o conceito generalizado de norma deve ser considerado. Considerando que l_p é a norma de um vetor \mathbf{x} e que $p \leq 1$ ela pode ser definida como [1]:

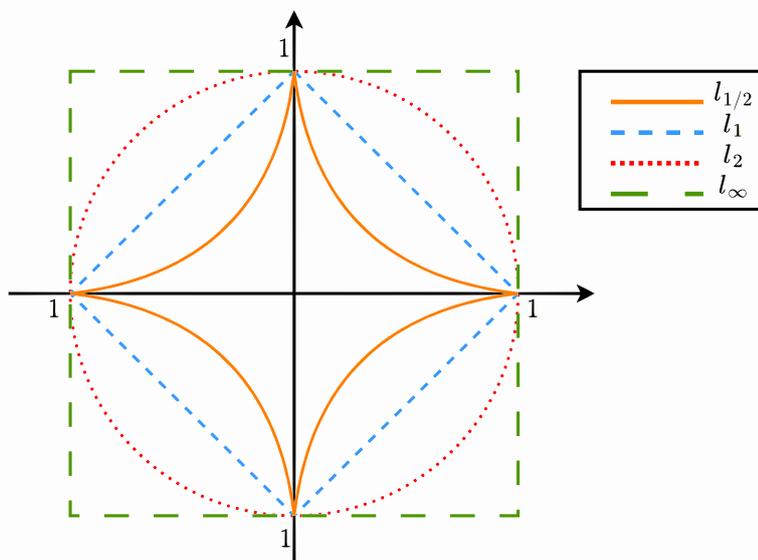
$$\|\mathbf{x}\|_p = \left(\sum x_i^p \right)^{1/p}. \quad (2.6)$$

Dessa maneira, considerando que o sistema linear mostrado na equação 2.3 possui uma solução esparsa, ela é dada por $\|\mathbf{x}\|_0$. Como o sistema possui um número infinito de soluções, o problema baseia-se em procurar no espaço de soluções aquela mais esparsa. Esse processo pode ser definido como um processo de minimização do tipo:

$$\begin{aligned} &\text{minimize} && \|\mathbf{x}\|_0, \\ &\text{tal que} && \mathbf{y} = \mathbf{Ax}. \end{aligned} \tag{2.7}$$

O problema mostrado na equação 2.8 é considerado como NP-difícil [19, 20], requerendo uma quantidade considerável de recursos computacionais para achar a solução adequada, até mesmo para problemas pequenos. Para contornar tal problema, muitas técnicas foram desenvolvidas para obter uma boa solução com menor consumo possível de tempo e de processamento computacional. As técnicas que utilizam algoritmos de vantagem, do inglês

Figura 2.1: Representação geométrica das normas. A definição é feita na equação 2.6. Tal definição estende-se para $p \rightarrow \infty$. Para $0 \leq p \leq 1$ tal definição para de ser identificada como norma. O uso de l_0 resume-se a identificar o número de elementos não nulos de \mathbf{x} , sendo a verdadeira medida do nível de esparsidade da solução. Apesar das regiões delimitadas pelas métricas apresentarem formas geométricas diversas, elas são comumente chamadas de bolas na literatura [1]



greedy algorithm, focam na regularização da norma l_0 de \mathbf{x} . As componentes do vetor de soluções \mathbf{x} são adicionadas sequencialmente até que a equação 2.8 seja totalmente satisfeita. Um bom exemplo dessa técnica é a perseguição de correspondência, ou do inglês *Matching Pursuit* (MP) [21]. Antes de adicionar uma nova componente a \mathbf{x} , a técnica verifica se a componente obtida anteriormente adequa-se ao sistema linear mostrado na equação 2.3. A abordagem MP possui a possibilidade de ficar presa em uma determinada iteração dependendo da atualização obtida. Para contornar essa situação o algoritmo de perseguição de adequação ortogonal, do inglês *Orthogonal Matching Pursuit* (OMP) foi desenvolvido, inserindo funções de custo [22]. Outras variações notáveis são o STOMP e o ROMP [3].

Já a classe de algoritmos de aplicação de limiar foca nas restrições presentes no sistema linear mostrado na equação 2.3, minimizando o erro residual de $\|\mathbf{Ax} - \mathbf{y}\|_2$ a cada interação. A esparsidade da solução é garantida pela limitação dos coeficientes da solução encontrada. A técnica de limitação rígida iterativa, *Interactive Hard Thresholding* é uma bom exemplo de aplicação dessa técnica [23]. Ela zera os coeficientes que apresentam valores desprezíveis entre as iterações, minimizando o erro residual. A solução obtida não é satisfatória enquanto as componentes não nulas não forem corretamente identificadas. Uma variação dessa técnica que vale a pena ser citada é o algoritmo de perseguição de limitação rígida, *Hard Thresholding Pursuit* (HTP) [20].

Outra maneira de obter soluções esparsas é utilizando o relaxamento convexo. O fato do problema de minimização de norma ser um problema NP-difícil é devido ao fato da norma l_0 ser um problema não convexo, ou seja, sem uma solução única. Entretanto, existe uma maneira de transformar o problema de minimização mostrado na equação 2.7 em outro mais fácil de se resolver, minimizando a norma l_1 em vez da l_0 , ou seja:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{x}\|_1, \\ \text{tal que} \quad & \mathbf{y} = \mathbf{Ax}. \end{aligned} \tag{2.8}$$

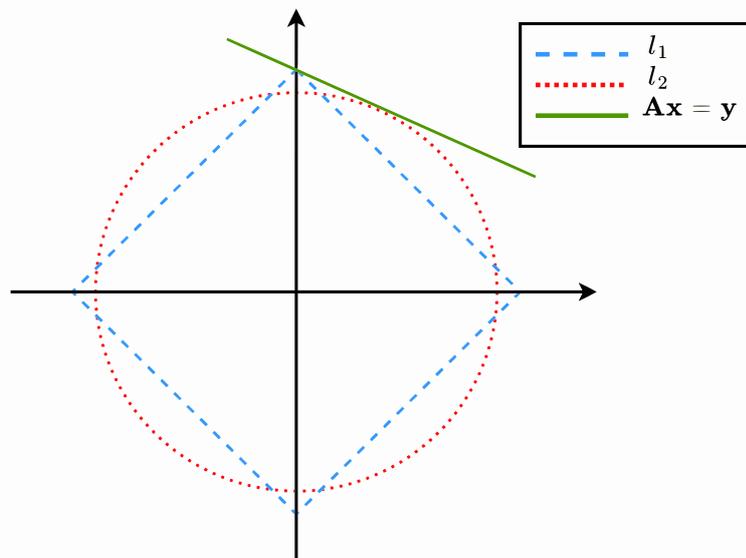
A solução do problema de minimização 2.8 é chamado de perseguição de base, ou *Basis Pursuit* (BP). Essa técnica seleciona colunas da matriz \mathbf{A} para formar uma sub-matriz que pode ser usada para resolver a equação 2.3. Caso o vetor \mathbf{x} seja esparsos o suficiente, a solução encontrada no relaxamento l_1 também será a solução para l_0 [18]. Essa condição pode ser atingida, já que a norma l_1 pode gerar soluções esparsas [20]. Isso pode ser visto na figura 2.2, a solução para as normas l_1 e l_2 são os locais de intersecção entre a linha e as bolas para cada norma. Nota-se que para bola l_1 o plano toca bem em um dos cantos, essa condição leva a uma das coordenadas ser 0 e a solução ser esparsa, por outro lado, para a norma l_2 isso não ocorre e por consequência leva a componentes não nulas.

Relaxamento Convexo para Normas l_1 e l_2 para Dados Ruidosos

A busca de soluções esparsas usando a normas l_1 e l_0 culminou nas técnicas discutidas no item anterior. Contudo, elas não consideram o fato do que medições experimentais estão sempre corrompidas por ruído, fazendo que o problema de minimização mostrado na equação 2.8 ser inadequado para recuperação de campo eletromagnético utilizando dados reais obtidos por meio de instrumentos de medição. Para conseguir executar a recuperação do campo nesse cenário utilizando a norma l_1 uma nova abordagem baseada em problema de minimização com regularização deve ser utilizada, sendo formulada como:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{x}\|_1, \\ \text{tal que} \quad & \mathbf{y} = \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \varepsilon. \end{aligned} \tag{2.9}$$

Figura 2.2: Visualização da esparsidade induzida pelo processo de minimização por norma l_0 de sistemas lineares.



Nessa ϵ denota a tolerância de erro. Essa técnica de minimização é chamada de perseguição de base com remoção de ruído, ou *Basis Pursuit Denoising* (BPDN). O nome de tal técnica vem de do fato ela ainda procura pelas colunas da matriz A para montar a melhor submatriz e obter a solução, contudo a restrição imposta pela iniquidade diminui o efeito do ruído experimental. A resolução do algoritmo BPDN é obtida de forma iterativa da seguinte maneira:

$$\begin{aligned} & \text{minimize} && \mathbf{x}, \\ & \text{tal que} && \mathbf{y} = \|\mathbf{Ax} - \mathbf{y}\|_2 + \lambda \|\mathbf{x}\|_1. \end{aligned} \quad (2.10)$$

Outra estratégia para diminuição da influência de ruído é a minimização na norma euclidiana ou regularização de Tikhonov-Phillips. Em termos matemáticos a técnica é formulada da seguinte forma:

$$\begin{aligned} & \text{minimize} && \mathbf{x}, \\ & \text{tal que} && \mathbf{y} = \|\mathbf{Ax} - \mathbf{y}\|_2 + \lambda \|\mathbf{x}\|_2. \end{aligned} \quad (2.11)$$

O papel de λ , ou fator de regularização, é idêntico em ambas as estratégias: controlar a quantidade de informação conhecida *a priori* sobre x para recuperar o mesmo. Tal papel é um problema bem conhecido e amplamente investigado. Nota-se que as equações 2.10 e 2.11 caso $\lambda = 0$ reduzem-se ao problema de mínimos quadrados.

Número de Amostras e Diagrama de Transição de Fase

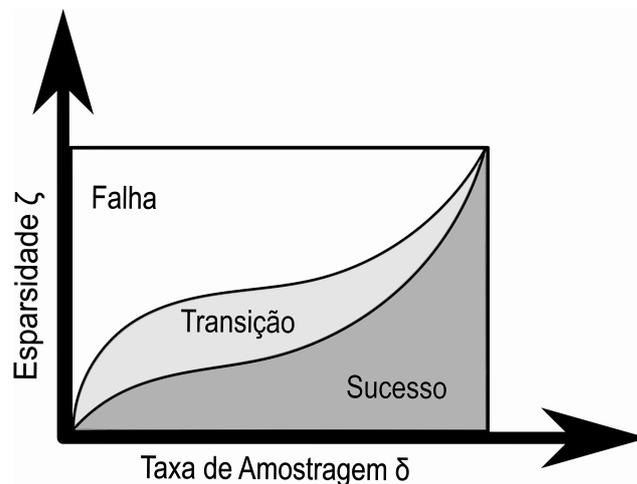
A recuperação esparsa dos coeficientes Q_{smn} permite a diminuição significativa da quantidade total de número de amostras quando comparada com a técnica analítica, além de garantir uma reconstrução de campo adequada e precisa do campo. Os diagramas de transição

de fase, *Phase Transition Diagrams* (PTD), são ferramentas gráficas que permitem visualizar em que condições um sistema subdeterminado será resolvido. Tais condições são [3]:

- A taxa de amostragem δ . O número de medidas para o número de coeficientes a serem determinados, compondo o eixo das abcissas;
- O nível de esparsidade ζ , ou seja, o número de coeficientes não nulos presentes na solução para um determinado número de medições, compondo o eixo das ordenadas.

O formato geral de tal diagrama é mostrado na figura 2.3. Como discutido em itens anteriores, quanto mais esparsa a solução, $\zeta \rightarrow 0$, menos pontos experimentais são necessários para recuperar a mesma. Contudo esse cenário raramente é replicado [3]. Originalmente, os diagramas PTD foram obtidos para matrizes de detecção A gaussianas [24]. Entretanto, eles podem ser usados para aplicações como a reconstrução do campo eletromagnético em que a matriz A foi obtida pela SWE.

Figura 2.3: Diagrama de transição para sistemas esparsos. Quanto mais esparsa for o vetor x menos pontos experimentais são necessários para recuperá-lo.



2.1.1.2 Reconstrução de Campo pelo Teorema de Equivalência

Como mostrado em 5.5 o campo elétrico pode ser descrito pelas fontes presentes dentro dos limites físicos da antena. Dessa maneira o princípio da Equivalência pode ser utilizado para reconstrução de campos eletromagnéticos. Esse conceito é extremamente implementado em análise de antenas como mostrado em [2, 25], em que a principal ideia é substituir a AUT por um conjunto de distribuições de corrente elétrica J_s e magnética M_s em uma superfície equivalente s [26]. Tal superfície pode ser modelada a partir do princípio da Equivalência de Schelkunoff [2, 25], muito utilizado para modelagem teórica de antenas de abertura, ou pelo princípio da Equivalência de Love. Dessa maneira, as distribuições de

corrente elétrica e magnética de superfície podem ser descritas como:

$$\begin{aligned} \mathbf{J}_s &= \hat{\mathbf{n}} \times (\mathbf{H}_1 - \mathbf{H}), \text{ e} \\ \mathbf{M}_s &= -\hat{\mathbf{n}} \times (\mathbf{E}_1 - \mathbf{E}). \end{aligned} \quad (2.12)$$

em que $\hat{\mathbf{n}}$ o vetor normal a s , enquanto \mathbf{E} e \mathbf{H} são os campos dentro de s e \mathbf{E}_1 e \mathbf{H}_1 são os campos externos à s . Por definição, os campos são expressos como a transformada de Fourier das correntes de superfície:

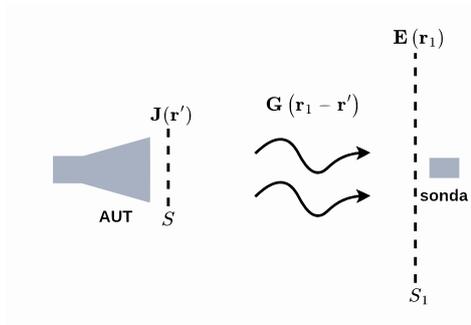
$$\mathbf{E}(\mathbf{r}) = \int_s \mathbf{J}_s(\mathbf{r}') \times \mathbf{G}(\mathbf{r} - \mathbf{r}') d s + \int_s \mathbf{M}_s(\mathbf{r}') \times \mathbf{G}(\mathbf{r} - \mathbf{r}') d s. \quad (2.13)$$

Nessa equação, $\mathbf{G}(\mathbf{r} - \mathbf{r}')$ é a já apresentada função de Green [2]. Utilizando-se do método dos momentos [27], *Method of Moments* (MoM), o método transforma a equação integral mostrado na equação 2.13 em uma combinação linear através da projeção de funções de base. Dessa maneira o campo eletromagnético pode ser escrito o em sistema linear do tipo 2.3 em que [28]:

$$\mathbf{y} = \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix}, \quad \mathbf{A} = \frac{k}{\sqrt{\eta}} \begin{bmatrix} Z_{J_u}^\theta & Z_{J_v}^\theta & Z_{M_u}^\theta & Z_{M_v}^\theta \\ Z_{J_u}^\phi & Z_{J_v}^\phi & Z_{M_u}^\phi & Z_{M_v}^\phi \end{bmatrix}, \text{ e } \mathbf{x} = \begin{bmatrix} J_u \\ J_v \\ M_u \\ M_v \end{bmatrix}, \quad (2.14)$$

em que u e v são as funções de base aplicada na integral mostrada na equação 2.13. Devido ao fato de ser implementada por uma abordagem matricial, todos os métodos discutidos de inversão podem ser utilizados para resolver tal problema, e alguns deles foram apresentados 2.1.1.1 para a reconstrução utilizando SWE. A aplicação de tal técnica é investigada em várias configurações de aquisição de dados experimentais [29] e superfícies equivalentes [28], e é exemplificada na figura 2.4.

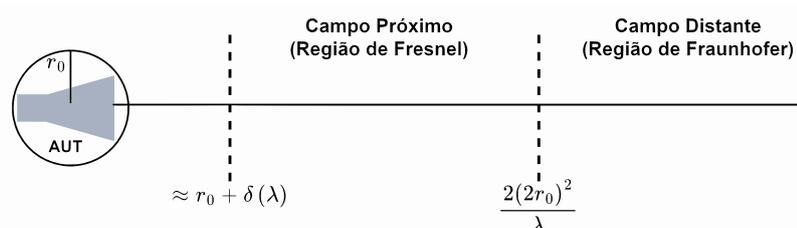
Figura 2.4: Exemplo de caracterização experimental de campo radiado baseado no princípio da Equivalência.



2.1.2 Técnicas de Medição do Campo Eletromagnético

Existem várias maneiras de executar as medições do campo eletromagnético. A escolha depende basicamente do tipo da antena e de sua aplicação [3]. Tais técnicas são separadas em duas vertentes: medições em campo distante e campo próximo. Essa separação advém da distância de separação entre a AUT e a sonda de aquisição. Essas duas regiões são mostradas na figura 2.5. Na região de campo distante, que ocorre para distâncias de observação acima da de Fraunhofer, o campo eletromagnético assume a forma de ondas planas em que o decréscimo da amplitude é inversamente proporcional à distância de propagação. Logo, as variações da magnitude do campo depende exclusivamente das coordenadas esféricas (θ, ϕ). Já na região de campo próximo, localizada na região intermediária entre a região reativa e a de campo distante a dependência do campo estende-se também à coordenada radial, o que leva ao uso de estratégias de reconstrução mais complexas e procedimentos de medição mais rigorosos.

Figura 2.5: Regiões de medições em torno da antena sob estudo de esfera mínima r_0 e λ o comprimento de onda de operação. $\delta(\lambda)$ denota a margem da região reativa [2], a sua determinação é extremamente difícil.



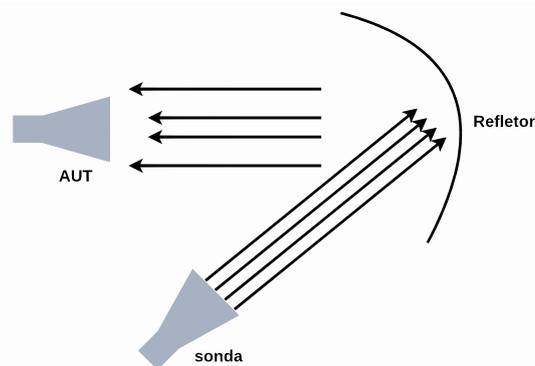
Medições em Campo Distante

Devido a aproximação de onda plana [4, 25, 2], o campo elétrico é mensurado de maneira direta. A reprodução da condição de campo distante geralmente é complicada e isso fica mais evidente para antenas de grandes dimensões, pois obter a distância r_e igual ou superior a distância de campo distante é impraticável em regiões fechadas, e a condução de tal procedimento em locais abertos é extremamente difícil devida à interferência do ambiente circundante. Dessa maneira, para poder conduzir essa classe de caracterização em ambientes controlados, é possível obter a condição de campo distante com auxílio elementos colimantes. Tais sistemas são chamados de caracterizadores de antenas em alcances compactos, ou *Compact Antenna Test Range (CATR)* [30] em que um exemplo é mostrado na figura 2.6. Esses equipamentos conseguem fornecer uma região de ondas planas para a AUT [31], utilizando um sistema de refletores ou lentes para colimar as ondas emitidas pela sonda de caracterização, sem a necessidade de grandes distâncias [32].

Apesar de serem equipamentos que permitem medições bastante precisas, eles são bastante caros, pois os sistemas de colimação (refletores e lentes) devem ser projetados e cali-

brados com alta precisão e devem estar dimensionados para o ambiente com os quais estão alocados. Além disso, o operador que conduz o roteiro de medições deve garantir a região de exposição entre o refletor e a AUT, ou região silenciosa [25, 4] para a condução correta do procedimento. Adicionalmente, a dimensão do elemento colimante é diretamente proporcional ao tamanho da região silenciosa, e sua forma define a frequência máxima de caracterização, além da fonte de campo deve garantir a simetria do diagrama de radiação e do centro de fase. Tais problemas limitam o uso da caracterizações em campo distante, principalmente para antenas que operam em frequências acima de 300 GHz.

Figura 2.6: Ilustração de um caracterizador de antena de alcances compactos CATR utilizando um sistema refletor. Tal instrumento permite fazer caracterizações de campo distante em espaços reduzidos.



Medições em Campo Próximo

Devido às desvantagens da realização de medições em campo distante, as medições em campo próximo são mais utilizadas, principalmente em técnicas de reconstrução baseadas em SWE, já que essa técnica identifica os coeficientes Q_{smn} utilizando transformações de campo. Tais transformações são chamadas assim pois transformam as medições do campo próximo para distante, *Near Field to Far Field Transformation* (NFFT), sendo também chamadas de técnicas indiretas [30]. Para as medições de campo próximo existem três tipos de estratégias de amostragem [25], em que para todas elas, correções devem ser executadas em algum grau em relação a sonda, como explicado em [4].

2.1.2.1 Estratégias de Amostragem de Dados em Campo Próximo

A amostragem de dados em medições em campo próximo pode ser feita em diferentes geometrias. Em termos gerais, as geometrias utilizadas são a planar, a cilíndrica e a esférica. A simetria planar é utilizada geralmente na reconstrução de campo baseada no teorema da Equivalência e no método analítico usando a SWE. Tal estratégia possui uma taxa de amostragem espacial constante, ou seja, $\Delta x = \Delta y$ em uma malha de amostragem bidimensional

planar, contudo simetrias polares também são utilizadas [25]. A taxa máxima de amostragem na configuração planar é $\Delta x_{max} = \Delta y_{max} = \lambda/2$. Nesse procedimento de varredura geralmente a AUT fica estática enquanto a sonda de aquisição descreve a trajetória na superfície de medição. Esse processo é ilustrado na figura 2.7. A estratégia de varredura planar possui a vantagem de ser a mais matematicamente simples quando comparada com as suas contrapartes cilíndrica e esférica [2], contudo a sua principal desvantagem é o fato de que o campo ser reconstruído em um setor angular limitado [25]. Já a varredura cilíndrica é ilustrada na

Figura 2.7: Ilustração da aquisição de dados com simetria planar. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.

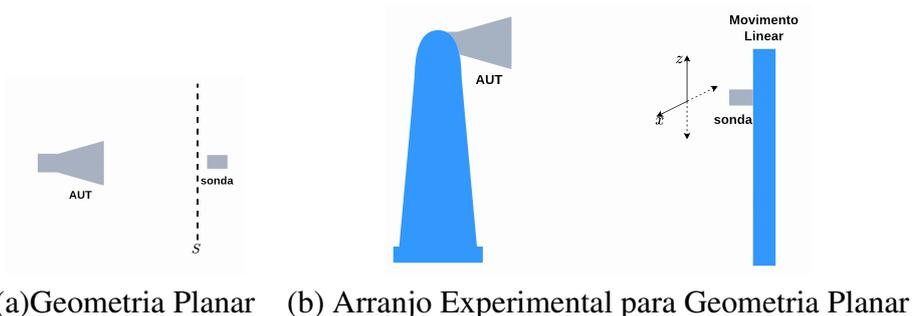
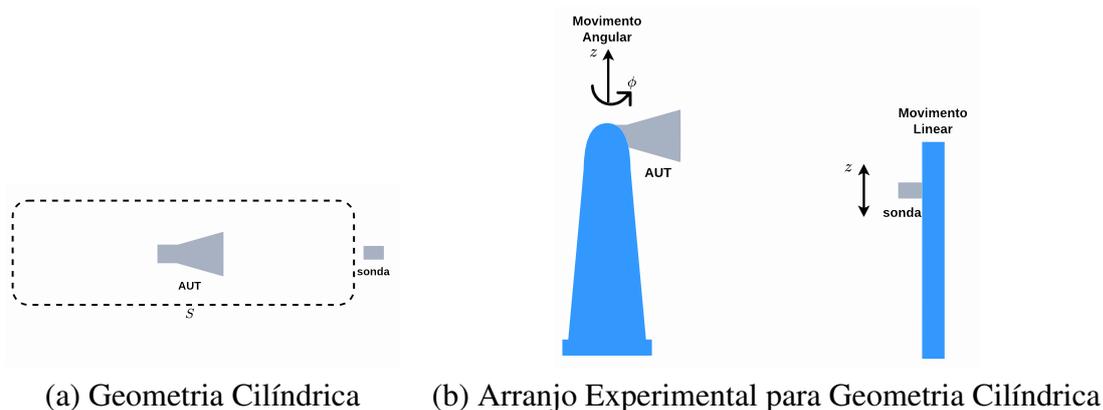


figura 2.8. Ela é muito utilizada para antenas de feixe estreito e possui um comportamento dual em relação às vantagens e desvantagens, ou seja, apesar de eliminar o uso de setores angulares limitados, a sua complexidade aumenta de forma significativa devido as funções matemáticas usada na reconstrução das ondas, incluem funções cilíndricas de Hankel [25]. As taxas de amostragem em tal geometria de superfície são [25]:

$$\Delta\phi = \frac{\lambda}{2r_0 + \lambda}, \text{ e} \quad (2.15)$$

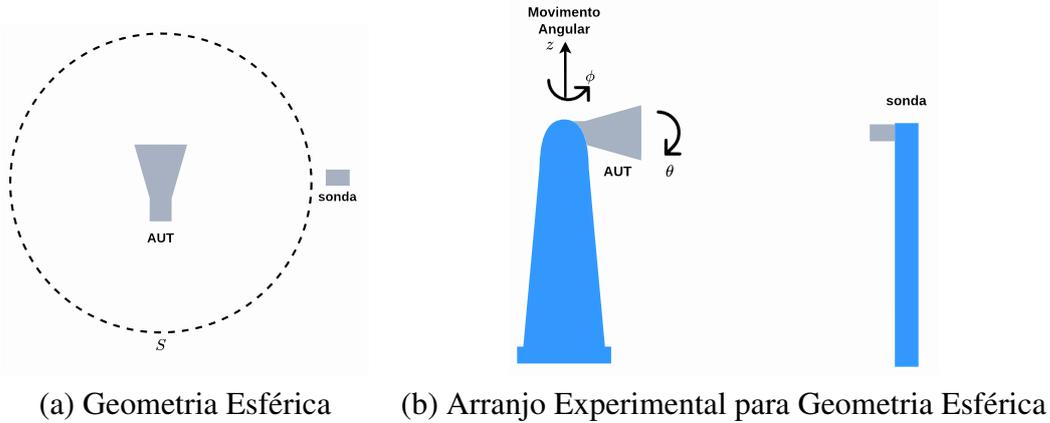
$$\Delta z = \frac{\lambda}{2},$$

Figura 2.8: Ilustração da aquisição de dados com simetria cilíndrica. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.



nessa equação r_0 é o raio da menor cilindro que engloba a antena. No processo de caracterização geralmente a sonda realiza movimentos lineares no eixo z enquanto a antena descreve os movimentos rotacionais em ϕ . A amostragem esférica é a abordagem mais eficiente quando comparada com as demais, pois a sua malha de amostragem engloba toda antena, podendo recuperar com alto grau de fidelidade o campo eletromagnético, sendo bastante utilizada em técnicas baseadas em SWE [16]. Além disso, ela apresenta diversas variações no que diz respeito a taxa de amostragem, além de ser extremamente precisa e bem documentada [33]. A figura 2.9 ilustra a amostragem esférica. A mais clássica técnica de

Figura 2.9: Ilustração da aquisição de dados com simetria esférica. A figura (a) ilustra a geometria de tal estratégia de varredura enquanto a figura (b) ilustra um arranjo experimental típico.



amostragem esférica é a equiangular. Essa estratégia, assim como a planar e a cilíndrica, possui uma taxa constante em θ e em ϕ . Segundo [25] elas podem ser obtidas como:

$$\Delta\phi = \frac{\lambda}{2r_0 + \lambda}, \text{ e} \quad (2.16)$$

$$\Delta\theta = \frac{\lambda}{2r_0 + \lambda}.$$

Essa técnica é geralmente utilizada com o método analítico de reconstrução utilizando SWE, e usa uma grande quantidade de pontos:

$$P = 2(2N + 1)(N + 1), \quad (2.17)$$

em que N é a já citada ordem de truncamento da SWE. Essa técnica de amostragem possui a desvantagem de ter uma sobre amostragem dos polos da esfera. Contudo ela permite uma recuperação bastante eficiente do campo eletromagnético e escala de maneira eficiente para altas ordem de truncamento [33, 3].

Com o objetivo de evitar sobre amostragem nos polos, a estratégia de amostragem de redundância mínima, *Minimal Redundant Samplig* (MRS), foi desenvolvida. Ela propõe a aquisição de dados mantendo um ângulo sólido de amostragem constante em θ e ϕ . Essa abordagem foi validada por [34] em aplicações com campos de banda limitada. As taxas de

amostragem em θ e ϕ seguem a seguinte definição matemática:

$$\begin{aligned}\Delta\theta &= \frac{2\pi}{2(\chi\beta r_0) + 1}, \text{ e} \\ \Delta\phi &= \frac{2\pi}{2(\chi_\theta\beta \sin\theta) + 1},\end{aligned}\tag{2.18}$$

em que $\beta = 2\pi/\lambda$, χ é o fator de sobre amostragem geralmente maior que 1, e r_0 o raio da esfera mínima. Já χ_θ é definida como:

$$\chi_\theta = 1 + (\chi - 1) \sin \theta^{-2/3}.\tag{2.19}$$

As relações mostradas na equação 2.18 estão relacionadas com a largura de banda e permitem a amostragem ser feita respeitando o teorema de amostragem de Nyquist [16].

Outra abordagem que possui o objetivo de não obtenção de dados redundantes é a estratégia iglu conforme descrito em [16]. Essa abordagem também garante um ângulo sólido constante pois todos os ângulos ϕ de amostragem seguem a seguinte formulação matemática:

$$\Delta\phi = \frac{\Delta\theta}{\sin\theta}.\tag{2.20}$$

Essa estratégia possui uma característica interessante quando se aplica a mesma regra de amostragem para θ mostrada para técnica MRS, ela produz posições amostragem bem semelhantes em ϕ [16], a diferença reside que a técnica MRS possui uma leve sobre amostragem nos polos. Já a estratégia de Fibonacci, ou técnica de amostragem espiral [35], as posições da superfície de medição seguem um padrão posicional baseado na sequência de Fibonacci, definindo um parâmetro auxiliar:

$$h_p = -1 + \frac{2(p-1)}{P-1},\tag{2.21}$$

em que P é o número total de pontos para $1 \leq p \leq P$, as posições de amostragem são definidas como:

$$\begin{aligned}\theta_p &= \arccos h_p, \text{ e} \\ \phi_p &= \left(\phi_{p-1} \frac{36}{\sqrt{P}} \frac{1}{\sqrt{1-h_p^2}} \right).\end{aligned}\tag{2.22}$$

Para $2 \leq p \leq P-1$ e $\phi_P = 0$.

2.1.3 Métricas de Avaliação

Para validar o campo eletromagnético reconstruído seja para avaliar novas técnicas, configurações de reconstrução ou para escolha do modo de operação da AUT é necessário utilizar métricas de avaliação. Para a reconstrução do campo eletromagnético, a comparação entre

dois padrões, o recuperado e o de referência, pode ser feito de duas maneiras: ponto a ponto ou global. A abordagem ponto a ponto fornece um novo padrão de radiação dependendo da ferramenta usada, já as técnicas globais retornam um valor único para mensurar a similaridade entre os dois padrões, tanto para medir a precisão do processo ou para otimização [3]. Por questões didáticas $E_1(\theta, \phi)$ e $E_2(\theta, \phi)$, ou simplesmente E_1 e E_2 , irão ilustrar os campos de referência e reconstruído respectivamente, enquanto \mathbf{y}_1 e \mathbf{y}_2 os conjuntos de discretizados em torno da mesma superfície de amostragem.

2.1.3.1 Técnicas Ponto a Ponto

Existem diferentes técnicas de comparação ponto a ponto, e cada uma delas irá evidenciar diferentes aspectos entre os campos comparados. As métricas mais comuns são [3]:

$$\begin{aligned} \Delta_{lin}(\theta, \phi) &= |E_1(\theta, \phi) - E_2(\theta, \phi)| && \text{Diferença Linear,} \\ \Delta_{W_{lin}}(\theta, \phi) &= W_{lin} \Delta_{lin} && \text{Diferença Linear Ponderada} \\ \\ \Delta_{log}(\theta, \phi) &= 20 \log_{10} |E_1(\theta, \phi)| - 20 \log_{10} |E_2(\theta, \phi)| && \text{Diferença Logarítmica} \\ \Delta_{W_{log}}(\theta, \phi) &= W_{log} \Delta_{log} && \text{Diferença Logarítmica Ponderada} \end{aligned} \quad (2.23)$$

As funções W_{lin} e W_{log} são usadas para diminuir a influência de valores entre as diferenças entre os campos comparados não relevantes.

2.1.3.2 Métricas Globais

De maneira geral, qualquer norma matemática pode ser utilizada como uma métrica, já que o campo está discretizado [3]. Contudo para reconstrução de campo algumas normas mais adequadas foram desenvolvidas para evidenciar melhor o contraste entre o campo recuperado e o campo de referência. A primeira métrica global é o campo residual, que utiliza da norma l_2 da diferença entre \mathbf{y}_1 e \mathbf{y}_2 para mensurar o erro entre os campos, enfatizando regiões em que a diferença entre eles é maior [7]. Matematicamente é definido como:

$$\epsilon(\mathbf{y}_1, \mathbf{y}_2) = \frac{\|\mathbf{y}_1 - \mathbf{y}_2\|_2}{\|\mathbf{y}_1\|_2}. \quad (2.24)$$

Para estimação de \mathbf{y}_1 ser considerada boa o valor $\epsilon(\theta, \phi)$ tem que ser próximo de zero. Já a métrica grau de correlação, em vez de mensurar a diferença entre os campos, determinar o grau de semelhança entre eles, ou seja quanto essa métrica é igual a unidade, isso indica que \mathbf{y}_1 e \mathbf{y}_2 são iguais [7, 36]. Dessa maneira, essa métrica indica se ambos os campos seguem as mesmas variações e é definida como:

$$\delta(\mathbf{y}_1, \mathbf{y}_2) = \frac{|\langle \mathbf{y}_1, \mathbf{y}_2 \rangle|}{\|\mathbf{y}_1\|_2^2 \|\mathbf{y}_2\|_2^2}. \quad (2.25)$$

Nessa equação, $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle$ é o produto interno entre \mathbf{y}_1 e \mathbf{y}_2 . Já o nível equivalente de ruído, *Equivalent Noise Level* (ENL), que pode ser definido como o erro normalizado médio entre \mathbf{y}_1 e \mathbf{y}_2 . O uso dessa métrica é indicada quando se espera um grau de aproximação entre os dois vetores \mathbf{y}_1 e \mathbf{y}_2 de maneira uniforme em uma determinada região [37]. A operação de média em tal técnica possui a desvantagem de não evidenciar diferenças muito pequenas, mesmo que ela esteja no lóbulo principal do campo em situações exista diferenças seja pequena em outras regiões. Contudo, tal métrica possui a vantagem de ser de fácil interpretação, podendo facilitar comparações quando vários cenários são analisados. Matematicamente ela pode ser definida como:

$$\text{ENL}(\mathbf{y}_1, \mathbf{y}_2) = \frac{\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2}{\text{média}(\|\mathbf{y}_1\|_2)}. \quad (2.26)$$

2.2 Recuperação de Fase

Esta seção aborda a classe de técnicas de reconstrução de campos eletromagnéticos utilizando somente medições em amplitude chamada recuperação de fase, do inglês *Phase Retrieval* (PR). Matematicamente, tal técnica pode ser representada como:

$$\text{encontre } \mathbf{x} \text{ para } \mathbf{y} = |\mathbf{A}\mathbf{x}|, \quad (2.27)$$

em que $\mathbf{x} \in \mathbb{C}^n$ é o vetor com contém o sinal de interesse, $\mathbf{A} \in \mathbb{C}^{m \times n}$ é matriz de detecção e $\mathbf{y} \in \mathbb{R}^m$ é o vetor de dados experimentais [28]. O problema de recuperação de fase está presente nos mais diversos campos de pesquisa como a cristalografia com imagens de raios-x [38], processamento de sinais [39], caracterização experimental de antenas [40] e ótica [41].

A recuperação da fase é por definição classificado como um problema não convexo [28] e NP-difícil (*NP-hard* em inglês). Por este motivo, diversas abordagens para solucionar a equação 2.27 foram desenvolvidas nas últimas décadas. Esta seção tem como objetivo apresentar técnicas de recuperação de fase, desde o clássico algoritmo de Gerchberg-Saxton até as mais recentes técnicas baseadas em programação semi-definida.

2.2.1 Estado da Arte

2.2.1.1 Algoritmo de Gerchberg-Saxton

A primeira e notória técnica desenvolvida para solucionar a recuperação de fase é o algoritmo de Gerchberg-Saxton (GB) ou método de dupla varredura. Essa abordagem iterativa utiliza dois planos de medição para encontrar a distribuição de fase em ambos os planos. Em outras palavras o algoritmo GB é um método de retropropagação, com objetivo de solucionar

um problema do tipo:

$$\mathbf{y} e^{j\tilde{\phi}} = \mathbf{A}\mathbf{x}, \quad (2.28)$$

em que $\tilde{\phi}$ são as fases a serem determinadas [28] e o vetor \mathbf{y} são as amplitudes nos planos de medição. No âmbito de caracterização de antenas as superfícies de varredura podem apresentar geometrias retangulares, cilíndricas ou esféricas. Combinações entre tais geometrias podem ser utilizadas para a resolução do problema [42]. Em termos gerais, o algoritmo de Gerchberg-Saxton pode ser resumido da seguinte maneira: Nesse algoritmo, $\tilde{\mathbf{y}}_1$ e $\tilde{\mathbf{y}}_2$ são a

Algoritmo 1: Algoritmo de Gerchberg-Saxton para Recuperação de Fase

Dados: $|\mathbf{y}_1|$, $|\mathbf{y}_2|$ as amplitudes nas superfícies S_1 e S_2 , as matrizes de detecção \mathbf{A}_1 e \mathbf{A}_2 e suas respectivas pseudo-inversas $\tilde{\mathbf{A}}_1^\dagger$ e $\tilde{\mathbf{A}}_2^\dagger$ para as superfícies S_1 e S_2 , palpite da distribuição das fases $\tilde{\phi}^0$, a tolerância de erro tol e a quantidade máxima de iterações max_{iter} .

início

```

# Inserindo palpite inicial;
 $\tilde{\mathbf{y}}_1 \leftarrow |\mathbf{y}_1| e^{j\tilde{\phi}^0}$ ;
 $iter \leftarrow 0$ ;
repita
    # Obtendo vetor  $\mathbf{x}$  em  $S_1$ ;
     $\mathbf{x} \leftarrow \tilde{\mathbf{A}}_1^\dagger \tilde{\mathbf{y}}_1$ ;
    # Propagando para  $S_2$ ;
     $\tilde{\mathbf{y}}_2 \leftarrow \mathbf{A}_2 \mathbf{x}$ ;
    Obtendo a fase de  $\tilde{\mathbf{y}}_2$  ;
     $\tilde{\phi}_2 \leftarrow \text{fase}(\tilde{\mathbf{y}}_2)$ ;
    # Fazendo correção de fase em  $\tilde{\mathbf{y}}_2$ ;
     $\tilde{\mathbf{y}}_2 \leftarrow |\mathbf{y}_2| e^{j\tilde{\phi}_2}$ ;
    # Obtendo vetor  $\mathbf{x}$  em  $S_2$ ;
     $\mathbf{x} \leftarrow \tilde{\mathbf{A}}_2^\dagger \tilde{\mathbf{y}}_2$ ;
    # Propagando para  $S_1$ ;
     $\tilde{\mathbf{y}}_1 \leftarrow \mathbf{A}_1 \mathbf{x}$ ;
    # Obtendo a fase de  $\tilde{\mathbf{y}}_1$  ;
     $\tilde{\phi}_1 \leftarrow \text{fase}(\tilde{\mathbf{y}}_1)$ ;
    # Fazendo correção de fase em  $\tilde{\mathbf{y}}_1$ ;
     $\tilde{\mathbf{y}}_1 \leftarrow |\mathbf{y}_1| e^{j\tilde{\phi}_1}$ ;
    # Cálculo de erro, mudança da norma;
     $\epsilon \leftarrow \left\| \mathbf{y}_1^{(-)} - \tilde{\mathbf{y}}_1 \right\|$ ;
    # Armazenando o novo valor de  $\mathbf{y}_1^{(-)}$ ;
     $\mathbf{y}_1^{(-)} \leftarrow \tilde{\mathbf{y}}_1$ ;
até  $\epsilon \leq tol$  ou  $iter = max_{iter}$ ;
Saída: Vetor do sinal de interesse  $\mathbf{x}$ 

```

fim

estimação do campo nas superfícies S_1 e S_2 e $\tilde{\phi}_1$ e $\tilde{\phi}_2$ são as fases recuperadas para ambas as superfícies. Já o erro associado ϵ é o que determina a convergência do algoritmo. Ide-

almente, as fases em S_1 e S_2 estarão totalmente recuperadas quando ϵ é nulo. Contudo, tal cenário não acontece na prática. De acordo com o teorema de Parseval, a convergência do algoritmo GB será garantida se soma das amplitudes da superfície S_1 seja ao menos igual à soma das amplitudes da superfície S_2 , o que garante que o erro decaia a cada iteração ou, na pior das hipóteses, mantenha-se constante [10].

Tal condição é necessária para que o método consiga recuperar a grandeza de interesse \mathbf{x} , contudo não é suficiente. Isso é explicado pelo problema de recuperação de fase ser um problema não linear e não convexo [15, 43], ou seja, com infinitas soluções [10]. Esse comportamento é explicado pelo número limitado de amostras, e por consequência encontrar o mínimo global da equação 2.27 não é um processo fácil, o que leva o algoritmo GB a ficar preso em algum mínimo local. A maneira mais direta para contornar tal situação é aumentar o número de amostras, o que aumenta a chance de alcançar uma solução mais próxima da exata, contudo, tal manobra exige grandes recursos computacionais aumentando o número de iterações [10].

O algoritmo GB é bastante sensível ao vetor de palpite inicial da distribuição de fase ϕ^0 . Nas primeiras implementações mostradas em [10], o vetor ϕ^0 era iniciado com uma distribuição randômica com valores entre -180° e 180° , o que poderia levar o algoritmo à falha. Ao longo do tempo novas abordagens para obter distribuições de fase iniciais adequadas foram desenvolvidas, sendo a mais simples delas é iniciar a distribuição ϕ^0 com valores nulos e não viesados. No caso da aplicação do algoritmo GB na caracterização de antenas, uma boa técnica é iniciar tal distribuição com a distribuição de fase de uma antena dipolo [44].

2.2.1.2 Algoritmos de Redução de Erro

O algoritmo de redução de erro, do inglês *Error-Reduction Algorithm* (ER), é umas das primeiras implementações desenvolvidas com o intuito de melhorar o algoritmo de Gerchberg-Saxton. Primeiramente implementada para aplicações do método de varredura à astronomia [11], o método ER é aplica restrições em ambos os planos de medição para alcançar resultados mais precisos, funcionando da seguinte maneira: considerando que $\tilde{\mathbf{y}}_n^k$ o sinal propagado ou em S_1 ou em S_2 , para $n = 1, 2$ respectivamente, e que k denota a k -ésima iteração, a propagação seguirá a seguinte regra [11]:

$$\tilde{\mathbf{y}}_n^{k+1} = \begin{cases} \mathbf{A}_n \mathbf{x} & \mathbf{x} \notin \gamma \\ 0 & \mathbf{x} \in \gamma \end{cases} . \quad (2.29)$$

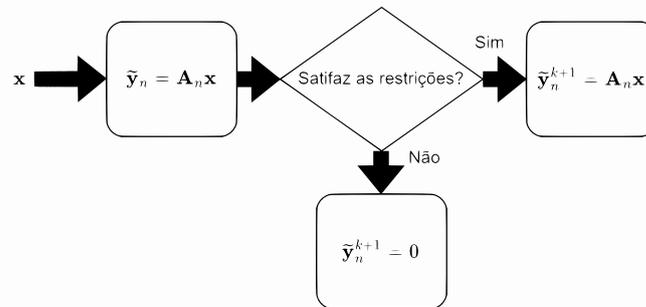
Aqui γ é a região que inclui todos os pontos de $\tilde{\mathbf{y}}_n^k$ que não satisfazem as restrições impostas. Em termos gerais, a principal restrição é que $\tilde{\mathbf{y}}_n$ não deve ser negativo. O uso de restrições adicionais pode ser utilizado para melhorar o rendimento do algoritmo, como exemplificado por [11] que impõe que o comprimento da estimacão não ultrapasse o tamanho do objeto conhecido. Já em [45] que utiliza o algoritmo ER em processamento de imagens, impõe a restrição em que intensidades das imagens em ambos os planos de medição mantenham-se

constantes durante todo o processo, visto que as mesmas são conhecidas.

No que compete ao erro ϵ , o mesmo deve cair ou manter-se constante durante a recuperação da fase, isso é esperado já que o algoritmo ER é derivado do antecessor Gerchberg-Saxton. Contudo, devido a imposição das restrições, a recuperação de fase que utiliza a técnica de redução de erro possui uma convergência mais rápida do que o algoritmo GB [11, 12], já que o erro decai rapidamente durante as primeiras iterações.

Já no que diz respeito ao vetor de palpite inicial de distribuição de fase, o mesmo também pode ser inicializado com uma distribuição aleatória, o que não é muito recomendável devido à natureza não linear e não convexa do problema de recuperação de fase, o que pode levar a soluções inadequadas.

Figura 2.10: Fluxo de atualização do algoritmo iterativo de entrada e saída. Caso na propagação o vetor \mathbf{x} satisfaça as restrições o campo receberá o resultado de tal operação, caso contrário ele será atualizado com o resultado da propagação adicionado de um fator de correção.



2.2.1.3 Algoritmo de Entrada e Saída

Os algoritmos de Gerchberg-Saxton e de redução de erro apresentam uma convergência lenta a medida que o número de interações aumenta. O algoritmo de entrada e saída, do inglês *Basic Input-Output Algorithm* (BIO), possui como objetivo superar tal problema. Ele é muito similar ao algoritmo ER, pois ambos selecionam a estimacão $\tilde{\mathbf{y}}_n^{k+1}$ de acordo com as restrições impostas. Contudo eles diferem em um aspecto: enquanto no algoritmo ER $\tilde{\mathbf{y}}_n^k$ é considerada como a melhor soluçao após a propagação, no algoritmo BIO a estimacão da onda é tratada com uma possibilidade que pode ser usada na próxima iteraçao.

Essa flexibilidade de escolha da melhor estimacão para a próxima iteraçao permite que a estimacão de fase por métodos de varredura melhore a sua convergência, principalmente nos estágios finais do processo. De fato, as abordagens baseadas nas técnicas GB e ER possuem como característica que a mudança aplicada na estimacão $\tilde{\mathbf{y}}_n^k$ terá um impacto linear no processo de propagação. A não linearidade inserida pelo método BIO irá adicionar termos ao vetor \mathbf{x} em que caso for executada de maneira adequada, levará a um processo de recuperação de fase a um melhor rendimento.

De acordo com [12], essa operação não linear pode ser pensada com uma entrada $\tilde{\mathbf{y}}_n^k$ e

uma saída $\tilde{\mathbf{y}}_n^{k'}$ que sempre irá satisfazer as restrições impostas no processo de propagação, e consequentemente na estimação da distribuição de fase. A mudança desejada para saída é:

$$\Delta\tilde{\mathbf{y}}_n^k = \begin{cases} 0, & \mathbf{x} \notin \gamma \\ -\tilde{\mathbf{y}}_n^{k'} & \mathbf{x} \in \gamma. \end{cases} \quad (2.30)$$

Nessa equação γ é um conjunto de pontos de $\tilde{\mathbf{y}}_n^{k'}$ que violam as restrições impostas na propagação. A equação 2.30 mostra que caso as restrições sejam satisfeitas, a saída do sistema não será alterada, caso contrário a saída será zero. Dessa maneira, a próxima saída para estimação será:

$$\tilde{\mathbf{y}}_n^{k+1} = \begin{cases} \tilde{\mathbf{y}}_n^k, & \mathbf{x} \notin \gamma \\ \tilde{\mathbf{y}}_n^k - \beta\Delta\tilde{\mathbf{y}}_n^k, & \mathbf{x} \in \gamma, \end{cases} \quad (2.31)$$

em que β é uma constante de proporcionalidade que determina o impacto da mudança no processo de estimação da fase. A equação 2.31 descreve o algoritmo BIO matematicamente. Como $\tilde{\mathbf{y}}_n^{k'}$ na propagação já satisfaz as restrições, a saída pode ser considerada que foi gerada a partir de si mesma considerada como entrada. Uma outra maneira de obter a estimação é:

$$\tilde{\mathbf{y}}_n^{k+1} = \begin{cases} \tilde{\mathbf{y}}_n^{k'}, & \mathbf{x} \notin \gamma \\ \tilde{\mathbf{y}}_n^{k'} - \beta\tilde{\mathbf{y}}_n^{k'}, & \mathbf{x} \in \gamma, \end{cases} \quad (2.32)$$

em que caso $\beta = 1$ a equação 2.32 reduz-se a equação 2.29 que caracteriza o algoritmo ER. Uma outra maneira de obter a estimação usada para próxima iteração é usando uma combinação da linha superior da equação 2.32 e a linha inferior da equação 2.31:

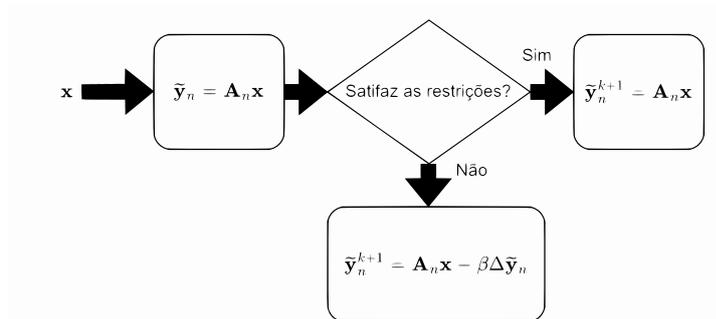
$$\tilde{\mathbf{y}}_n^{k+1} = \begin{cases} \tilde{\mathbf{y}}_n^{k'}, & \mathbf{x} \notin \gamma \\ \tilde{\mathbf{y}}_n^k - \beta\Delta\tilde{\mathbf{y}}_n^k, & \mathbf{x} \in \gamma. \end{cases} \quad (2.33)$$

Essa formulação resulta no algoritmo de entrada e saída híbrido, ou do inglês *Hybrid Input-Output Algorithm* (HIO), que é uma tentativa de evitar uma estagnação que ocorre especificamente no algoritmo ER, já que este trabalha com a condição de não alterar a sua velocidade de convergência mesmo nque esteja longe da solução. Nesse aspecto o algoritmo HIO mesmo na situação em que \mathbf{x} esteja na direção oposta da solução por mais de uma iteração, a taxa de convergência continua sendo alterada até a distribuição de fase convergir para uma estimação $\tilde{\mathbf{y}}_n$ adequada.

2.2.1.4 Algoritmo de Elevação de Fase

A abordagem clássica de dupla varredura de Gerchberg-Saxton possui uma implementação direta e pode ser alterada para satisfazer determinadas restrições dando origem às técnicas ER, BIO e HIO, que possuem como objetivo melhorar a convergência do algoritmo GB. Contudo tais técnicas necessitam de restrições bem estabelecidas, um bom palpite inicial ou um palpite não enviesado da distribuição de fase. Além disso, o fato do problema de PR

Figura 2.11: Fluxo de do algoritmo de entrada e saída. Caso na propagação o vetor \mathbf{x} satisfaça as restrições o campo receberá o resultado de tal operação, caso contrário ele será atualizado com valor nulo.



possuir infinitas soluções impacta bastante o uso de tais técnicas iterativas.

Tal problema de não unicidade da solução da distribuição de fase, faz necessária a utilização de uma grande quantidade de amostras, que apesar de mitigar os problemas citados, aumenta o tempo do processo e por consequência o uso de recursos computacionais. O algoritmo de elevação de fase, ou do inglês *Phaselift Algorithm* (PL) em contraste as técnicas de dupla varredura, tende a convergir à uma solução única. De modo geral, o método de recuperação de fase PL é baseado nas etapas [13]:

- Utilização de múltiplas superfícies, que proveem diferentes perspectivas da distribuição de fase da onda a ser caracterizada. Essas diferentes perspectivas podem fornecer uma solução única;
- Formulação do problema de recuperação de fase como um problema de relaxamento, transformando o problema de obter um vetor a partir de restrições quadráticas em um problema matricial de posto unitário com restrições afins, afrouxando um problema combinatório em um problema convexo.

Essa transformação da formulação matricial do problema de recuperação de fase é o que nomeia o algoritmo PL. Nos itens a seguir serão discutidos com mais detalhes as etapas envolvidas para recuperação da distribuição de fase utilizando a técnica PL.

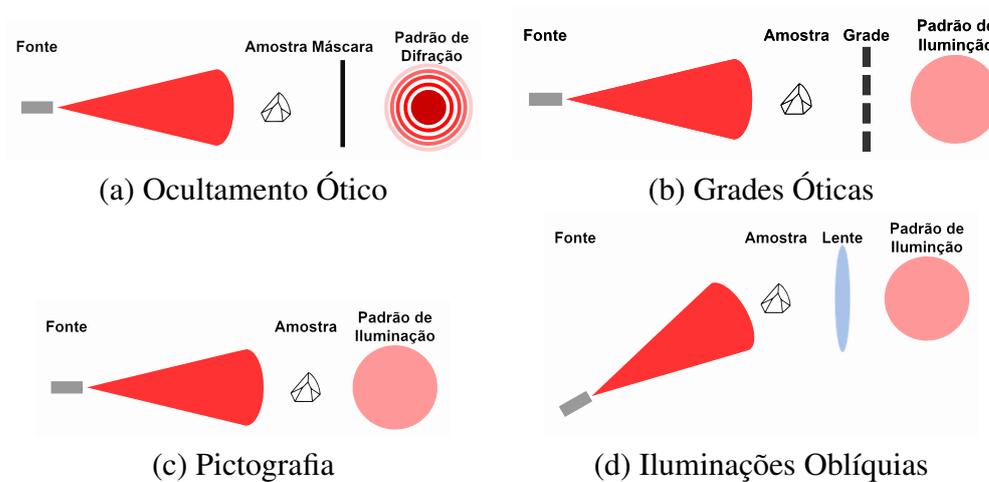
Técnicas de Aquisição

O algoritmo PL utiliza de várias perspectivas da onda eletromagnética:

$$\mathbf{y}(\phi) = \mathbf{w}(\Phi)e^{j\phi}, \quad (2.34)$$

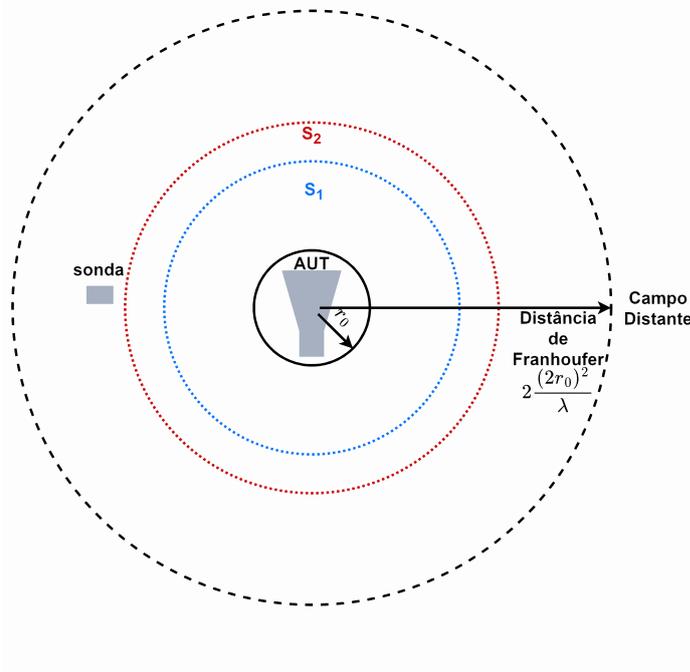
em que \mathbf{w} simboliza a variação inserida na distribuição com fase Φ , que deve ser selecionada dependendo da aplicação. A equação 2.34 mostra o processo como uma modulação, que por definição altera a frente de fase da onda eletromagnética. Para obter as diferentes visões da distribuição de fase ϕ , as principais técnicas são [13]:

Figura 2.12: Técnicas de aquisição de padrões de iluminação para resolução do problema de recuperação de fase via relaxação convexa. A figura (a) refere-se à técnica de ocultamento óptico, a figura (b) à técnica de grades óticas, figura (c) ao uso de pictografia e a figura (d) ilustra o uso de iluminações oblíquas.



- **Ocultamento Óptico:** Este método modifica a frente de fase da onda eletromagnética inserindo uma máscara ótica na frente da amostra a ser escaneada, como ilustrado na figura 2.12-(a). Muito utilizado em cristalografia e processamento de imagens, nela a amostra é escaneada a partir de deslocamentos da máscara, gerando assim variações na distribuição de fase. Esse método é muito similar a pictografia, a diferença reside que enquanto a pictografia escaneia o padrão gerado pela máscara, o ocultamento escaneia a amostra em si;
- **Grades Óticas:** O método é muito similar ao ocultamento óptico. A diferença reside que em vez de usar uma máscara, é utilizada uma estrutura em grade para gerar a variação da distribuição de fase, como evidenciado na figura 2.12-(b). Outro detalhe é que amostra pode ser colocada ou na frente ou atrás dessa estrutura;
- **Pictografia:** Aqui múltiplos padrões de iluminação são escaneadas como mostrado na figura 2.12-c, sendo comum uma sobreposição entre padrões adjacentes;
- **Iluminações Oblíquas:** Esta técnica utiliza raios que iluminam a amostra em diferentes ângulos como esquematizado na figura 2.12-d, com a possibilidade de uso de múltiplas iluminações simultâneas.
- **Superfícies de Medição:** Assim como nas técnicas de dupla varredura, múltiplas superfícies de escaneamento podem ser utilizadas, visto que para ondas eletromagnéticas de elementos radiantes a distribuição de fase está diretamente ligada com a distância de propagação [2]. A figura 2.13 exemplifica o processo de aquisição de dados em amplitude para este caso.

Figura 2.13: Técnicas de aquisição da amplitude de ondas eletromagnéticas para resolução do problema de recuperação de fase via relaxação convexa. Tal técnica também é utilizada nos algoritmos iterativos do tipo Gerchberg-Saxton. A diferença reside na quantidade, enquanto na técnica iterativas são utilizadas duas esferas, na relaxação convexa utiliza-se múltiplas medições.



Processo de Elevação

Para \mathbf{x} pertencente a \mathbb{C}^n , e considerando que o módulo das medições assume a forma:

$$y_k = |\langle \mathbf{a}_k, \mathbf{x} \rangle|^2, \quad k = 1, 2, \dots, m. \quad (2.35)$$

Nessa equação, \mathbf{a}_k denota o k -ésimo vetor da matriz de detecção \mathbf{A} . Dessa maneira as medições quadráticas podem ser elevadas e interpretadas como medições lineares em torno de uma matriz de posto unitário do tipo $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$, em que \mathbf{x}^\dagger é a hermitiana de \mathbf{x} da seguinte maneira:

$$\begin{aligned} |\langle \mathbf{a}_k, \mathbf{x} \rangle|^2 &= \text{Tr}(\mathbf{x}^\dagger \mathbf{a}_k^\dagger \mathbf{a}_k \mathbf{x}) \\ \text{Tr}(\mathbf{x}^\dagger \mathbf{a}_k^\dagger \mathbf{a}_k \mathbf{x}) &= \text{Tr}(\mathbf{a}_k \mathbf{a}_k^\dagger \mathbf{x} \mathbf{x}^\dagger) \\ \text{Tr}(\mathbf{a}_k \mathbf{a}_k^\dagger \mathbf{x} \mathbf{x}^\dagger) &= \text{Tr}(\mathbf{A}_k \mathbf{X}), \end{aligned} \quad (2.36)$$

em que \mathbf{A}_k é a matriz de posto unitário formada por $\mathbf{a}_k \mathbf{a}_k^\dagger$, e \mathbf{X} por $\mathbf{x}\mathbf{x}^\dagger$. Dessa maneira, o problema de recuperação de fase assume uma forma matricial:

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x} \quad (2.37)$$

Utilizando \mathcal{A} para denotar o operador linear para mapear as matrizes definidas \mathbf{A}_k e \mathbf{X} em $\text{Tr}(\mathbf{A}_k \mathbf{X})$ para $k = 1, 2, \dots, n$, o problema PR é equivalente a:

$$\begin{aligned} &\text{encontre} && \mathbf{X} \\ &&& \mathcal{A}(\mathbf{X}) = \mathbf{y}, \\ &\text{sujeito a} && \mathbf{X} \geq 0, \\ &&& \text{posto}(\mathbf{X}) = 1. \end{aligned} \tag{2.38}$$

Dessa maneira, a recuperação da distribuição de fase é problema de minimização em torno de pedaços afins do cone semi-definido positivo de recuperação matricial. Logo, o sistema linear mostrado na equação 2.37 é extremamente subdeterminado. Entretanto, como o algoritmo de elevação de fase procura por soluções com de posto baixo, o limite de medidas de amplitude deve ser no mínimo de $3n - 2$ para garantir unicidade, em que n é número de variáveis a serem determinadas em \mathbf{x} .

Recuperação de Fase via Programação Convexa

De acordo com [13], obter a solução de minimização de posto mostrado na equação 2.38 é um processo difícil. Dessa maneira em vez de utilizar uma funcional de posto, a técnica PL utiliza da norma do traço:

$$\begin{aligned} &\text{minimize} && \text{Tr}(\mathbf{X}), \\ &\text{sujeito a} && \mathcal{A}(\mathbf{X}) = \mathbf{y}, \\ &&& \mathbf{X} \geq 0, \end{aligned} \tag{2.39}$$

Essa técnica fornece a base necessária para utilização da programação semi-definida (SDP), do inglês *Semidefinite Programming*, já que condição $\mathbf{X} \geq 0$ implica que \mathbf{X} seja a hermitiana semi-definida positiva. Com essa condição o problema torna-se convexo [13]. Além disso, a formulação apresentada na equação 2.39 apresentam boa performance empírica, já que o operador \mathcal{A} não necessita obedecer às condições adicionais na maioria das aplicações, possuindo uma alta probabilidade de recuperar a grandeza de interesse \mathbf{x} com uma distribuição de fase global.

O resultado é obtido a partir de um número de medições de $n \log(n)$, onde n é o número de coeficientes a serem determinados. Outro benefício da normalização de traço é a possibilidade de lidar com ruído presente na aquisição de medições de amplitude, em que tal ruído pode ser modelado utilizando diferentes distribuições de probabilidade. Supondo que o vetor de medições \mathbf{y}_k possua uma uma distribuição de probabilidade $p(\boldsymbol{\mu})$, em que $\boldsymbol{\mu} = \mathbf{A}\mathbf{x}$ representa o valor da média das amostras. Dessa maneira, pode-se considerar que as medições em amplitude assumem a forma $\boldsymbol{\mu}_k = |\langle \mathbf{a}_k, \mathbf{x} \rangle|^2$. Sendo assim, a clássica abordagem

de maximização assume a forma:

$$\begin{aligned} & \text{maximize} && p(\mathbf{y}, k) \\ & \text{sujeito a} && \mathbf{A}\mathbf{x} = \boldsymbol{\mu}. \end{aligned} \tag{2.40}$$

Otimizando $\boldsymbol{\mu}$ e \mathbf{x} . Aplicando a técnica de elevação de fase e o comportamento monotônico da função logarítmica, o problema é reescrito da seguinte maneira:

$$\begin{aligned} & \text{minimize} && -\log [p(\mathbf{y}, \mathbf{x})] \\ & \text{sujeito a} && \boldsymbol{\mu} = \mathcal{A}(\mathbf{X}) \\ & && \mathbf{X} \geq 0, \\ & && \text{posto}(\mathbf{X}) = 1. \end{aligned} \tag{2.41}$$

Que é um problema não tratável [13]. Uma outra formulação pode ser utilizada para regularizar o problema e permitir a sua solução, que é:

$$\begin{aligned} & \text{minimize} && -\log [p(\mathbf{y}, \boldsymbol{\mu})] + \lambda \text{Tr}(\mathbf{X}), \\ & \text{sujeito a} && \boldsymbol{\mu} = \mathcal{A}(\mathbf{X}) \\ & && \mathbf{X} \geq 0, \\ & && \text{posto}(\mathbf{X}) = 1. \end{aligned} \tag{2.42}$$

Que otimiza $\boldsymbol{\mu}$ e \mathbf{X} , encontrando-o tal que $\log [p(\mathbf{y}, \mathcal{A}(\mathbf{X}))]$ seja mínimo. A constante de multiplicação λ é o fator de regularização. Considerando que a distribuição probabilidade seja logarítmica côncava, a equação 2.42 é convexa e com solução. Em [13] mostra a equação 2.42 para algumas distribuições conhecidas que apresentam distribuições logarítmicas côncavas. Primeiramente considerando uma distribuição de Poisson do tipo:

$$\sum_k [-\boldsymbol{\mu}_k + \mathbf{y}_k \log (\boldsymbol{\mu}_k)], \tag{2.43}$$

o problema apresentado na equação 2.42 torna-se:

$$\begin{aligned} & \text{minimize} && \sum_k [\boldsymbol{\mu}_k - \mathbf{y}_k \log (\boldsymbol{\mu}_k)] + \lambda \text{Tr}(\mathbf{X}) \\ & \text{sujeito a} && \boldsymbol{\mu} = \mathcal{A}(\mathbf{X}) \\ & && \mathbf{X} \geq 0, \\ & && \text{posto}(\mathbf{X}) = 1, \end{aligned} \tag{2.44}$$

para uma sequência de amostras independentes \mathbf{y}_k que formam a distribuição. Outro exemplo é a distribuição gaussiana, com distribuição logarítmica côncava:

$$-\sum_k \frac{1}{\sigma_k^2} (\mathbf{y}_k - \boldsymbol{\mu}_k)^2. \tag{2.45}$$

Logo, a formulação na equação 2.42 torna-se:

$$\begin{aligned} & \text{minimize} && \sum_k \frac{1}{\sigma_k^2} (\mathbf{y}_k - \boldsymbol{\mu}_k)^2 + \lambda \text{Tr}(\mathbf{X}) \\ & \text{sujeito a} && \boldsymbol{\mu} = \mathcal{A}(\mathbf{X}) \\ & && \mathbf{X} \geq 0, \\ & && \text{posto}(\mathbf{X}) = 1, \end{aligned} \quad (2.46)$$

em que σ_k é o desvio padrão associado. Caso a equação 2.46 apresenta uma matriz diagonal com elementos σ_k^2 , a mesma pode ser reescrita da seguinte forma:

$$- \sum_k \frac{1}{\sigma_k^2} (\mathbf{y}_k - \boldsymbol{\mu}_k)^2, \quad (2.47)$$

em que a formulação na equação 2.42 torna-se:

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \{ -\mathcal{A}(X) * \sum^{-1} [\mathbf{y} - \mathcal{A}(X)] \} + \lambda \text{Tr}(\mathbf{X}) \\ & \text{sujeito a} && \boldsymbol{\mu} = \mathcal{A}(\mathbf{X}) \\ & && \mathbf{X} \geq 0, \\ & && \text{posto}(\mathbf{X}) = 1. \end{aligned} \quad (2.48)$$

Essas formulações são convexas e retornam para condição para inexistência de ruído quando $\lambda \rightarrow 0^+$. Assim como foi apresentado nos algoritmos iterativos [11, 12], a técnica de elevação de fase também pode ser melhorada com restrições, como mostrado por [13], que com o objetivo de impor valores reais, usa a restrição que os elementos da matriz \mathbf{X} sejam positivos, ou seja:

$$\mathbf{X}_{ij} \geq 0, \quad (2.49)$$

mostrando que tal técnica é flexível e que pode ser adaptada para utilização nas mais diferentes aplicações de utilização de vários tipos de restrições.

2.2.1.5 Algoritmo de Maximização de Fase

O algoritmo de maximização de fase, ou do inglês *Phasemax Algorithm* (PM), de otimização convexa para medições de amplitude quadráticas. Diferentemente da elevação de fase, a técnica PA executa a relaxamento na dimensão original do sinal. Considerando \mathbf{x} é o, a formulação das medições quadráticas pode ser feita de maneira similar a apresentada anteriormente:

$$\mathbf{y}_k = |\langle \mathbf{a}_k, \mathbf{x} \rangle|^2 + \alpha_k, \quad k = 1, 2, \dots, m. \quad (2.50)$$

Nessa equação, $\mathbf{a}_k \in \mathbb{C}^n$ são os vetores de detecção, $\alpha_k \in \mathbb{R}^n$ é o ruído associado para $k \geq n$. Para resolver o problema, a técnica PM utiliza um vetor $\hat{\mathbf{x}}$, da seguinte forma [46]:

$$\begin{aligned} & \text{maximize} && \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\mathbb{R}} \\ & \text{sujeito a} && |\langle \mathbf{a}_k, \mathbf{x} \rangle| \leq \mathbf{y}_k, \end{aligned} \quad (2.51)$$

em que $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\mathbb{R}}$ denota a parte real do produto interno entre o vetor dos coeficientes a serem determinados e sua estimação. A equação 2.51 demonstra a principal ideia por trás da maximização de fase: encontrar o vetor \mathbf{x} mais próximo da estimação $\hat{\mathbf{x}}$ e que ao mesmo tempo satisfaça a restrição imposta pela equação 2.50, empregando para isso técnicas para diminuir os limites inferiores da distribuição de probabilidade, em que a PM consegue recuperar o sinal de interesse sem afetar as restrições quadráticas envolvidas na aquisição de amplitude, [46].

Isso se aplica independente da ambiguidade existente na distribuição de fase do sistema. No cenário livre da influência do ruído, o ângulo entre o vetor dos coeficientes e sua estimação é definido como:

$$\hat{\text{ângulo}}(\mathbf{x}, \hat{\mathbf{x}}) = \arccos \left(\frac{\langle \mathbf{x}, \hat{\mathbf{x}} \rangle}{\|\mathbf{x}\|_2 \|\hat{\mathbf{x}}\|_2} \right). \quad (2.52)$$

Considerando que:

$$\beta = 1 - \frac{2}{\pi} \hat{\text{ângulo}}(\mathbf{x}, \hat{\mathbf{x}}), \quad (2.53)$$

é uma constante que calcula a acurácia de aproximação, a probabilidade de o algoritmo PM conseguir recuperar o vetor de que contém os coeficientes de interesse é:

$$p_{\mathbb{C}}(m, n) \geq 1 - e^{-\frac{(\beta m - 4n)^2}{2m}}, \quad (2.54)$$

para todo $\beta m > 4n$ [46]. Dessa maneira caso $m > 4n/\beta$ para $\beta > 0$ a técnica de elevação de fase terá sucesso em recuperar \mathbf{x} com probabilidade não nula. Além disso, caso o vetor de estimação não seja ortogonal ao vetor \mathbf{x} de dimensão fixa n , é possível aumentar a probabilidade de sucesso da técnica PM aumentando a quantidade de medições m [46].

A técnica de elevação é muito semelhante a técnica de busca de base [47, 48]:

$$\begin{aligned} & \text{minimize} && \|\mathbf{z}\|_1, \\ & \text{sujeito a} && \hat{\mathbf{x}} = \mathbf{A}\mathbf{B}^{-1}\mathbf{z}, \end{aligned} \quad (2.55)$$

em que $\mathbf{B} = \text{diag}(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)$ e \mathbf{A} é a matriz de detecção. Dessa maneira, caso a técnica seja bem-sucedida as fases do vetor de solução $\mathbf{z} \in \mathbb{C}^n$ conterà as fases perdidas durante o processo de aquisição, em que [46]:

$$\mathbf{y}_k = \text{fase}(\mathbf{z})\mathbf{b}_m = \langle \mathbf{a}_k, \mathbf{x} \rangle, \text{ para } k = 1, 2, \dots, m. \quad (2.56)$$

Com $\text{fase}(\mathbf{z}) = \mathbf{z}/|\mathbf{z}|$ para $\mathbf{z} \neq 0$ e $\text{fase}(0) = 1$. Tal relação demonstra a conexão entre a recuperação de fase e a recuperação de sinal a partir de matrizes esparsas e que a busca de fase pode ser utilizada para processos que usam medições em amplitude. Para obter o vetor de estimação $\hat{\mathbf{x}}$ várias técnicas podem ser utilizadas como inicializador espectral [49] e suas variantes otimizadas [50, 51], o inicializador *Null* [52], o método provedor de ortogonalidade [53] e o clássico mínimos quadrados [54], com alta possibilidade de recuperação de fase, caso for fornecida uma quantidade adequada de medições ao algoritmo.

2.2.1.6 Algoritmo de Separação da Fase

A técnica de separação de fase, do inglês *Phasecut* (PC), é mais uma técnica para resolver o problema de recuperação de fase através do relaxamento de problemas matriciais com restrições quadráticas. Diferente das técnicas PL e PM, a técnica PC separa de forma explícita a amplitude e a fase da medição experimental, com foco na otimização das variáveis associadas à distribuição de fase.

Considerando o cenário livre da influência do ruído nas medições, a técnica PC procura um vetor \mathbf{z} que contenha a distribuição de fases da seguinte maneira:

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{A}\mathbf{x} - \text{diag}(\mathbf{y})\mathbf{z}\|_2^2, \\ & \text{sujeito a} \quad |\mathbf{z}_i| = 1. \end{aligned} \quad (2.57)$$

Logo, ambos os vetores \mathbf{z} e \mathbf{x} pertencem à \mathbb{C}^n . A equação 2.57 mostra que a otimização é feita em ambos os vetores \mathbf{z} . Tal formulação permite resolver o problema de minimização em \mathbf{x} utilizando mínimos quadrados:

$$\mathbf{x} = \mathbf{A}^\dagger \text{diag}(\mathbf{y})\mathbf{z}. \quad (2.58)$$

Substituindo a equação 2.58 na equação 2.57 obtém-se:

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{A}\mathbf{A}^\dagger \text{diag}(\mathbf{y})\mathbf{z} - \text{diag}(\mathbf{y})\mathbf{z}\|_2^2, \\ & \text{sujeito a} \quad |\mathbf{z}| = 1. \end{aligned} \quad (2.59)$$

A parte da minimização da equação 2.59 pode ser melhorada da seguinte forma:

$$\begin{aligned} \|\mathbf{A}\mathbf{A}^\dagger \text{diag}(\mathbf{b})\mathbf{z} - \text{diag}(\mathbf{y})\mathbf{z}\|_2^2 &= \|(\mathbf{A}\mathbf{A}^\dagger - \mathbf{I}) \text{diag}(\mathbf{y})\mathbf{z}\|_2^2 \\ \|(\mathbf{A}\mathbf{A}^\dagger - \mathbf{I}) \text{diag}(\mathbf{y})\mathbf{z}\|_2^2 &= \mathbf{z}^* \text{diag}(\mathbf{y}^T) \widetilde{\mathbf{M}} \text{diag}(\mathbf{y})\mathbf{z}. \end{aligned} \quad (2.60)$$

Nessa equação, $\widetilde{\mathbf{M}} = (\mathbf{A}\mathbf{A}^\dagger - \mathbf{I})(\mathbf{A}\mathbf{A}^\dagger - \mathbf{I}) = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$. Considerando que $\mathbf{M} = \text{diag}(\mathbf{y}^T) \widetilde{\mathbf{M}} \text{diag}(\mathbf{y})$ o problema de minimização torna-se:

$$\begin{aligned} & \text{minimize} \quad \mathbf{z}^* \mathbf{M} \mathbf{z}, \\ & \text{sujeito a} \quad |\mathbf{z}_k|^2, \end{aligned} \quad (2.61)$$

para $k = 1, 2, \dots, n$. Por definição, a matriz \mathbf{M} é hermitiana semidefinida positiva, e que a formulação $(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)$ denota a projeção ortogonal da imagem de \mathbf{A} .

Dessa maneira, a equação 2.61 minimiza no vetor de fases \mathbf{z} a norma da componente $\text{diag}(\mathbf{y})\mathbf{z}$, mesmo ela não sendo a imagem de \mathbf{A} . Após a transformação do problema de recuperação de fase em um problema de minimização quadrática, o algoritmo de separação

otimiza cada elemento do vetor de distribuição de fase \mathbf{z} . Tal problema é equivalente a:

$$\begin{aligned} & \text{minimize} && \bar{\mathbf{z}}_k \mathbf{M}_{kk} \mathbf{z}_k + 2 \operatorname{Re} \left(\sum_{j \neq 1} \bar{\mathbf{z}}_j \mathbf{M}_{jk} \mathbf{z}_k \right), \\ & \text{sujeito a} && |\mathbf{z}_k|^2. \end{aligned} \quad (2.62)$$

Nessa equação, \mathbf{z}_k são coeficientes a serem determinados e \mathbf{z}_j permanecem constantes. Devido ao fato de $|\mathbf{z}_k| = 1$, a resolução de:

$$\begin{aligned} & \text{minimize} && \operatorname{Re} \left(\mathbf{z}_k \sum_{j \neq 1} \mathbf{M}_{kj} \bar{\mathbf{z}}_j \right), \\ & \text{sujeito a} && |\mathbf{z}_k| = 1. \end{aligned} \quad (2.63)$$

que pertence o termo de minimização da equação 2.62 implica que o valor de \mathbf{z}_k seja:

$$\mathbf{z}_k = - \frac{\sum_{j \neq 1} \mathbf{M}_{kj} \bar{\mathbf{z}}_j}{\left| \sum_{j \neq 1} \mathbf{M}_{kj} \bar{\mathbf{z}}_j \right|}, \quad (2.64)$$

para cada valor de $k = 1, 2, \dots, n$, o que faz \mathbf{z} a solução ótima para equação 2.60, que pode ser resolvido utilizando abordagens iterativas parecidas com a técnica GB, a diferença reside no número de operações. Na técnica PC a quantidade de operações é de n^2 para atualizar todos os elementos \mathbf{z}_k , ao passo que no algoritmo de Gerchberg-Saxton e derivados utiliza de $n \log n$ operações para o mesmo propósito.

De forma mais generalizada, pode-se definir $\mathbf{Z} = \mathbf{z}\mathbf{z}^*$ e dessa maneira a equação 2.61 torna-se:

$$\begin{aligned} & \text{minimize} && \operatorname{Tr}(\mathbf{Z}\mathbf{M}), \\ & \text{sujeito a} && \operatorname{diag}(\mathbf{Z}) = 1, \text{ e} \\ & && \operatorname{posto}(\mathbf{Z}) = 1. \end{aligned} \quad (2.65)$$

O relaxamento convexo pode ser aplicado à equação 2.66 retirando a restrição do posto da seguinte forma:

$$\begin{aligned} & \text{minimize} && \operatorname{Tr}(\mathbf{Z}\mathbf{M}), \\ & \text{sujeito a} && \operatorname{diag}(\mathbf{Z}) = 1, \text{ e} \\ & && \mathbf{Z} \succeq 0. \end{aligned} \quad (2.66)$$

Logo, caso a solução seja de posto unitário o relaxamento é convexo e mostra que o vetor \mathbf{Z} para $\mathbf{Z} = \mathbf{z}\mathbf{z}^*$ é a solução otimizada para a PR. No cenário em que o posto é maior que um, um autovetor \mathbf{v} de \mathbf{Z} é utilizado como uma solução aproximada, enquanto $\mathbf{Z} - \mathbf{v}\mathbf{v}^T$ fornece a incerteza em torno dos coeficientes de \mathbf{v}_k .

Entretanto, a abordagem SDP mostrada acima é raramente projetada para lidar com problemas que envolvem matrizes hermitianas. Neste caso uma reformulação de programas que lidam com números do domínio complexo para programas equivalentes no domínio real \mathbb{R}^{2n} .

Definindo um operador:

$$\mathcal{T}(\mathbf{C}) = \begin{bmatrix} \text{Re}(\mathbf{C}) & -\text{Im}(\mathbf{C}) \\ \text{Im}(\mathbf{C}) & \text{Re}(\mathbf{C}) \end{bmatrix}, \quad (2.67)$$

em que \mathbf{C} é uma matriz pertencente a \mathbb{C}^n . Considerando agora outra matriz \mathbf{D} , obtém-se:

$$\text{Tr}[\mathcal{T}(\mathbf{C})\mathcal{T}(\mathbf{D})] = 2 \text{Tr}(\mathbf{CD}). \quad (2.68)$$

Dessa maneira, o autovalor $c = a + bi$ é o autovetor de \mathbf{C} , que possui autovalor λ , se, e somente se:

$$\begin{bmatrix} a \\ b \end{bmatrix} \text{ e } \begin{bmatrix} -b \\ a \end{bmatrix}, \quad (2.69)$$

são os autovetores de \mathcal{T} , com auto valor λ , significado que $\mathbf{C} \geq 0$ se, e somente se, $\mathcal{T}(\mathbf{C}) \leq 0$, formulando a programação semi-definida de recuperação de fase em \mathbb{R}^{2n} de matrizes simétricas como:

$$\begin{aligned} & \text{minimize} && \text{Tr}[\mathcal{T}(\mathbf{X})] \\ & \text{sujeito a} && \mathbf{X}_{kk} + \mathbf{X}_{(n+k)(n+k)} = 2, \text{ e} \\ & && \mathbf{X} \succeq 0. \end{aligned} \quad (2.70)$$

Nessa equação, $\mathbf{X}_{kj} = \mathbf{X}_{(n+k)(n+j)}$ e $\mathbf{X}_{(n+k)j} = -\mathbf{X}_{k(n+j)}$, para $k, j = 1, 2, \dots, n$. Devido às simetrias do operador $\mathbf{T}(\mathbf{M})$ as restrições que reforçam tais características podem ser relaxas, transformando a equação 2.70 a uma formulação semelhante àquela apresentada para a técnica de separação de fase em \mathbb{R}^{2n} , ou seja:

$$\begin{aligned} & \text{minimize} && \text{Tr}[\mathbf{T}(\mathbf{A})\mathbf{X}], \\ & \text{sujeito a} && \text{diag}(\mathbf{X}), \text{ e} \\ & && \mathbf{X} \succeq 0. \end{aligned} \quad (2.71)$$

Para $\mathbf{X} \in \mathbb{R}^{2n}$. Essa formulação traz benefícios computacionais para a determinação da distribuição de fase, e pode ser resolvida de diversas maneiras como o método dos pontos interiores, indicado para problemas de pequena escala, métodos de primeira ordem para quando o número de incógnitas começa a ultrapassar uma ordem de grandeza de 10^2 , a técnica da descida do bloco de coordenadas que é mais indicada quando os limites complexos não estão explicitamente definidos.

2.3 Aprendizado de Máquina

O conceito de aprendizado de máquina, do inglês *Machine Learning* (ML), pode ser definido como a utilização de algoritmos e modelos estatísticos para que computadores possam compreender uma determinada base de dados, estabelecer correlações e executar tarefas específicas após um processo de aprendizagem [55]. Esta seção apresenta os conceitos gerais

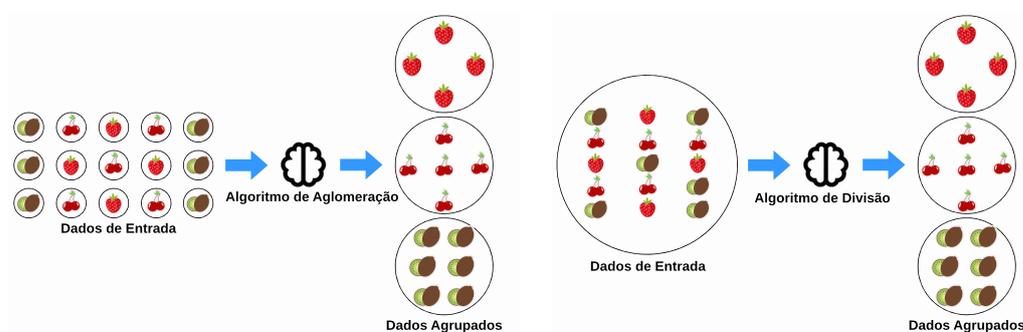
de aprendizagem de máquina como os tipos de estratégias e algoritmos de aprendizagem, tendo como foco a regressão supervisionada, além de discutir sobre testes estatísticos e de hipóteses que visam principalmente certificar a integridade dos modelos gerados para previsões e selecionar os melhores de acordo com as necessidades e expectativas do operador.

2.3.1 Aprendizagem Não Supervisionada

A aprendizagem não supervisionada, do inglês *Unsupervised Learning* (UL), denota a classe de estratégias de ML que não possui a necessidade de indicação de quais são as variáveis independentes, também conhecidas como elementos preditores ou *labels*, e a variável dependente, também conhecida como alvo ou *target*, [55] utilizando técnicas de agrupamento para detectar similaridade entre amostras e agrupá-las, não necessitando de uma base de dados de treinamento para o processo de aprendizagem, em que uma nova entrada será alocada a um determinado grupo quando suas características são compatíveis com o mesmo [56].

A não necessidade de uma base de dados de treinamento é a principal diferença entre a aprendizagem não supervisionada e a supervisionada. Além disso, tais técnicas não possuem como garantia obter uma solução global ótima para a aplicação para qual estão sendo usados, apesar de utilizar técnicas de otimização [56]. Tal comportamento pode ser explicado que para obter uma solução global, todos os grupos presentes da base de dados devem ser analisados, o que não é viável mesmo que a quantidade de dados seja baixa, sendo necessário a utilização de aproximações heurísticas no processo. A utilização das técnicas de UL precisam atender alguns requisitos, que são: seleção das amostras a serem agrupadas, determinação dos critérios de agrupamento, escolha adequada da métrica que mede o nível de similaridade entre as amostras e utilização do algoritmo que executará o agrupamento. As estratégias de UL podem ser separadas em dois grupos: a abordagem hierárquica ou a

Figura 2.14: Exemplificação do funcionamento dos algoritmos de agrupamento por aglomeração, figura (a) e divisão, figura (b). Tais técnicas são chamadas de algoritmos hierárquicos pois organizam os dados de maneira bem definida em agrupamentos específicos.



(a) Aglomeração

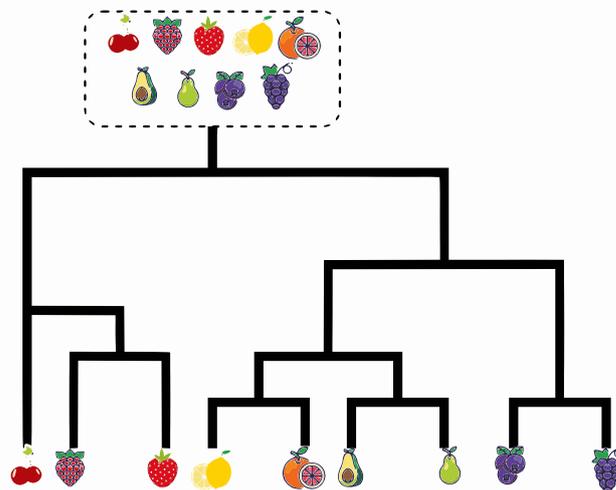
(b) Divisão

abordagem de particionamento [56]. No que diz respeito à abordagem hierárquica ela ainda pode ser separada em algoritmos de aglomeração e de divisão. No processo de aprendiza-

gem por aglomeração, o algoritmo inicia com grupos com apenas um elemento, na medida que o número de iterações aumenta as classes unitárias mais semelhantes são alocados em novas classes maiores, isso é exemplificado na figura 2.14-(a). Em contrapartida na divisão todas as amostras começam em um único grupo, que é dividido em outros que contém elementos semelhantes à medida que o número de iterações aumenta como ilustrado na figura 2.14-(b). Ambas as abordagens organizam os dados em estruturas com hierarquia bem definida, com uma estrutura em cascata em árvore que engloba desde as classes mais gerais até aquelas extremamente bem agrupadas e específicas, um exemplo é mostrado na figura 2.15. Dessa maneira é possível visualizar a interrelação entre os diferentes grupos, e o seu grau de proximidade.

Já as técnicas de particionamento possuem um funcionamento que lembra em algum grau a divisão, contudo enquanto a divisão pode ser classificada como um algoritmo determinístico o particionamento pode ser considerado como um método estocástico [56]. Primeiramente um número finito de classes são criadas e uma quantidade de amostras são alocadas nos mesmos, e à medida que iterações ocorrem seguindo uma série de critérios estabelecidos, as amostras são enviadas para os grupos para os quais suas características têm uma maior compatibilidade. Essa dinâmica é executada até que cada grupo tenha uma distribuição mais homogênea, ou seja, quando não há mudanças significativas na classe, ou até que o limite de iterações seja atingido.

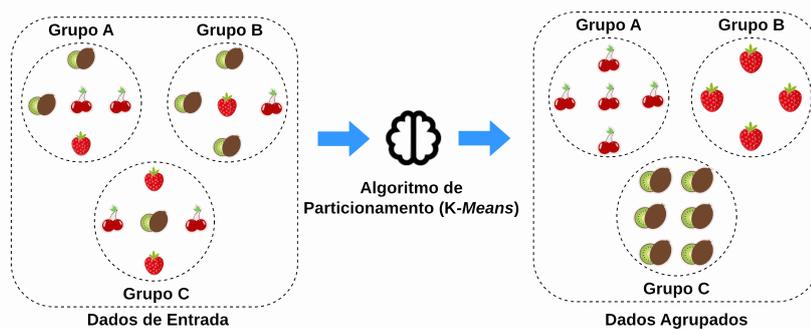
Figura 2.15: Exemplificação de uma estrutura em cascata gerada pelos algoritmos de agrupamento hierárquico. Tal árvore permite a identificação do nível de interrelação entre os diferentes grupos e o grau de proximidade entre eles.



A técnica mais famosa de particionamento é a *k-means* [56, 57]. Essa técnica separa os dados em k grupos e tenta minimizar a norma euclidiana entre as amostras e o centro do grupo, como ilustrado na figura 2.16. A técnica *k-means* ainda permite selecionar qual é o centro de um determinado grupo e a quantidade de amostras contidas no mesmo. Entretanto, apesar das técnicas de agrupamento hierárquico ou de particionamento serem as mais utilizadas, elas não são as únicas. Os modelos de mistura gaussiana particionam os dados em

agrupamentos que são organizados seguindo a distribuição normal evidenciando subpopulações no banco de dados [56], enquanto os modelos de agrupamento difusos os agrupamentos se sobrepõem formando uma espécie de diagrama de Venn, dessa maneira os dados podem pertencer a grupos com características distintas ao mesmo tempo [56].

Figura 2.16: Princípio de funcionamento do algoritmo *k-Means*. Os dados são organizados de maneira aleatória em um número finito k de grupos com centro C e raio r . À medida que as iterações ocorrem o cálculo da distância entre os elementos e os centros dos grupos é calculada e a sua localização é alterada para uma mais adequada dependendo do resultado.



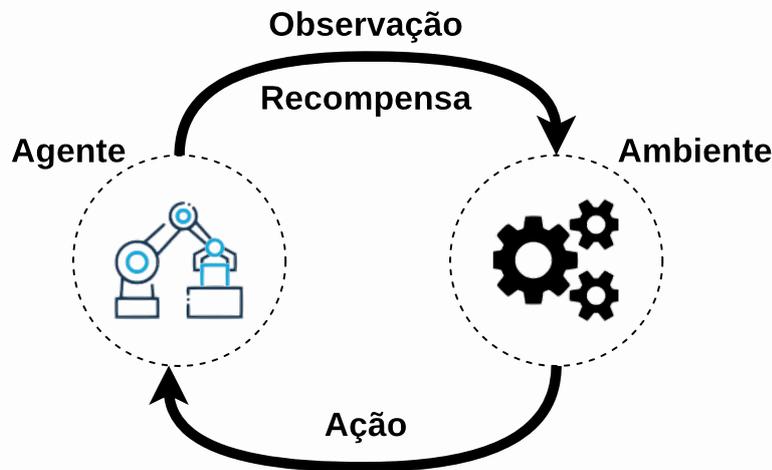
2.3.2 Aprendizagem por Reforço

A aprendizagem por reforço, do inglês *Reinforcement Learning* (RL), é a classe de técnicas de aprendizagem de máquina em que um objeto interage com o meio buscando maximizar a quantidade de recompensas, reforçando um determinado comportamento a partir das respostas à interação com meio ambiente ao seu redor [55]. Dessa maneira a inteligência artificial por tentativa e erro, consegue determinar qual a melhor maneira para executar uma tarefa específica. Esse processo é ilustrado pelo fluxograma apresentado na figura 2.17.

De acordo com [58], o ambiente de interação pode ser virtual (simuladores, jogos etc.) ou o próprio mundo real. O processo de aprendizagem por reforço pode ser resumido da seguinte maneira: em um determinado tempo t , a inteligência artificial (IA) observa um certo estado s_t do ambiente, tomando uma determinada ação a_t e recebendo uma recompensa r_t . Em seguida o estado é alterado para s_{t+1} , em que cada mudança de estado a inteligência artificial acumula as experiências e aprende com elas. Caso o processo tenha sido bem-sucedido, a IA receberá uma recompensa positiva, caso contrário receberá uma negativa, ou seja, uma punição. Esse processo continua até maximizar os resultados obtidos em todos estados apresentados pelo ambiente. Essa dinâmica apresentada pela aprendizagem por reforço permite classificá-la com um processo de decisão de Markov, ou *Markov Decision Process* (MDP) [58], em que o resultado de saída de uma ação em um determinado estado é resultado somente deste, não tendo relação alguma com os estados e ações passadas.

Por definição, um processo MDP necessita de cinco elementos geralmente denotados por (s, a, p, r, ω) , em que s denota o estado, a a ação, p a punição, r a recompensa e ω

Figura 2.17: Fluxograma que ilustra o processo de aprendizagem por reforço. O agente ou inteligência artificial observa o estado do ambiente, interage com ele por meio de uma ação, recebendo uma recompensa positiva ou negativa dependendo se a mesma foi ou não bem-sucedida, dessa maneira o agente consegue aprender a executar uma determinada tarefa para qual está sendo treinado.



ω é um número escalar entre 0 e 1 denominado com fator de desconto, responsável pela maximização das recompensas futuras [58]. O processo matemático envolvido na RL é o mapeamento dos valores das recompensas obtidas desde o início de um estado até a mudança para o próximo. Esse mapeamento é executado por uma função chamada função de valor $v(s)$ ou função de Bellman que é definida como:

$$v(s) = \max (r_t + \omega v(s_{t+1})), \quad (2.72)$$

que determina a melhor saída para um estado como a soma da recompensa atual com alguma influência do estado seguinte, que é ponderado pelo fator ω . Dessa maneira a inteligência artificial consegue determinar o melhor caminho até a execução da tarefa, acumulando os parâmetros e aprendendo durante o processo.

Existem outras maneiras de aplicar a aprendizagem por reforço. O algoritmo de programação dinâmica (*Dynamic-Programming Approach*), que considera que se a IA irá executar uma determinada tarefa em número finito de passos h , utilizando uma abordagem bayesiana para o processo de aprendizagem [59]. Tal técnica assume um conjunto de distribuição para alguns parâmetros p_i , em que a escolha mais natural é uma distribuição uniforme entre 0 e 1, e um mapeamento dos chamados estados de confiança, que são resumos das experiências da inteligência artificial em que cada um pode ser representada como uma tabulação de ações e recompensas $(a_1, r_1, \dots, a_n, r_n)$ e denotam qual ação a_n foi executada com uma recompensa r_n .

Considerando que $V^*(a_1, r_1, \dots, a_n, r_n)$ denota a quantidade restante de recompensas para um determinado número de passos. No caso $\sum_n a_n = h$ isso demonstra que não existem mais recompensas disponíveis para serem retiradas, ou seja, $V^*(a_1, r_1, \dots, a_n, r_n) = 0$. Tal

definição é a base de uma abordagem recursiva, em que caso o valor V^* seja conhecido para todos os estados de confiança restantes, é possível calcular o valor de V^* para qualquer estado de confiança em qualquer instante da seguinte forma:

$$V^*(a_1, r_1, \dots, a_n, r_n) = \max_n [\rho_n (a_1, r_1, \dots, a_n + 1, r_n + 1, \dots, a_n, r_n) + (1 + \rho_i) (a_1, r_1, \dots, a_n + 1, r_n, \dots, a_n, r_n)], \quad (2.73)$$

em que ρ_n é a probabilidade subjetiva de uma ação a_n para uma recompensa r_n . No caso de uma distribuição uniforme ou distribuição do tipo beta, tal probabilidade pode ser calculada da seguinte forma:

$$\rho_n = \frac{r_n + 1}{a_n + 1}. \quad (2.74)$$

De acordo com [59], o custo de preencher V^* é diretamente proporcional ao produto entre o número de estados de confiança e de ações.

A técnica de indexação de Gittins é outro método de aprendizagem por reforço, que propõe encontrar de maneira ótima a melhor ação em cada etapa para problemas de bandidos multiarmados [60], agindo no critério de punições [59]. Considerando um número de vezes que uma determinada ação a vezes a quantidade de vezes que a mesma recebe uma recompensa r , e que para certos valores de punição (recompensa negativa) foi utilizada uma tabela para cada ação n , $I(a_n, r_n)$. Tal tabela de verificação representa uma medida comparativa dos valores combinados da recompensa esperada para uma determinada ação e o valor da informação que poderá ser obtida como consequência da escolha de tal ato baseado no histórico de recompensas.

Em [59] é defendido que a escolha do índice n de maior valor garante um melhor balanço entre uma investigação mais igualitária, dando a mesma oportunidade para todas as ações, ou uma investigação iterativa e que dá maior peso para aquelas que apresentaram os melhores resultados. Por causa dessa característica, tal abordagem tem muitas vantagens em aplicações complexas, como manipuladores robóticos de recompensa imediata. Outro ramo da RL é a aprendizagem *automata*, que é baseada em máquinas de estados *automatas*. Um exemplo de RL utilizando tal estratégia é o algoritmo recompensa inata linear [59]. Considerando que p_n é a probabilidade da IA tomar uma ação a_n , ela assume a forma:

$$\begin{aligned} a_n &= p_n + \alpha (1 - p_n) \\ a_m &= p_m - p_m \alpha, \end{aligned} \quad (2.75)$$

para $m \neq n$. Essa situação ocorre quando a ação tomada é bem-sucedida. Na situação contrária, p_m permanece inalterada. Caso o vetor resultante seja composto por zeros e um único 1, o algoritmo irá convergir com probabilidade 1. De maneira geral, nem sempre o algoritmo consegue convergir para a ação correta. Para evitar tal cenário, a constante de proporcionalidade α pode ser manipulada para assumir valores arbitrariamente pequenos.

Os algoritmos de vantagem, do inglês *greedy algorithms*, também podem ser utilizados

na aprendizagem por reforço. Tais estratégias tentam encontrar a ação que retornará a recompensa de maior valor. Essa escolha pode parecer a mais vantajosa em um primeiro momento, contudo no cenário em que a uma ação não tão benéfica para a IA é tomada, mas com um valor de destaque de recompensa pode se ser sempre escolhida no processo de aprendizagem ofuscando a ação mais adequada, mas que não gera uma recompensa de alto valor [59]. Uma maneira de mitigar tal problema é abordagem heurística de otimismo em face da incerteza, que mantém a escolha da estratégia de maior recompensa, contudo parâmetros convictos e otimistas são colocados em todas as recompensas, evitando que alguma ação seja excluída sem um motivo bem convincente. De fato, de acordo com [59] essa abordagem diminui em muito o risco de tal ocorrência. Isso é bastante investigado nos métodos de exploração de intervalo, exploração de bônus, exploração dirigida à curiosidade e exploração de varredura mecanizada.

Já as estratégias randômicas utilizam a ação a que irá gerar a maior recompensa de maneira aleatória com probabilidade p . No geral tais estratégias começam o processo de aprendizagem com alto valores de p , mas que diminuem lentamente nas próximas iterações. Apesar de simples implementação, tal estratégia pode dar prioridade a ações pouco proveitosas. Em [59] é apresentado a estratégia de Boltzmann, que mitiga o problema da seguinte maneira: a recompensa esperada r é usada para determinar uma ação a , $ER(a)$, de acordo com a seguinte distribuição de probabilidade:

$$P(a) = \frac{e^{\frac{ER(a)}{T}}}{\sum_{a'} e^{\frac{ER(a')}{T}}}, \quad (2.76)$$

em que $a' \in A$ em que A simboliza o conjunto de todas as ações. O parâmetro T , chamado de temperatura por [59], decresce lentamente. Essa estratégia possui uma boa performance quando existe uma grande separação entre as ações, caso contrário ele apresenta dificuldade na seleção da melhor ação. Além disso, caso o parâmetro T não seja alterado manualmente a convergência será lenta.

A última classe de algoritmos abordadas por [59] são aquelas baseadas em intervalos que utilizam informações probabilísticas secundárias da ação a ser tomada, por exemplo a variância. O algoritmo de intervalos de Kaelbling que armazena dados estatísticos de todas as ações a_n ligados ao números de tentativas e o número de sucessos, em que as probabilidades de todas as ações são calculadas e caso o resultado de uma determinada ação seja superior a um grau de confiabilidade de $1 - \alpha$, em que α é o erro mínimo aceitável, ela é selecionada. Essa metodologia funciona muito bem em abordagens empíricas, como no problema na determinação de tratamentos com o menor número de experimentos possíveis.

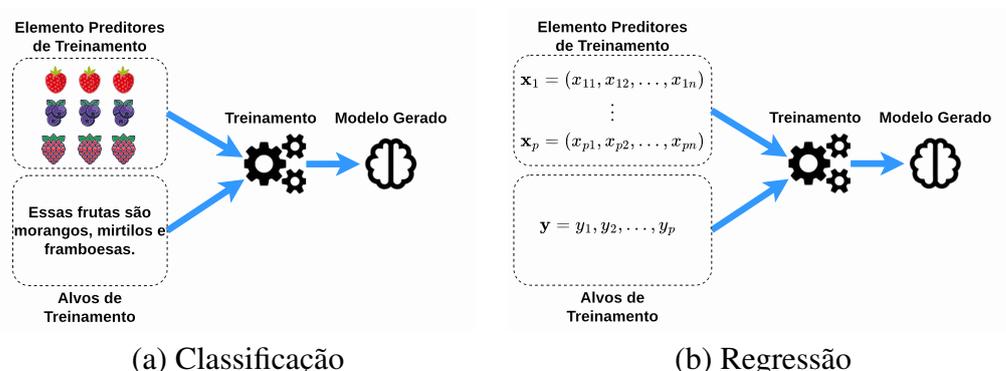
2.3.3 Aprendizagem Supervisionada

A aprendizagem de máquina supervisionada, ou *Supervised Learning* (SL), é o terceiro tipo de abordagem para geração de inteligências artificiais. Diferentemente da aprendizagem

não supervisionada e por reforço, a técnica SL trabalha com dados separados por classes, em que as variáveis independentes e dependente são determinadas por um operador externo que guia como as relações serão estabelecidas entre as entradas e saídas [55].

Dessa maneira, para o processo de aprendizagem dos algoritmos SL seja bem-sucedido é necessário a construção de uma estrutura de dados bem definida e que estabeleça de maneira clara quais são os elementos preditores e a variável alvo, tal estrutura é a base de dados ou *database* [61]. A partir dessa estrutura os algoritmos são treinados, e caso tudo ocorra de maneira adequada, o resultado obtido são modelos de inteligência artificial que podem ser utilizados para fazer previsões a partir de novas entradas. Geralmente a aprendizagem su-

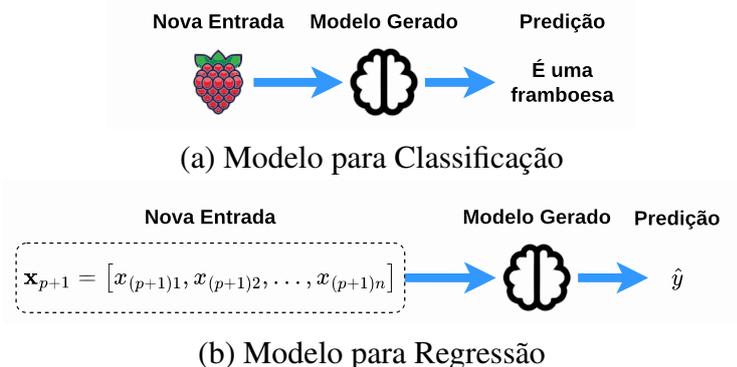
Figura 2.18: Ilustração do processo de aprendizagem supervisionada para as aplicações de classificação, figura (a), e regressão, figura (b). Neste tipo de aprendizagem de máquina é necessário indicar quais são as variáveis independentes e dependente no processo de treinamento.



pervisionada é utilizada na resolução de problemas de classificação ou regressão [55, 61]. A classificação pode ser definida como o treinamento de modelos de inteligência artificial para detectar a qual classe uma determinada entrada pertence, como mostrado na figura 2.18-(a). Os problemas de classificação podem apresentar dois tipos de variável alvo: binária ou multiclasse. Em problemas de alvos binários geralmente os modelos são gerados para classificar os resultados em dois grandes grupos, por exemplo um modelo de inteligência artificial pode ser gerado para classificar se um aluno foi aprovado ou não dependendo de suas notas ao longo do ano. Já nos problemas multi-classe a variável alvo pode pertencer a mais de um grupo, isso é ilustrado na figura 2.19-(a). Tal abordagem é muito comum em aplicações de detecção de objetos, doenças e reconhecimento facial. Por outro lado a regressão trabalha com dados contínuos, em que os modelos são gerados e treinados para estabelecer a melhor relação entre elementos preditores e a variável alvo [55] como ilustrado nas figuras 2.18-(b) e 2.19-(b). Dessa maneira a regressão baseada em SL é muito utilizada para previsão de preços, valores de ações, número de vendas, detecção de padrões de consumo além de conseguir estabelecer correlações e obter saídas condizentes e cruciais nos mais diversos campos de aplicação.

Devido a quantidade enorme de dados e aplicações presentes nos mais diversos campos, é muito comum que seja necessário usar múltiplas abordagens para conseguir resolver os

Figura 2.19: Ilustração da obtenção das predições tanto para classificação e regressão, figuras (a) e (b) respectivamente. Após a obtenção do modelo na etapa de treinamento é possível usar o mesmo para executar predições para novas entradas.



problemas de interesse de maneira ótima. Neste contexto, tanto para classificação e para regressão existem muitos algoritmos de aprendizagem de máquina que podem ser usados, sendo os mais comuns a regressão linear [62], as árvores de decisão [63], florestas randômicas [64], máquinas de vetores de suporte [65], k -vizinhos próximos [66] e redes neurais [67, 68].

2.3.3.1 Regressão Linear

A regressão linear, *Linear Regression* (LR), é o modelo mais clássico e famoso de aprendizagem de máquina. De maneira geral tal algoritmo é utilizado para criação de modelos em que a relação entre os elementos preditores e o alvo resulta em uma combinação linear do tipo [69, 70, 62]:

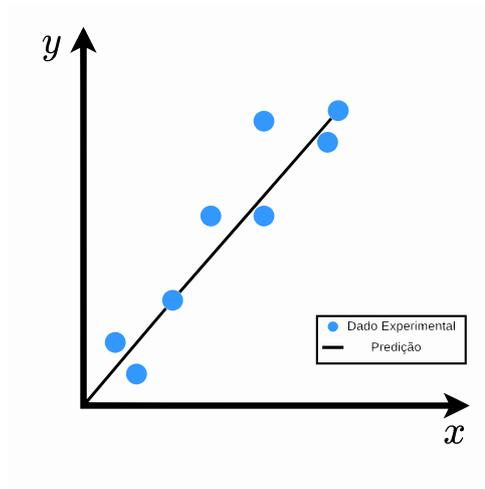
$$\hat{y}(w, x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n, \quad (2.77)$$

em que o objetivo é determinar os coeficientes w_n que melhor se adequam aos dados, obtendo a melhor estimacão \hat{y} , como ilustrado na figura 2.20. Ou seja, o algoritmo adequa tais coeficientes minimizando o quadrado das somas entre os alvos obtidos experimentalmente y e as estimacões \hat{y} resolvendo o famoso problema de mínimos quadrados:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{x}\mathbf{W} - \mathbf{y}\|_2^2, \\ \text{tal que} \quad & \mathbf{x}\mathbf{W} - \mathbf{y} \leq \epsilon. \end{aligned} \quad (2.78)$$

Em que \mathbf{x} simboliza o vetor composto pelo elementos preditores, as variáveis independentes, \mathbf{W} a matriz composta pelos coeficientes w e \mathbf{y} são as observações experimentais e ϵ denota o limite de erro aceitável entre a estimacão \hat{y} e \mathbf{y} . No processo de resolução da RL é muito comum também utilizar estratégias regularizadas, como mostrado em [69]. Outra estratégia de regressão linear para SL é a estratégia LASSO (*Linear Absolute Shrinkage and Selection Operator*) [71], que é muito utilizada em casos em que os coeficientes são esparsos, como em aplicações de *compressive sensing*. Em termos gerais, tal técnica pode ser definida da

Figura 2.20: Ilustração do processo de aprendizagem de máquina para regressão linear. A linha preta simboliza o melhor resultado do processo de aprendizagem obtido a partir da resolução das equações 2.78 e 2.79 para os pontos experimentais em azul.



seguinte maneira:

$$\text{minimize } \frac{1}{2n} \|\mathbf{x}\mathbf{W} - \mathbf{y}\|_2^2 + \beta \|\mathbf{W}\|_1, \quad (2.79)$$

$$\text{tal que } \mathbf{x}\mathbf{W} - \mathbf{y} \leq \epsilon,$$

em que β é um coeficiente geralmente menor que 1 e $\|\mathbf{W}\|_1$ é a norma l_1 de \mathbf{W} .

2.3.3.2 Árvores de Decisão

As árvores de decisão, do inglês *Decision Trees* (DT), são um tipo de algoritmo de aprendizagem supervisionada não paramétrica, utilizando regras simples inferidas da base de dados para criar modelos de inteligência artificial e fazer as previsões [63, 69]. Como todo algoritmo de SL, o algoritmo DT recebe como entrada um vetor com as variáveis independentes para obter uma saída de classificação ou regressão, como ilustrado na figura *chap2 – fig8*. Quanto mais complexa a árvore, maior o grau de complexidade das regras e mais bem adaptado o modelo gerado é aos dados.

Matematicamente, para um vetor $\mathbf{x} \in \mathbb{R}^n$, em que n simboliza o número de variáveis independentes e $\mathbf{y} \in \mathbb{R}^m$, para $0 < m \leq n$ o algoritmo DT cria partições de elementos semelhantes de maneira recursiva. Para visualizar tal dinâmica é preciso considerar um conjunto de dados O_l pertencente a um nó l com n_l amostras [69]. Para cada candidato de separação $c = (j, t_l)$ com aspecto j e limiar t_l , o particionamento de O_l em dois outros grupos $O_l^L(c)$ e $O_l^R(c)$, em que L denota o grupo ao lado esquerdo do nó pós separação e R o lado direito, é feito da seguinte maneira:

$$\begin{aligned} O_l^L(c) &= (x, y) & \text{caso } x_j \leq t_l, \text{ e} \\ O_l^R(c) &= O_l / O_l^L(c) & \text{caso } x_j \geq t_l. \end{aligned} \quad (2.80)$$

Logo, a qualidade de um candidato de separação em um nó l é calculada a partir de uma função de custo ou impureza H , que é selecionada dependendo da aplicação (regressão ou classificação), com a forma [72]:

$$G(O_l, c) = \frac{n_l^L}{n_l} H [O_l^L(c)] + \frac{n_l^R}{n_l} H [O_l^R(c)], \quad (2.81)$$

Para aplicações de regressão a função de custo H pode ser [69]:

$$\begin{aligned} H(O_l) &= \frac{1}{n_l} \sum_{n_l} (y - \hat{y}_l)^2, & \text{Erro Quadrático Médio, e} \\ H(O_l) &= \frac{1}{n_l} \sum_{n_l} \left[y \log \left(\frac{y}{\hat{y}_l} \right) - y + \hat{y}_l \right], & \text{Poisson} \\ H(O_l) &= \frac{1}{n_l} \sum_{n_l} |y - \hat{y}_l| & \text{Erro Absoluto Médio.} \end{aligned} \quad (2.82)$$

A qualidade G pode ser utilizada para minimizar os parâmetros da impureza de maneira recursiva:

$$c^* = \min [G(O_l, c)], \quad (2.83)$$

para os subconjuntos $O_l^L(c^*)$ e $O_l^R(c^*)$ até a profundidade máxima permitida, para n_l menor que número mínimo de amostras ou para $n_l = 1$. O critério de regressão, isto é, a métrica utilizada para selecionar uma determinada amostra pertence a um dos subconjuntos esquerdo e direito, também depende da função H . Para o erro quadrático médio e a distribuição de Poisson:

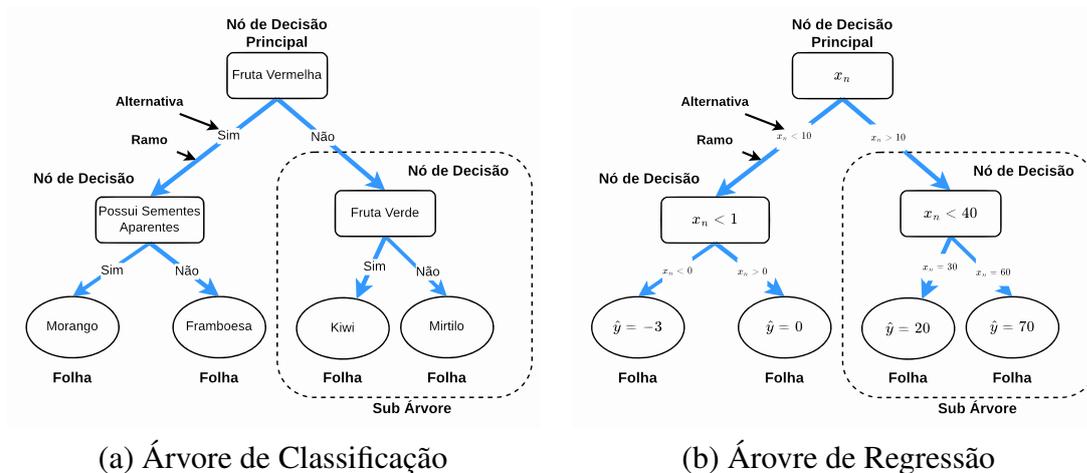
$$\hat{y} = \frac{1}{n_l} \sum_{n_l} y, \text{ e} \quad (2.84)$$

para o erro absoluto médio:

$$\text{mediana}(y_l) = \text{mediana}(y). \quad (2.85)$$

O algoritmo DT possui uma série de vantagens em relação aos algoritmos de aprendizagem supervisionada. O método é de fácil compreensão devido ao seu aspecto visual em árvore, necessitam de pouca manipulação (normalização, substituição de dados faltantes etc.) na base de dados e algumas árvores de decisão funcionam bem até com dados faltantes. Além disso o custo computacional do uso dos modelos de inteligência artificial é proporcional ao logaritmo do número de dados [69]. Por outro lado, as árvores podem perder poder de generalização criando regras complexas demais, deixando os modelos inúteis para informações de fora da base de dados e o modelo gerado pode não ter um comportamento contínuo suave. Além disso, o uso de árvores de decisão pode ser classificado como NP-completo, que é a classe de problemas computacionais mais complexa de se resolver, pois engloba todos os problemas complexos em um processo de decisão [73, 74].

Figura 2.21: Esquemática de funcionamento do algoritmo de árvores de decisão tanto para classificação, figura (a), e regressão, figura (b). A partir do nó principal os dados são separados a partir da confirmação ou rejeição da regra seguindo a equação 2.80.



2.3.3.3 Floresta Randômica

O algoritmo de floresta randômica, ou *Random Forest* (RF), pode ser definido como uma combinação de duas ou mais árvores de decisão para obter melhores modelos de classificação ou regressão. Devido ao fato dessa abordagem utilizar um conjunto de árvore de decisão, todos os conceitos explicados anteriormente para tal algoritmo também são extremamente relevantes nas florestas randômicas. De maneira mais formal, a predição \hat{y} é construída a partir de uma coleção de predições modelos base h_1, h_2, \dots, h_n [72]:

$$\hat{y} = \frac{1}{n} \sum_n h_n. \quad (2.86)$$

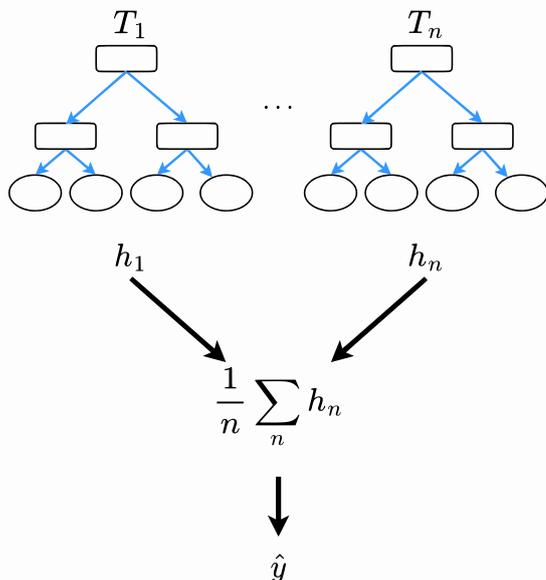
Logo, a n -ésima árvore de decisão é denominada $T(x, J)_n$ em que J é conjunto de variáveis aleatórias e que são independentes de n . Esse processo é ilustrado na figura 2.22.

2.3.3.4 K-Vizinhos Próximos

O algoritmo de k -vizinhos próximos, ou *K-Nearest Neighbors* (KNN) é baseado no cálculo da distância entre as amostras. De forma mais precisa a distância entre um novo ponto e àqueles usados para treinamento é calculada, fazendo a predição de acordo com a mesma, caso a distância entre a nova amostra e os pontos vizinhos seja pequena ele pertence àquele subconjunto, caso contrário ele pode pertencer a outro [69, 66], como ilustrado na figura 2.23-(a).

O número de pontos das variáveis independentes pode ser uma constante k , essa característica é que o nome do algoritmo, ou pode ser baseado na densidade de pontos de um subconjunto local, essa variante é chamada de aprendizagem de raio fixo. A distância d pode

Figura 2.22: Princípio de funcionamento do algoritmo de floresta randômica. A predição \hat{y} é obtida pela combinação das predições intermediárias h_n produzidas pelas T_n árvores de decisão.



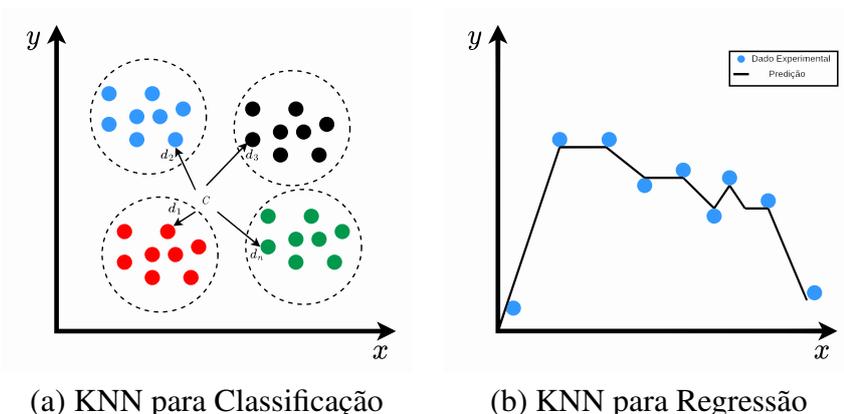
ser calculada a partir de qualquer métrica ℓ_n [1, 69]:

$$d = \|\hat{y} - y\|_n, \quad (2.87)$$

em que geralmente a métrica euclidiana, $n = 2$, é a escolha mais comum. O KNN é uma técnica de inteligência artificial não generalista que simplesmente gera seus modelos a partir de um conjunto fechado. Assim como muitos algoritmos de aprendizagem de máquina o algoritmo KNN pode ser utilizado tanto para classificação e regressão, em que a regressão o resultado final da predição é determinada pela média dos pontos vizinhos, isso é, o ponto é alocado a um determinado subconjunto se, e somente se, o seu valor corresponde à média de tal grupo [69], o comportamento do KNN quando usado para regressão é exemplificado na figura 2.23-(b).

Existem duas maneiras de os pontos vizinhos poderem influenciar o processo de aprendizagem: uniforme ou ponderada. No processo uniforme todos os pontos usados no treinamento possuem pesos idênticos, ou seja, o nível de influência é o mesmo para toda a base de dados. Já na ponderada pode-se atribuir pesos distintos aos pontos alterando a sua influência no aprendizado, por exemplo pode-se aumentar o peso de influência dos pontos mais próximos e negligenciar os mais distantes. Já no que se refere ao cálculo de distância pode-se utilizar três estratégias: a força bruta, a árvore-KD e a bola métrica. A estratégia força bruta calcula todas as distâncias entre pares formados pelo banco de dados, em que para n amostras de dimensão D o tempo de computação é proporcional à $O[nD]^2$ [69]. Tal abordagem é muito eficiente para base de dados de pequena escala, e ficando cada vez mais ineficiente à medida que o número de n aumenta. Para mitigar tal situação, a estratégia árvore-KD foi

Figura 2.23: Esquematização de funcionamento do algoritmo KNN tanto para classificação, figura (a), e regressão, figura (b). Para a classificação, os dados são agrupados em seções semelhantes dado a distância dos pontos ao centro C . Já para a regressão o algoritmo obtém o seu modelo a partir da otimização das distâncias entre os pontos experimentais.



desenvolvida. Tais estruturas foram desenvolvidas para reduzir o número de cálculos das distâncias entre os pontos guardando informações importantes de distância. A ideia por trás da estratégia é intuitiva: considerando três pontos, f , g e h . Caso g for muito distante de f e muito próximo de h não é necessário calcular a distância entre f e h de maneira explícita [69].

Essa maneira de priorização diminui o tempo de treinamento, pois possui custo computacional do tipo $O[Dn \log(n)]$ que é um melhora significativa quando comparada à estratégia força bruta. Esse custo pode ser ainda menor, em torno de $O[\log(n)]$ caso as árvores usadas puderem assumir dimensões arbitrárias. Contudo, a efetividade das árvores-KD decresce de forma vertiginosa para $D > 20$. Tal fenômeno é chamado de maldição de funcionalidade [69]. Nessa situação o uso da bola métrica é a melhor alternativa, visto que a mesma é recomendada para problemas com dimensões $D > 20$, pois diferentemente das árvores-KD que possuem simetria retangular, as bolas métricas são constituídas de hiper esferas. Apesar dessa simetria de solução ser mais computacionalmente custosa do que as árvores-KD, a eficiência pode ser muito superior nos casos de bases de dados bem estruturadas. Essa técnica divide os dados em nós de centro C e raio r em que eles ficam alocados em uma hiper esfera definida por C e r [69]. Dessa maneira a pesquisa por candidatos para pertencerem a tal subconjunto é limitado a pontos que respeitam a desigualdade triangular [17]:

$$|x + y| \leq |x| + |y|, \quad (2.88)$$

permitindo calcular a distância entre o ponto de interesse, o centroide e os limites inferiores e superiores do subconjunto.

2.3.3.5 Máquina de Vetores de Suporte

O algoritmo de máquina de vetores de suporte, ou *support vector machines* (SVM), é outro algoritmo de aprendizagem supervisionada focado na criação de subconjuntos. De maneira geral, o algoritmo SVM busca pelo melhor hiperplano que separa os dados em dois subconjuntos maximizando as suas margens, em que os pontos que estão em tais fronteiras são considerados como vetores de suporte que ajudam a definir os dois subgrupos criados [65, 69, 75]. O hiperplano está no centro definido pelas margens de separação e define o critério de separação. Isso é ilustrado na figura 2.24.

Matematicamente, considerando um conjunto p com x_n elementos preditores de treinamento e y_n e valores alvo:

$$p = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n). \quad (2.89)$$

Para:

$$\begin{aligned} y_n &= 1, \text{ se } x_n \in A, \text{ ou} \\ y_n &= -1, \text{ se } x_n \in B, \end{aligned} \quad (2.90)$$

em que A e B simbolizam os dois subconjuntos em que os dados podem pertencer. Dessa maneira uma função de decisão $D(x)$ é obtida no processo de treinamento, e que uma nova entrada x vai pertencer a A caso $D(x) > 0$ e caso $D(x) < 0$ x pertencerá a B [76]. A partir dessa relação, a função de decisão deve ser uma combinação linear do tipo:

$$D(x) = \sum_{n=1}^N w_n f(x) + b. \quad (2.91)$$

Dessa maneira, w_i e b são parâmetros ajustáveis e $f(x)$ é uma função contínua dependente de x , formando o que é chamado por [76] os classificadores do tipo polinomial. Neste contexto, a função de decisão é definida como:

$$D(x) = \sum_{n=1}^N \alpha_n K(x_n, x) + b, \quad (2.92)$$

em que α_n são coeficientes a serem determinados a partir dos dados de treinamento x_n . A função K é comumente chamada de núcleo, do inglês *kernel*, que podem ser funções potenciais ou funções de base radiais, *Radial Basis Functions*, que em certas condições possuem uma expansão em série na forma:

$$K(x_n, x) = \sum_n f_n(x_n) f_n(x). \quad (2.93)$$

Outra característica é que a função de *kernel* geralmente assume a forma $K(x_n, x) = (x_n x + 1)^q$, ou seja, é uma expansão polinomial de $f(x)$ de ordem q [76]. Satisfeita a equação 2.93, as equações 2.92 e 2.91 representam a mesma função de decisão e dessa maneira

os coeficientes w_n podem ser obtidos da seguinte maneira:

$$w_n = \sum_{n=1}^p \alpha_n f(x_n), \quad (2.94)$$

em que w_n são comumente chamados de parâmetros diretos e α_n parâmetros duais. A função D também é extremamente importante para a determinação das margens. Segundo [76] a distância entre o elemento de entrada x e o hiperplano definido pela função de separação é:

$$d = \frac{D(x)}{|x|}. \quad (2.95)$$

Assumindo que existe uma margem M entre a fronteira da classe, o padrão de treinamento irá satisfazer a seguinte desigualdade:

$$M \geq y_n \frac{D(x_n)}{|w|}. \quad (2.96)$$

Lembrando que no processo de treinamento do SVM o objetivo é maximizar o valor de M , é necessário encontrar w tal que:

$$\begin{aligned} &\text{maximize} && M, \\ &\text{tal que,} && y_n D(x_n) \geq M, \end{aligned} \quad (2.97)$$

para $n = 1, 2, \dots, p$. Aqueles valores de treinamento que estão atados a fronteira M são os já citados vetores de suporte que dão nome à técnica. Dessa maneira, o problema de determinação do hiperplano é um problema de maximização de mínimo:

$$\text{maximize} \quad \text{mínimo} [y_n D(x_n)]. \quad (2.98)$$

Logo, os valores das equações 2.97 e 2.98 são fixados com objetivo de encontrar uma solução dentre conjunto infinito delas. Para evitar o escalonamento causado pela fixação do módulo de w , é possível impor a seguinte restrição:

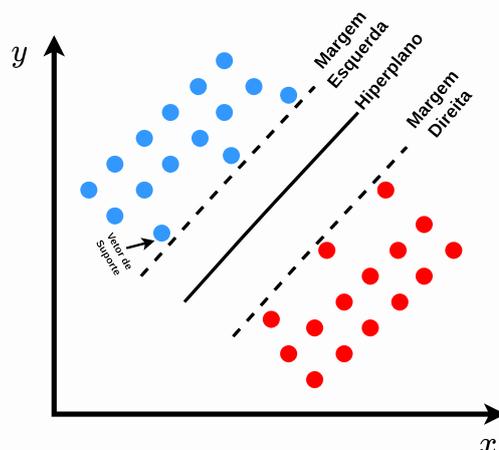
$$M |w| = 1. \quad (2.99)$$

Dessa maneira, maximizar o valor de M necessariamente implica na minimização de $|w|$. Logo, o problema reduz-se a encontrar o valor mínimo de $|w|^2$ nas condições de $y_n D(x_n) \geq 1$ em que a margem máxima é $M^* = 1/|x|$. Esses conceitos são válidos tanto para classificação quanto para regressão. Entretanto, na regressão o problema da maximização torna-se [75, 76]:

$$\begin{aligned} &\text{minimize} && |w|^2, \\ &\text{tal que,} && |y_n \langle w_n, x_n \rangle - b| \geq \epsilon. \end{aligned} \quad (2.100)$$

Nessa equação, $\langle w_n, x_n \rangle$ denota o produto interno entre w_n e x_n .

Figura 2.24: Princípio de funcionamento do algoritmo de SVM. O algoritmo SVM busca pelo melhor hiperplano que separa os dados em dois subconjuntos maximizando as margens, em que os pontos experimentais próximos as margens são chamados de vetores de suporte.



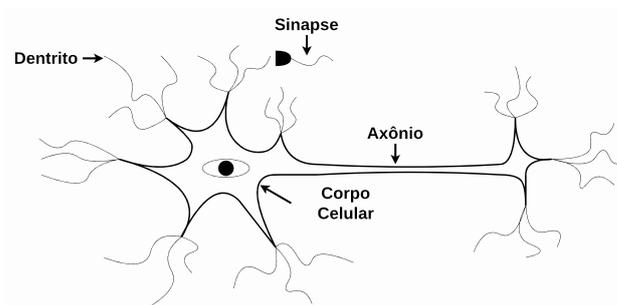
2.3.3.6 Redes Neurais Perceptron

Redes neurais perceptron também conhecida como *Multi-Layer Perceptron* (MLP) é uma classe de algoritmos de aprendizagem de máquina supervisionada que busca uma função f a partir da base de dados tal que $f : \mathbb{R}^m \rightarrow \mathbb{R}^o$ em que m representa a dimensão das entradas, ou variáveis e o a dimensão da saída [69].

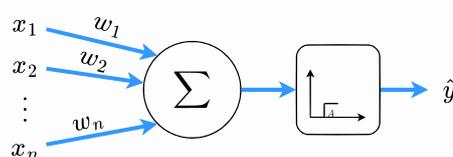
O algoritmo MLP baseia-se em redes neurais biológicas, em que o principal agente é o neurônio [77]. Tal célula é composta por um corpo celular ou *soma* e duas estruturas em estilo de árvore chamadas de axônio e dendrito, como mostrando na figura 2.25-(a). Como toda célula eucarionte, o corpo celular contém toda a informação genética além de também ser responsável pela produção das substâncias necessárias para o funcionamento do neurônio. Os dendritos por outro lado são responsáveis por receber as informações enviadas por outras células nervosas, enquanto o axônio é responsável por transmitir impulsos nervosos para outras células. Essas estruturas podem se ramificar dando origem a fios e subcadeias e em suas terminações existem as sinapses, que são as estruturas que se localizam entre dois neurônios, mas especificamente entre um dendrito da célula transmissora e o axônio da receptora. Quando o impulso nervoso chega até as sinapses, estruturas químicas chamadas de neurotransmissores são liberadas, em que a célula pode ou não emitir sinais elétricos para a célula seguinte dependendo do neurotransmissor recebido [77]. A partir dessa premissa foi criado um modelo matemático chamado de neurônio artificial, que possui como objetivo reproduzir o funcionamento de um neurônio biológico. De maneira geral, o neurônio artificial calcula uma soma ponderada dos n sinais de entradas x_1, x_2, \dots, x_n e gera uma saída y que assume o valor 1 caso o resultado for maior que certo limiar u e 0 na situação contrária, ou seja [67, 77, 69, 68]:

$$y = A \left(\sum_{n=1}^N w_n x_n - u \right), \quad (2.101)$$

Figura 2.25: Ilustração de um neurônio biológico, figura (a), e um artificial, figura (b). O neurônio é a célula nervosa responsável pela transmissão de impulsos elétricos no sistema nervoso, sendo ativado ou não dependendo do estímulo aplicado. Já o neurônio artificial é um modelo matemático que busca reproduzir tal comportamento.



(a) Neurônio Biológico



(b) Neurônio Artificial

em que w_n são os pesos aplicados às entradas x_n , como demonstrado na figura 2.25-(b). O limiar u é também denotado como w_0 , sendo uma entrada constante. Devido ao comportamento binário do neurônio artificial a função A , ou função de ativação, deve ser uma função do tipo limiar [77], em que geralmente as funções mais utilizadas são a sigmoide ou a gaussiana para classificação e para a regressão a função identidade, contudo outras podem ser utilizadas dependendo da simetria do problema [77, 69]. Assim como nos sistemas biológicos os neurônios artificiais são interligados para formar redes, isso é mostrado na figura 2.26. Tais redes artificiais são organizadas da seguinte forma [67, 69, 68]:

- Camada de entrada: são os neurônios que recebem os valores das x_n entradas;
- Camadas ocultas: são as camadas dos neurônios que ficam entre a camada de entrada e a camada de saída;
- Camada de saída: essa camada é a responsável por apresentar os resultados produzidos pela rede.

O objetivo do treinamento de uma rede neural é para um dado conjunto de entrada x encontrar um conjunto de pesos w que reproduza as saídas y , em que x e y são os dados de treinamento. O processo de obtenção dos pesos geralmente é realizado com técnica da descida do gradiente com retropropagação. Essa técnica aplica o gradiente em relação ao parâmetro que precisa ser encontrado, os pesos, a uma função de perda (*loss function*) L , ou seja:

$$w_n^{k+1} = w_n^k - \eta [\alpha \partial_w R(w) + \partial_w L(w)]. \quad (2.102)$$

Nessa equação, k simboliza a iteração atual, η simboliza a taxa de aprendizado, α uma constante e R o termo de regularização [69]. A função L assume formas diferentes dependendo da aplicação do algoritmo MLP. No caso da regressão:

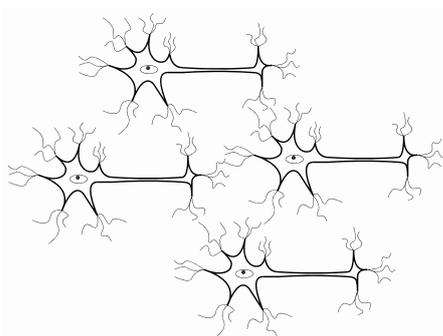
$$L(y, \hat{y}, \mathbf{w}) = \frac{1}{2n} \sum_{n=0}^N \|y - \hat{y}\|_2^2 + \frac{\alpha}{2n} \|\mathbf{w}\|_2^2. \quad (2.103)$$

A distribuição de pesos \mathbf{w} é iniciada com valores aleatórios que são minimizados à medida que o número de iterações k aumenta, calculando a função de perda L e retropagando os resultados da camada de saída para as camadas anteriores, atualizando todos os pesos dos neurônios e minimizando a função de perda [69]. De maneira formal, o gradiente de L é calculado pela distribuição de pesos \mathbf{w} , e dessa maneira, a equação 2.102 torna-se:

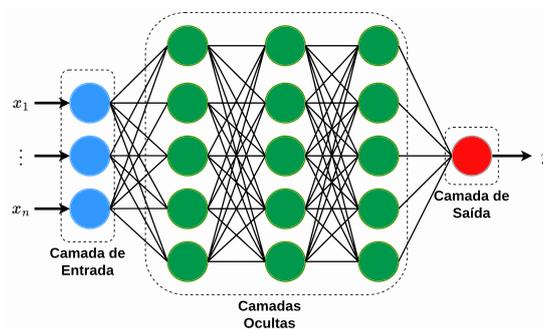
$$\mathbf{w}^{k+1} = \mathbf{w}^k - \eta \nabla_{\mathbf{w}} L, \quad (2.104)$$

para $0 < \eta < 1$. A descida do gradiente para quando o algoritmo atinge o limite máximo de iterações k ou quando a função L assume valores extremamente baixos. O algoritmo MLP possui como vantagens a possibilidade de produzir distribuições não lineares e a capacidade de aprendizado em tempo real [67, 77, 69, 68]. Entretanto, as camadas ocultas podem apresentar funções de perda não convexas e por consequência elas podem apresentar diferentes distribuições de peso \mathbf{w} devido a inicialização das mesmas ser aleatória. Por consequência é necessária uma rotina de adequação por tentativa e erro do número de camadas ocultas para atingir seu potencial máximo [69]. Além disso, a complexidade computacional escala extremamente rápido à medida que a dimensão dos elementos preditores aumenta.

Figura 2.26: Redes neurais. A figura (a) ilustra uma rede formada pelas células nervosas, enquanto a figura (b) a rede artificial. A camada de entrada é a responsável por receber as entradas x_n , enquanto as camadas ocultas são responsáveis pela adaptabilidade da rede aos dados. Finalmente a camada de saída é responsável por entregar a saída \hat{y} , ou seja, a predição.



(a) Rede Neural Biológica



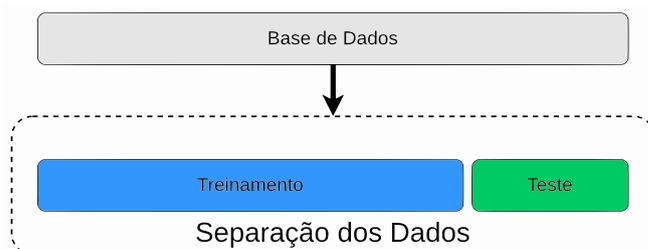
(b) Rede Neural Artificial

2.3.4 Validação Cruzada

O processo de aprendizagem supervisionada, diferente das técnicas UL e RL, necessita de uma sequência de aprendizagem para que os algoritmos consigam aprender e gerar modelos de inteligência artificial para realizar previsões. Para obter tais modelos, a prática mais comum é a separação da base de dados em um conjunto de treinamento e outro de validação, como ilustrado na figura 2.27, geralmente em uma proporção de 70% para treinamento e 30% para teste [69]. Apesar dessa prática ser útil para realizar testes preliminares nos algoritmos de aprendizagem supervisionada, o uso indevido dela pode trazer vários prejuízos no processo de obtenção dos modelos de inteligência artificial.

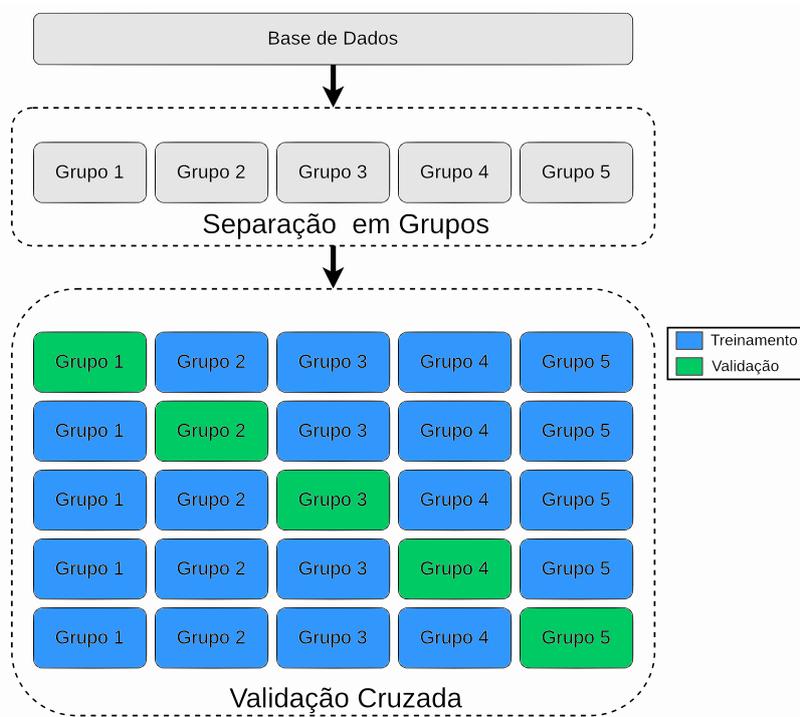
De fato, o uso da mesma sequência para aprendizado e validação pode fazer a inteligência artificial gerada está super adaptada aos dados usados para treiná-la, tornando-a ineficiente para fazer previsões para dados não visto antes, apesar de apresentar resultados ótimos para determinada métrica de desempenho. Tal problema é chamado de *overfitting* [69]. Para evitar o *overfitting* e garantir robustez ao modelo de inteligência artificial gerado, usa-se uma técnica chamada de validação cruzada ou *cross-validation* (CV) [78]. Essa técnica separa a base de dados em um número limitado de grupos, comumente chamados de *folds*, em que o próximo passo é selecionar um desses grupos como o subconjunto de validação enquanto os outros são utilizados para treinamento [78]. Esse processo de treinamento e teste é realizado em um número limitado de iterações, e em cada iteração todos os subconjuntos usados para treinamento são organizados de maneira aleatória para aumentar a variabilidade do processo. Todo esse processo é mostrado na figura 2.28. O resultado de validação intermediário para

Figura 2.27: Procedimento comum para execução do treinamento dos modelos de inteligência artificial. A base de dados é separada uma parte para treinamento e outra para teste, em uma proporção de 70% e 30% respectivamente para a maioria das aplicações.



uma determinada métrica é armazenado e será utilizado para a obtenção do resultado final, que é determinado pela média aritmética de todos os resultados obtidos durante o processo [69, 78]. Em termos gerais, a seleção das métricas de desempenho dependem da aplicação (classificação ou regressão). Para regressão as métricas mais adequadas para a validação são o coeficiente de determinação [79, 80], erro absoluto médio e erro quadrático médio [81] e a acurácia [82], dependendo da necessidade apresentada pela aplicação ou do desejo do operador.

Figura 2.28: Validação cruzada. Essa técnica de validação de modelos de inteligência artificial de aprendizagem supervisionada verifica a robustez e imunidade contra o problema de *overfitting* separando a base de dados em um número limitado de grupos, utilizando um deles para validação e os outros para treinamento. Tal procedimento é executado de maneira iterativa intercalando qual dos grupos criados será utilizado para validação para uma certa quantidade de testes.



2.3.4.1 Coeficiente de Determinação

O coeficiente de determinação também conhecido como métrica R^2 avalia a capacidade de adaptação aos dados de um modelo de inteligência artificial, medindo a proporção de variação da variável dependente a partir das previsões geradas [79, 80, 83]. Considerando um modelo de regressão linear:

$$y_n = w_n x_n + \epsilon, \quad (2.105)$$

para título de prova, em que y_n é a n ésima componente do vetor \mathbf{y} , x_i do vetor \mathbf{x} pertencente à base de dados, em que $n \times p$ em que p simboliza o número de medições experimentais, n a quantidade de variáveis independentes e w_n os coeficientes. A partir do processo de aprendizagem uma distribuição estimada dos pesos \hat{w}_n e com previsão $\hat{y}_n = \hat{w}_n x_n$ são geradas e dessa maneira pode-se definir erro quadrático associado [83]:

$$SE = \sum_{n=1}^N (y_n - \hat{y}_n)^2, \quad (2.106)$$

que determina a variação das previsões \hat{y} em relação aos valores reais y . Por outro lado, caso $n = 1$ e as entradas assumirem o valor $x_n = 1$, as previsões serão basicamente os

coeficientes obtidos pelo processo de aprendizagem, ou seja, $\bar{y}_n = \hat{w}_n$ com erro quadrático:

$$\overline{SE} = \sum_{n=1}^N (y_n - \bar{y}_n)^2, \quad (2.107)$$

computando assim a variação total das respostas. Considerando que em a sua totalidade o modelo gerado possui um termo de interceptação, o coeficiente de determinação pode ser definido como [83]:

$$R^2 = 1 - \frac{SE}{\overline{SE}}. \quad (2.108)$$

A demonstração realizada acima considerou um modelo de regressão linear, contudo nada limita o seu uso em modelos gerados por outros algoritmos. De modo geral, o valor máximo que o coeficiente R^2 pode assumir é 1 (100%), contudo isso é forte indício de *overfitting* [69]. De maneira geral, os melhores modelos geram valores entre 0,8 e 0,96 (80% e 96% respectivamente), demonstrando que a IA consegue descrever com certo grau de generalização o comportamento demonstrado pelos dados de treinamento, conseguindo fazer boas previsões a partir de entradas não vistas pelo modelo [83].

2.3.4.2 Erro Absoluto Médio e Erro Quadrático Médio

As métricas erro absoluto médio, ou *Mean Absolute Error* (MAE), e erro quadrático médio, *Mean Squared Error* (MSE) são responsáveis por determinar a diferença entre pares de resultados distintos obtidos da observação do mesmo fenômeno. Dado um subconjunto de validação com entradas x_n e observações experimentais y_p , o erro absoluto médio pode ser definido como [69, 81]:

$$\text{MAE} = \frac{1}{P} \sum_{p=1}^P |y_p - \hat{y}_p|. \quad (2.109)$$

Em que P é o número total de dados ou observações experimentais e \hat{y}_p é a predição obtida a partir dos dados de entrada x_n . A métrica MAE possui a mesma escala que os dados, e dessa maneira ela mostra uma faixa média em que as predições podem desviar da solução ideal do problema, em que valores muito elevados mostram incapacidade de o modelo adaptar-se ao comportamento mostrado pela base de dados. Já a métrica MSE é definida da seguinte forma [69, 81]:

$$\text{MSE} = \frac{1}{P} \sum_{p=1}^P |y_p - \hat{y}_p|^2. \quad (2.110)$$

Em que os resultados obtidos não estão na mesma dimensão que as predições. Contudo, devido ao fato de o resultado do cálculo do erro estar elevado ao quadrado, tal métrica destaca a tendência de os modelos obterem predições muito discrepantes, visto que a elevação ao quadrado do erro deixa bastante evidente diferenças com valores muito altos.

2.3.4.3 Acurácia

A acurácia é uma métrica com utilização mais comum em aplicações de classificação, mas também pode ser utilizada em aplicações de regressão. Basicamente tal métrica mede o quão próximo as predições \hat{y}_p está do valor real y_p da seguinte maneira [69, 82]:

$$A = [1 \dots 1] \cdot \frac{\mathbf{c}}{\text{card}(\mathbf{y})} \quad (2.111)$$

Nessa equação, \mathbf{c} é o vetor de distribuição de todas as predições y_p corretas, \mathbf{y} representa todos os valores conhecidos de y_p pertencentes ao subconjunto de validação e $\text{card}(\mathbf{y})$ representa o número de elementos que compõem o vetor \mathbf{y} . O valor máximo que A pode assumir é 1 ou 100%, contudo em tal cenário fica evidenciado que o modelo também sofre de *overfitting* [69]. Assim como a métrica R^2 valores considerados aceitáveis giram em torno de 0,8 a 0,96, ou seja, 80% e 96% respectivamente. De fato, a acurácia é muito parecida com a métrica R^2 , a diferença reside no seguinte fato: enquanto a métrica R^2 mensura o qual adaptado o modelo gerado está ao subconjunto usado para treinamento, a métrica acurácia mede o qual próximo da solução convexa do problema o modelo gerado está.

2.3.5 Testes de Hipóteses

A validação cruzada é uma das etapas mais importantes no processo de treinamento de algoritmos de inteligência artificial, pois valida a robustez do modelo gerado frente às variações e dados não antes vistos no processo de treinamento. Após a realização de tal procedimento, o próximo passo são os testes de hipóteses, que nada mais são que análises estatísticas efetuadas para validar ou refutar alguma suposição feita sobre alguma característica ou comportamento esperado da distribuição sob análise. Tais suposições são chamadas de hipótese nula, ou hipótese H_0 , que descreve o comportamento esperado e a hipótese alternativa, denominada como hipótese H_1 , que denota o comportamento contrário ao esperado [84].

No contexto da aprendizagem de máquina supervisionada, os testes de hipóteses são comumente usados para seleção dos modelos de inteligência artificial já que não é incomum que os resultados gerados durante o processo de validação cruzada não sejam estatisticamente diferentes entre dois ou mais algoritmos. Tal prática permite ao operador selecionar as IAs que mais se adequam às suas necessidades, sejam elas economia de recursos computacionais ou até a o uso da combinação dos melhores modelos com o objetivo de predições mais precisas [69]. Nesse cenário, os testes mais realizados são aqueles que envolvem análise de normalidade [85], análise de diferença populacional entre um grupo de distribuições [86] e a análise de diferença populacional entre pares de distribuições [87]. A seguir serão abordados tais testes, com foco para àqueles voltados a distribuições estatísticas normais.

2.3.5.1 Teste de Shapiro-Wilk

O teste de Shapiro-Wilk é um teste de normalidade que visa checar se um determinado conjunto de amostras $\mathbf{x} = x_1, x_2, \dots, x_n$ pertence ou não à uma distribuição normal [85, 88]. Basicamente o teste divide o quadrado de uma combinação linear dos dados amostrais pela estimação simétrica da variância, ou em termos matemáticos [88]:

$$W = \frac{\sum_{n=1}^N (w_n x_{n:N})^2}{\sum_{n=1}^N (x_n - \bar{x})^2}. \quad (2.112)$$

Em que \bar{x} simboliza a média do conjunto das amostras, w_n os coeficientes associados à x_n e $x_{n:N}$ são as estatísticas de ordem da amostra x_n . Os coeficientes w_n são obtidos da seguinte forma:

$$(w_1, w_2, \dots, w_n) = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{-1/2}}. \quad (2.113)$$

Em que \mathbf{m} é o vetor das estatísticas de ordem distribuídas de forma idêntica, definidos como:

$$(m_1, m_2, \dots, m_n) = N \begin{pmatrix} N-1 \\ N-1 \end{pmatrix} \int_0^1 \Phi(X) X^{n-1} (1-X)^{N-1} dx. \quad (2.114)$$

Em que X denota as amostras ordenadas de tamanho N com média zero e variância unitária. Já a matriz \mathbf{V} é a matriz de covariância que pode ser definida como:

$$\mathbf{V} = [m_{nm} - m_n m_m], \quad (2.115)$$

para $m = 1, 2, \dots, N$. A partir dessas definições pode-se estabelecer uma hipótese nula H_0 , em que as amostras avaliadas estão em distribuição gaussiana [88]. Essa hipótese será confirmada ou rejeitada dependendo se o valor de p associado ao valor da probabilidade W para número de graus de liberdade $GL = N - 1$ para um determinado número de amostras N for maior ou menor que um limiar α , também conhecido como confiabilidade do teste. No caso do teste de Shapiro-Wilk, caso o valor p seja maior que α a hipótese nula é confirmada e a distribuição é normal [85]. Geralmente os valores de α giram em torno de 0,01 e 0,05, ou em outras palavras, com graus de confiabilidade de 99% e 95% respectivamente [69]. A realização do teste de Shapiro a partir dos resultados dos testes de validação cruzada é o primeiro passo para obtenção dos melhores modelos de inteligência artificial, pois os testes de diferença populacional mais usados só são aplicados para populações normalmente distribuídas. Contudo, isso não quer dizer que as distribuições não normais não totalmente descartáveis, mas sim que outros testes mais adequados devem ser executados.

2.3.5.2 Teste de Análise de Variância

O teste de análise de variância, comumente conhecido como *One-Way Analyses of Variance* (One-Way ANOVA), possui como objetivo testar três ou mais populações, verificando

se existe uma diferença significativa a partir da comparação da variabilidade entre as distribuições [86]. Dessa maneira a hipótese nula H_0 do teste ANOVA afirma que todas as distribuições com médias $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ são idênticas, em que k denota o número de distribuições a serem comparadas. Para refutar tal hipótese basta que apenas umas das médias \bar{x}_k seja distinta das restantes. Uma característica interessante é que o teste ANOVA só é indicado para distribuições normais [69].

Para obter o resultado do teste ANOVA, uma série de passos devem ser executados. Considerando que x_{nk} é a n -ésima amostra pertencente a uma distribuição normal k e \bar{X} a média de todas as observações de todos os grupos, pode-se determinar a soma dos quadrados:

$$SS = \sum_N (x_{nk} - \bar{X})^2, \quad (2.116)$$

em que N é o número total de amostras. Posteriormente são calculadas outras duas somas quadráticas, a soma quadrada de tratamento e de erro:

$$SST = \sum_{k=1}^K N_k (\bar{x}_k - \bar{X})^2, \text{ e} \quad (2.117)$$

$$SSE = \sum_{k=1}^K (N_k - 1) \sigma_k^2,$$

em que N_k denota o número de amostras na k -ésima população, σ_k o seu desvio padrão e K o número total de populações em análise. As somas SS e SST possuem como objetivo verificar o quão longe as médias populacionais e as amostras em si estão longe de \bar{X} , dando uma ênfase maior naquelas em que a distância seja grande. Já a soma SSE determina a variabilidade de cada distribuição, também dando ênfase em populações com amostras bem distintas entre si. A seguir é definida a soma quadrática total:

$$SS_{Total} = SST + SSE. \quad (2.118)$$

O número total de graus de liberdade da análise é obtido pela soma pelos graus de liberdade de tratamento e erro, ou seja $DG_{Total} = DGT + DGE$, em que:

$$DGT = K - 1, \text{ e} \quad (2.119)$$

$$DGE = N - K.$$

Em seguida são calculadas as médias quadráticas de tratamento:

$$MST = \frac{SST}{DGT}. \quad (2.120)$$

E o erro é:

$$MSE = \frac{SSE}{DGE}. \quad (2.121)$$

Dessa maneira pode-se determinar a estatística F é que a razão entre MST e MSE [86]. Na análise ANOVA, o valor de F associado aos graus de liberdade de tratamento e erro determinam um número p , que geralmente é tabelado. Caso o valor de p da observação seja menor que um limiar α a hipótese nula é confirmada e não existe diferença entre as populações, caso contrário pelo menos uma das populações é diferente das outras. Assim como no teste de Shapiro-Wilk os valores de α variam entre 0,01 e 0,05, ou seja, os testes apresentam graus de confiabilidade entre 99% e 95% [69].

2.3.5.3 Teste de Tukey

Como apresentado anteriormente, o teste ANOVA possui como objetivo evidenciar que as distribuições normais de diferentes populações são diferentes entre si. Contudo ela possui a capacidade de informar que pelo menos uma das populações é diferente das demais, mas não informa quais delas em si apresentam tal comportamento [89]. O teste de Tukey, também conhecido como teste da Honestidade de Diferença Significativa ou *Honestity Significant Difference* (HSD), por outro lado busca evidenciar quais populações são realmente diferentes das demais, executando testes de hipóteses entre pares de distribuições normais [87].

As similaridades entre o teste ANOVAM e Tukey fica mais pronunciada pois a definição da soma quadrática dos erros SSE e o número de graus de liberdade DGE seguem as mesmas fórmulas apresentadas nas equações 2.117, 2.119 e 2.121, com número máximo de grupos $K = 2$. Dessa maneira pode-se determinar a estatística q_T da seguinte maneira [87]:

$$q_T = \frac{\bar{x}_{max} - \bar{x}_{min}}{MSE}, \quad (2.122)$$

em que \bar{x}_{max} é a maior média entre \bar{x}_1 e \bar{x}_2 . Dessa maneira a diferença honesta significativa pode ser calculada da seguinte maneira:

$$HSD = q_T \sqrt{\frac{MSE}{\sigma}}, \quad (2.123)$$

em que σ denota o desvio padrão de todas as amostras das duas populações envolvidas no processo. Assim como os testes de Shapiro e ANOVA, o processo de confirmação ou não da hipótese nula depende se o parâmetro p associado a distribuição q_T supera ou não um limiar α . No teste HSD a hipótese nula é que não existe diferença estatística entre o par de distribuições normais para $p > \alpha$. Normalmente o valor de α é de 0,05 fazendo que o teste HSD tenha um grau de confiabilidade de 95% [69]. Apesar de ser uma boa prática executar o teste de pares de Tukey após o teste ANOVA, não é uma condição obrigatória. Tal equívoco é muito comum no processo de avaliação dos modelos de inteligência artificial obtidos na resolução de um determinado problema.

Capítulo 3

Reconstrução do Campo Eletromagnético para Dados Faltantes e Ruidosos

A reconstrução do campo eletromagnético a partir de dados experimentais é extremamente importante. Dessa maneira, este capítulo apresenta todo o desenvolvimento associado a implementação da técnica de reconstrução do campo eletromagnético utilizando a técnica de Thikonov-Philips, desde como determinar a melhor distribuição para amostragem em um setor circular específico, as grandezas a serem estimadas e como o número de condicionamento é a métrica determinante para todo o processo.

3.1 Expansão em Campo Distante

O campo eletromagnético $E^s = E_\theta^s \hat{\theta} + E_\phi^s \hat{\phi}$ amostrado em um ambiente de caracterização é composto por duas componentes ortogonais à direção de propagação e compõem as duas componentes lineares do campo elétrico E obtido pela aproximação de campo distante, com algum ruído de medição adicional ϵ . Como mostrado em 5.5, tal campo é sobreposição de modos esféricos. Logo, a expressão do campo na aproximação de campo distante r , isolando a dependência azimutal, é escrita em termos de uma série de Fourier do tipo:

$$E = \frac{e^{-jk_0r}}{k_0r} \left(E_\theta^\infty \hat{\theta} + E_\phi^\infty \hat{\phi} \right) \left(1 + \mathcal{O} \left(\frac{1}{k_0r} \right) \right) \quad (k_0r \rightarrow \infty). \quad (3.1)$$

Para:

$$E_{\theta|\phi}^\infty = \sum_{m=-\infty}^{\infty} E_{\theta|\phi}^m e^{jm\phi}, \quad (3.2)$$

$$E_{\theta|\phi}^m = \frac{k_0}{\sqrt{\eta_0}} (-1)^{|j|} \sum_{n=\mu}^{\infty} h_{mn} j^n \sum_{s=1,2} q_{smn}^{(4)} t_{mn}^{s|(3-s)},$$

para $\mu = \max(1, |m|)$, $k_0 = 2\pi/\lambda_0$ e:

$$\eta_0 = \sqrt{\frac{\epsilon_0}{\mu_0}}, \quad (3.3)$$

que é a admitância específica do espaço livre. Os termos h e t são definidos como:

$$h_{mn} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m, \text{ with } \left(-\frac{m}{|m|}\right)^m \Big|_{m=0} \triangleq 1, \text{ e} \quad (3.4)$$

$$t_{mn}^1(\theta) = m \frac{\bar{P}_n^{|m|}(\cos\theta)}{\sin\theta}, t_{mn}^2(\theta) = -\partial_\theta \bar{P}_n^{|m|}(\cos\theta),$$

em que:

$$\bar{P}_n^m(\cos\theta) = \sqrt{\frac{2n+1}{2}} \sqrt{\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta), \quad (3.5)$$

em que $n!$ é o fatorial associado à n -ésima ordem de m -ésimo grau da função de Legendre, definidos como:

$$P_n^m(x) = (1-x^2)^{m/2} \partial_{x^m} P_n(x), \quad (3.6)$$

em que:

$$P_n(x) = \frac{1}{2^n n!} \partial_{x^n} (x^2-1)^n. \quad (3.7)$$

A partir das relações de ortogonalidade das funções associadas de Legendre [90] segue $\forall (n, n') \in \mathbb{N}^2, (n, n') \neq (0, 0)$:

$$\int_{-1}^1 \partial_t P_n^m(t) \partial_t P_{n'}^m(t) \frac{dt}{1-t^2} = \left(\frac{2n(n+1)}{2n+1} - m\right) \frac{(n+m)!}{(n-m)!} \delta_{n,n'}, \text{ e} \quad (3.8)$$

$$\int_{-1}^1 P_n^m(t) P_{n'}^m(t) \frac{dt}{1-t^2} = \frac{1}{m} \frac{(n+m)!}{(n-m)!} \delta_{n,n'},$$

com consequência direta de,

$$\int_0^\pi (t_{mn}^1 t_{mn'}^1 + t_{mn}^2 t_{mn'}^2)_{(\theta)} \sin\theta d\theta = \delta_{n,n'}, \text{ e} \quad (3.9)$$

$$\int_0^\pi (t_{mn}^1 t_{mn'}^2 + t_{mn}^2 t_{mn'}^1)_{(\theta)} \sin\theta d\theta = 0.$$

Quando o campo distante é definido no ângulo sólido completo em 4π esterradiano, as relações de ortogonalidade de Fourier-Legendre se mantêm. logo as expressões de diretividade D e a potência radiada média $\langle P_R^\infty \rangle_t$ através de uma esfera de raio maior que a distância de Fraunhofer são:

$$D = 4\pi \frac{\eta_0}{k_0^2} (|E_\theta^\infty|^2 + |E_\phi^\infty|^2) / \sum_{s,m,n} |q_{smn}^{(4)}|^2, \text{ e} \quad (3.10)$$

$$\langle P_R^\infty \rangle_t = \frac{1}{2} \sum_{s,m,n} |q_{smn}^{(4)}|^2, \quad (3.11)$$

com:

$$\sum_{s,m,n} \triangleq \sum_{s=1,2} \sum_{m=-\infty}^{\infty} \sum_{n=\mu}^{\infty}. \quad (3.12)$$

Na situação na qual o mapa do campo distante não cobre totalmente o ângulo sólido em 4π esterradiano, a versão clássica das expressões apresentadas acima deve ser apli-

cada. Como a série de Fourier-Legendre converge na maioria das vezes [91], seus coeficientes decaem mais rápido do que a sua dependência radial em n . Dessa maneira, como já discutido em 5.5, existe uma ordem máxima N em que a contribuição dos coeficientes esféricos é negligenciável, e dessa maneira os somatórios devem ser modificados para $\sum_{m=-\infty}^{\infty} \rightarrow \sum_{m=-N}^N$ e $\sum_{n=\mu}^{\infty} \rightarrow \sum_{n=\mu}^N$, em que a ordem de truncamento N , é uma função linear do comprimento elétrico $k_0 l$:

$$N = x_s k_0 l, \quad (3.13)$$

em que $2l$ é a maior dimensão da AUT e x_s é uma constante de sobre amostragem maior que a unidade, usualmente $1.00 \leq x_s \leq 1.30$ dependendo do autor. O comprimento elétrico está associado com a quantidade de dígitos significativos na série harmônica esférica, e caso sejam necessários mais dígitos a ordem de truncamento N deve ser aumentada.

3.2 Regularização de Tikhonov-Phillips

Com o objetivo de estimar o campo distante irradiado por qualquer antena pela minimização da potência residual entre o campo medido e o estimado, neste trabalho foi utilizado o esquema de regularização de Tikhonov-Phillips já que o problema foi formulado na forma $\min_x \|Ax - y\|^2$, em que a melhor estimacão é denominada \hat{x} , em que A denota um operador compacto no espaço de Hilbert, em que $\|\cdot\|$ é a norma.

3.2.1 Matriz de Posto Completo

O problema matricial de reconstrução do campo em que o ruído é considerado é resolvido quando se obtém um vetor de soluções x tal que $y = Ax + \epsilon$ em que $(x, y) \in \mathbb{C}^{N_x} \times \mathbb{C}^{N_y}$, $A \in \mathcal{M}_{N_y, N_x}(\mathbb{R})$ e $\epsilon \sim \mathcal{RV}(0_{N_y, 1}, \Omega)$ é uma variável aleatória centrada com valor esperado $\mathbb{E}\epsilon = 0_{N_y, 1}$ e com covariância $\text{cov}(\epsilon) = \Omega \succeq 0$ positiva semidefinida. Aqui, $\mathcal{RV}(\mu, \omega)$ se mantém para uma variável aleatória com valor esperado μ e covariância σ .

Sendo assim, no método de mínimos quadrados usado, $\min_x \|Ax - y\|_V^2$, em que $\|y\|_V = \sqrt{y^H V y}$ é a norma de Mahalanobis ponderada em que V é a matriz real positiva semidefinida ($V \succeq 0$) e diagonal, ou seja, $V^H = V^T = V$, em que o operador compacto A no espaço de Hilbert é uma matriz coluna independente, tornando-o um operador auto adjunto invertível (a pseudo-inversa de Moore-Penrose). Logo a solução ótima é $\hat{x} \triangleq (A^H V A)^{-1} A^H V y$, é também conhecido como estimador do método dos mínimos quadrados. Consequentemente, $\hat{x} - x = (A^H V A)^{-1} A^H V \epsilon$ é uma estimacão não enviesada de x , em que a sua matriz de covariância é $\text{cov}(\hat{x} - x) = (A^H V A)^{-1} A^H V \Omega V A (A^H V A)^{-1} = \Gamma = \Gamma^H \succeq 0$.

Dessa maneira fica definido o estimador $\hat{y} \triangleq A - \hat{x} = P y$ em que $P = A (A^H V A)^{-1} A^H V \in \mathcal{M}_{N_y, N_y}(\mathbb{R})$ que é a projeção de em A , em $P^2 = P$ e $PA = A$.

Já o seu operador ortogonal é a projeção do elemento nulo $M \triangleq I - P$ já que $M^2 = M$ e $MA = 0_{N_y}$, em que I é o operador identidade. Então, o estimador do erro residual é $\hat{\epsilon} \triangleq y - \hat{y} = M\epsilon$, sendo uma variável aleatória. Consequentemente, \hat{y} e $\hat{\epsilon}$ são também estimadores não viesados, com covariâncias $\text{cov}(\hat{y} - y) = \text{cov}(\hat{\epsilon}) = \mathbb{E}(\hat{\epsilon} \otimes \hat{\epsilon}^*) = M\Omega M^H$, em que $\mathbb{E}(A)$ representa a estatística esperada de A , e que $A^* = (A^H)^T$ é o seu complexo conjugado e $A \otimes B$ o produto tensorial entre A e B .

3.2.2 Mapa Angular Parcial: Matriz de Posto Incompleto

Em muitos cenários de caracterização experimental de antenas não é sempre um requisito amostrar e reconstruir o campo em toda esfera unitária. De fato, em alguns casos é necessário focar em um setor angular específico, por exemplo onde o lóbulo principal está localizado. Primeiramente, tal cenário leva a um número menor de amostras, e por esse motivo o sistema linear se torna uma matriz de posto incompleto, necessitando de informações *a-priori* para poder ser invertida. Uma maneira de fazer isso é aumentar a densidade de amostras dentro do setor circular com o objetivo de alcançar uma quantidade mínima necessária para obter uma matriz de posto unitário, em que o setor circular de interesse é delimitado por uma janela:

$$\theta_{min} \leq \theta \leq \theta_{max}, \text{ e } \phi_{min} \leq \phi \leq \phi_{max}. \quad (3.14)$$

Contudo, em situações em que A não é uma matriz independente devido ao número menor de amostras, o sistema anterior deve ser modificado para adicionar alguma informação *a-priori* conhecida. Dessa maneira a regularização de Tikhonov-Phillips consideram tal informação em sua solução da seguinte maneira:

$$x_\alpha = \min_{x \in \mathbb{C}^{N_x}} \left(\|Ax - y\|_V^2 + \alpha \|x - x_0\|_Q^2 \right), \quad (3.15)$$

em que V e Q são duas matrizes positivas diagonais, e $\alpha > 0$ e $\|x_0\|^2$ é a potência desejada. Neste trabalho é inserida uma nova abordagem do da regularização de Tikhonov-Phillips para matrizes de posto incompleto. Considerando a notação direta dos somatórios $A = A \oplus I_{N_x}$ para o operador compacto, $V_\alpha = V \oplus \alpha Q \succeq 0$, $x = x \oplus x_0$ e $y = y \oplus x_0$ o sistema da equação 3.15 volta a ser um problema de posto unitário mostrado em 3.2.1, ou seja, $x_\alpha = \min_x \|Ax - y\|_{V_\alpha}^2 = (A^H V A + \alpha Q)^{-1} (A^H V y + \alpha Q x_0)$.

3.3 Amostragem do Campo

3.3.1 Amostragem em ϕ

Para uma dada posição $\theta \in [0, \pi]$, o campo elétrico é amostrado em M posições angulares $\{\phi_m \in [0, 2\pi] | m = 1, 2, \dots, M\}$. Isso produz uma coletânea de amostras de campo para cada

polarização $\left\{ (E_\theta^s, E_\phi^s)_{(\theta, \phi_m)} \in \mathbb{C}^2 \mid m = 1, 2, \dots, M \right\}$. Para simplificar o processamento, as amostras são agrupadas da seguinte maneira:

$$\forall \theta \in [0, \pi], [E_\alpha^s]_{(\theta)} = \frac{e^{-jk_0 r}}{k_0 r} [E_\alpha^\infty]_{(\theta)} + [\varepsilon_{\phi\alpha}]_{(\theta)}, \quad (3.16)$$

em que:

$$[E_\alpha^\infty]_{(\theta)} = \Phi \cdot [E_\alpha^F]_{(\theta)}, \quad (3.17)$$

para:

$$\alpha \in \{\theta, \phi\},$$

$$[\varepsilon_\phi] = \{[\varepsilon_{\phi\theta}] \mid [\varepsilon_{\phi\phi}]\}, [\varepsilon_{\phi\alpha}] = \left[[\varepsilon_{\phi\alpha}]_{(\theta_1)} \mid \dots \mid [\varepsilon_{\phi\alpha}]_{(\theta_L)} \right],$$

$$\forall m = 1 \dots M, n = 1 \dots 2N + 1, (\Phi)_{m,n} = e^{j(n-N-1)\phi_m}, \quad (3.18)$$

$$[E_\alpha^s]_{(\theta)} = \begin{bmatrix} E_\alpha^s(\theta, \phi_1) \\ \vdots \\ E_\alpha^s(\theta, \phi_M) \end{bmatrix}, [\varepsilon_{\phi\alpha}]_{(\theta)} = \begin{bmatrix} \varepsilon_{\phi\alpha}(\theta, \phi_1) \\ \vdots \\ \varepsilon_{\phi\alpha}(\theta, \phi_M) \end{bmatrix},$$

resultando em:

$$\begin{aligned} [E_\alpha^F]_{(\theta)} &= \left[E_\alpha^{-N}(\theta) \cdots E_\alpha^N(\theta) \right]^T, \\ [E^F] &= \left[[E_\theta^F]_{(\theta_1)} \cdots [E_\theta^F]_{(\theta_L)} \mid [E_\phi^F]_{(\theta_1)} \cdots [E_\phi^F]_{(\theta_L)} \right], \\ [E^s] &= \left[[E_\theta^s]_{(\theta_1)} \cdots [E_\theta^s]_{(\theta_L)} \mid [E_\phi^s]_{(\theta_1)} \cdots [E_\phi^s]_{(\theta_L)} \right]. \end{aligned} \quad (3.19)$$

Nesse sentido, o campo obtido deve ser projetado nos modos esféricos para ser reconstruído. Em outras palavras, o máximo de potência deve ser transferida do campo medido para a SWE. Tal processo resulta na minimização da variância do erro residual, ou em outras palavras, resulta no uso da regularização de Tikhonov-Phillips explicado em 3.2.1. Deve se destacar o ponto em que a estimação pode ser obtida se, e somente, não houver uma multicolinearidade entre as variáveis de Φ que garantem que ela é uma matriz de posto completo, o que implica $M \geq 2N + 1$. Logo a correspondência entre a situação em análise e um sistema linear com ruído associado é:

$$A \leftrightarrow \Phi, X \leftrightarrow \frac{e^{-jk_0 r}}{k_0 r} [E^F]_{(\theta)}, Y \leftrightarrow [E^s]_{(\theta)}, \varepsilon \leftrightarrow [\varepsilon_\phi]_{(\theta)}, V \leftrightarrow V_\Phi. \quad (3.20)$$

O conjunto $\{\phi_m \in]0, 2\pi[\mid m = 1, 2 \dots M\}$ deve ser definido. Para determinar a melhor distribuição, o número de condicionamento c_Φ da matriz de Gram do tipo $\Phi^H V_\Phi \Phi$ é investigado de tal maneira que tal matriz deve o mais próximo possível de I para garantir o melhor con-

dicionamento possível do problema. Por definição, $c_\Phi = \|\Phi^H V_\Phi \Phi\|_2 \cdot \|(\Phi^H V_\Phi \Phi)^{-1}\|_2 = \frac{\sigma_{\max}(\Phi^H V_\Phi \Phi)}{\sigma_{\min}(\Phi^H V_\Phi \Phi)}$, em que $\|A\|_2$, $\sigma_{\max}(A)$ e $\sigma_{\min}(A)$, para a norma L_2 da matriz A e os valores máximos e mínimos singulares. Uma condição suficiente para a matriz de Gram deve ser no mínimo diagonal e o mais próximo possível da matriz identidade. Com $V_\Phi \in \mathcal{M}_{M \times M}(\mathbb{R})$.

Os elementos da matriz $\Phi\Phi^H$, $\forall p = 1, 2 \dots M, q = 1, 2 \dots M$, são $(\Phi\Phi^H)_{p,q} = \sum_{n=1}^{2N+1} e^{j(n-N-1)\phi_p} e^{-j(n-N-1)\phi_q} \in \mathbb{C}$, podem ser escritos como $(\Phi\Phi^H)_{p,q} = \frac{\text{sinc}((\phi_p - \phi_q)(2N+1)/2)}{\text{sinc}((\phi_p - \phi_q)/2)}(2N+1)$, onde $\text{sinc } u = \begin{cases} 1, & \text{if } u = 0 \\ \frac{\sin u}{u}, & \forall u \in \mathbb{R}^* \end{cases}$ sendo a função cardinal.

Para garantir que todos os elementos serão descartados, exceto os diagonais, é suficiente fixar $\phi_p - \phi_q = \frac{2m\pi}{2N+1}$, $m \in \mathbb{N}^*$ para $p \neq q$. Garantido que $M \geq 2N+1$ e que não exista um denominador nulo, e fixando $\{\phi_m = m\frac{2\pi}{2N+1} | m = 1, 2 \dots M\}$. fornece $\Phi\Phi^H = MI_M$. Os elementos da matriz $\Phi^H\Phi, \forall p = 1, 2 \dots 2N+1, q = 1, 2 \dots 2N+1$, são $(\Phi^H\Phi)_{p,q} = \sum_{m=1}^M e^{j(p-N-1)\phi_m} e^{-j(q-N-1)\phi_m} \in \mathbb{C}$, esses elementos podem ser escritos como $(\Phi^H\Phi)_{p,q} = \sum_{m=1}^M e^{j(q-p)\phi_m}$ usando a distribuição apresentada anteriormente $\Phi^H\Phi = MI_{2N+1}$.

Dessa maneira, escolhendo $V_\Phi = \frac{1}{M}I_M$, a expressão para a matriz de Gram é $\Phi^H V_\Phi \Phi = I_{2N+1}$, com número de condicionamento $c_\Phi = 1$. Esse resultado não é apenas um problema bem condicionado, mas é estatisticamente independente dos coeficientes resultantes quando a sua matriz de covariância é diagonal. Como consequência direta, arranjando as amostras em matrizes, os estimadores tornam-se:

$$\left[\hat{E}^F \right] = k_0 r e^{jk_0 r} (\Phi^H V_\Phi \Phi)^{-1} \Phi^H V_\Phi [E^s], \text{ and} \quad (3.21)$$

$$\left[\hat{E}^s \right] = P_\Phi [E^s], [\hat{\varepsilon}_\phi] = M_\Phi [\varepsilon_\phi] = M_\Phi \left([E^s] - \left[\hat{E}^s \right] \right),$$

em que:

$$\left[\hat{E}^F \right] = \left[\left[\hat{E}_\theta^F \right]_{(\theta_1)} \cdots \left[\hat{E}_\theta^F \right]_{(\theta_L)} \left[\hat{E}_\phi^F \right]_{(\theta_1)} \cdots \left[\hat{E}_\phi^F \right]_{(\theta_L)} \right], \quad (3.22)$$

$$\left[\hat{E}^s \right] = \left[\left[\hat{E}_\theta^s \right]_{(\theta_1)} \cdots \left[\hat{E}_\theta^s \right]_{(\theta_L)} \left[\hat{E}_\phi^s \right]_{(\theta_1)} \cdots \left[\hat{E}_\phi^s \right]_{(\theta_L)} \right],$$

e:

$$\left[\hat{\varepsilon}_\phi \right] = \left[\left[\hat{\varepsilon}_{\phi\theta} \right]_{(\theta_1)} \cdots \left[\hat{\varepsilon}_{\phi\theta} \right]_{(\theta_L)} \left[\hat{\varepsilon}_{\phi\phi} \right]_{(\theta_1)} \cdots \left[\hat{\varepsilon}_{\phi\phi} \right]_{(\theta_L)} \right], \quad (3.23)$$

$$\left[\hat{\varepsilon}_\phi \right] = \left[\left[\hat{\varepsilon}_{\phi\theta} \right]_{(\theta_1)} \mid \cdots \mid \left[\hat{\varepsilon}_{\phi\theta} \right]_{(\theta_L)} \mid \left[\hat{\varepsilon}_{\phi\phi} \right]_{(\theta_1)} \mid \cdots \mid \left[\hat{\varepsilon}_{\phi\phi} \right]_{(\theta_L)} \right],$$

com:

$$\Gamma_\phi = \Phi^H V_\phi \Omega_\phi V_\phi \Phi, \text{ and } \Omega_\phi = \mathbb{E}([\varepsilon_\phi] \otimes [\varepsilon_\phi]^*). \quad (3.24)$$

3.3.2 Amostragem em θ

Já em no plano θ , o campo eletromagnético é amostrado em $\{\theta_l | l = 1, 2 \dots L\}$. A análise da série de Fourier feita anteriormente é o começo da de Legendre. Logo, a expressão compacta das relações de campo é:

$$\forall m = -N \dots N, \quad (3.25)$$

em que:

$$F^m = \frac{k_0}{\sqrt{\eta_0}} \Theta_m \cdot J_m \cdot Q_m, \quad (3.26)$$

para:

$$\Theta_m \triangleq \left[\begin{array}{c|c} -T_m^1 & -T_m^2 \\ \hline j \cdot T_m^2 & j \cdot T_m^1 \end{array} \right] \quad (3.27)$$

$$\forall i \in \{1, 2\}, \forall l = 1 \dots L, \forall n = \mu \dots N, (T_m^i)_{l,n} \triangleq t_{mn}^i(\theta_l),$$

$$Q_{sm} \triangleq \begin{bmatrix} q_{sm\mu}^{(4)} \\ \vdots \\ q_{smN}^{(4)} \end{bmatrix}, Q_m \triangleq \begin{bmatrix} Q_{1m} \\ Q_{2m} \end{bmatrix}, \forall s \in \{1, 2\},$$

$$\epsilon_{\theta\alpha}^m \triangleq \left[\epsilon_{\theta\alpha(\theta_1)}^m, \dots, \epsilon_{\theta\alpha(\theta_L)}^m \right]^T, \epsilon_{\theta}^m \triangleq \begin{bmatrix} \epsilon_{\theta\theta}^m \\ \epsilon_{\theta\phi}^m \end{bmatrix}, \forall \alpha \in \{\theta, \phi\} \quad (3.28)$$

$$j_m \triangleq \text{diag} (h_{\mu,m} \cdot j^\mu, \dots, h_{N,m} \cdot j^N),$$

$$J_m \triangleq \left[\begin{array}{c|c} j_m & 0_{N+1-\mu} \\ \hline 0_{N+1-\mu} & j_m \end{array} \right],$$

para $[F] = [F^{-N} | \dots | F^N] = [\hat{E}^F]^T$. Aqui ϵ_{θ}^m é a variável aleatória centrada representando a transformada de Fourier discreta do ruído medido. A variância da transformada de Fourier do ruído residual medido é minimizada, resultando novamente na aplicação da técnica de regularização de Tikhonov-Phillips mostrada em 3.2. Para evitar a existência de multicolinearidade, o que significa Θ_m ser uma matriz de posto completo, é necessário que $L \geq N$, para $m = -N \dots N$. A correspondência logo assume a forma:

$$A \leftrightarrow \Theta_m, x \leftrightarrow \frac{k_0}{\sqrt{\eta_0}} J_m \cdot Q_m, y \leftrightarrow F^m, \epsilon \leftrightarrow \epsilon_{\theta}^m, V \leftrightarrow W_m \quad (3.29)$$

Seguindo o mesmo raciocínio, para determinar a melhor distribuição, o número de condicionamento c_{Θ} da matriz $\Theta_m^H W_m \Theta_m$, em que W_m é a matriz diagonal de pesos positiva semidefinida $W_m = \begin{pmatrix} W_L & 0_L \\ 0_L & W_L \end{pmatrix}$, $W_L = \text{diag} (v_1, \dots, v_L) \in \mathcal{M}_L(\mathbb{R})$. Fazendo a análise dessa matriz de bloco e aplicando a mesmo raciocínio feito no setor angular ϕ , a matriz

em bloco deve ser o mais próxima possível de uma matriz diagonal e dessa maneira possibilitar o melhor número de condicionamento possível. Um condição suficiente é que a matriz $\Theta_m^H W_m \Theta_m$ deve ser ao mesmo diagonal e ser o mais próxima possível da matriz identidade. Tais resultados seguem para a matriz de blocos $\Theta_m^H W_m \Theta_m, \forall m = -N, \dots, N$ em que:

$$\left(\begin{array}{c|c} [\Theta_m^H W_m \Theta_m]_{11} & [\Theta_m^H W_m \Theta_m]_{21} \\ \hline [\Theta_m^H W_m \Theta_m]_{21} & [\Theta_m^H W_m \Theta_m]_{11} \end{array} \right), \quad (3.30)$$

em que:

$$\begin{aligned} [\Theta_m^H W_m \Theta_m]_{11} &= T_m^1{}^H W_L T_m^1 + T_m^2{}^H W_L T_m^2, \\ [\Theta_m^H W_m \Theta_m]_{21} &= T_m^1{}^H W_L T_m^2 + T_m^2{}^H W_L T_m^1, \end{aligned} \quad (3.31)$$

com, $\forall (i, j) \in \{1, 2\}^2, \forall p, q = 1 \dots N$,

$$\left(T_m^i{}^H W_L T_m^j \right)_{p,q} = \sum_{l=1}^L t_{mp}^i(\theta_l) \cdot v_l \cdot t_{mq}^j(\theta_l).$$

Para conseguir a expressão para os pesos $\{\theta_l | l = 1, 2 \dots L\}$ da somatória os produtos internos devem ser analisados. De acordo com a relação 18.11.1 de [91], $\forall (m, n) \in \mathbb{N}^2, 0 \leq m \leq n, \forall x \in [-1, 1], P_n^m(x) = \left(\frac{1}{2}\right)_m (-2)^m (1-x^2)^{\frac{m}{2}} \mathcal{C}_{n-m}^{(m+\frac{1}{2})}(x)$ em que $(a)_m$ é o símbolo de Pochhammer, $\{\mathcal{C}_n^{(\alpha)}\}_{n \in \mathbb{N}}$ representa a ortogonal com respeito a $(1-x^2)^{\alpha-1/2}, x \in [-1, 1]$ família dos polinômios esféricos reais de Gegenbauer, também conhecido como polinômios Ultra Esféricos. Dessa maneira, $\mathcal{C}_n^{(\alpha)} \in \mathcal{P}_n(\mathbb{R})$ é um polinômio real de grau n , resultando em:

$$\begin{aligned} \forall (m, n, n', i, j, \theta) &\in \mathbb{N}^3 \times \{1, 2\}^2 \times [0, \pi], \\ t_{mn}^i(\theta) t_{mn'}^j(\theta) &= w(x) P_{i+j}(x) \Big|_{x=\cos \theta}, P_{i+j} \in \mathcal{P}_{m'}(\mathbb{R}), \\ w(x) &= (1-x^2)^{\mu-1}, m' = n + n' - 2m + i + j - 2. \end{aligned} \quad (3.32)$$

O método de integração em quadratura de Gauss é agora aplicado em relação aos polinômios associados aos pesos w . Logo, $\forall i = 2, 3, 4, \int_{-1}^1 w(x) P_i(x) dx = e_L + \sum_{l=1}^L w_l P_i(x_l)$ em que o erro é $e_L = \frac{P_i^{(2L)}(\xi)}{(2L)!} \|\mathcal{C}_L^{(\mu-\frac{1}{2})}\|^2$, onde o erro assume a forma $e_L = \frac{P_i^{(2L)}(\xi)}{(2L)!} \|\mathcal{C}_L^{(\mu-\frac{1}{2})}\|^2$, com $\xi \in]-1, 1[, \forall l = 1 \dots L, w_l = 1 / \sum_{k=0}^L \left(\mathcal{C}_k^{(\mu-\frac{1}{2})}(x_l) / \|\mathcal{C}_k^{(\mu-\frac{1}{2})}\| \right)^2$, conhecido como número de Christoffel, $\{-1 < x_1 < \dots < x_L < 1 | \forall l = 1 \dots L, \mathcal{C}_L^{(\mu-\frac{1}{2})}(x_l) = 0\}$ o L zeros ordenados do polinômio ortogonal $\mathcal{C}_L^{(\mu-\frac{1}{2})} \in \mathcal{P}_L(\mathbb{R})$. Escolhendo o grau L do polinômio de Gegenbauer para que $2L$ é maior que o maior grau dos polinômios P_1, P_2 e P_3 , o erro e_L desaparece e a igualdade entre a soma discreta e a contínua é exata. Uma condição suficiente é $L > N$. Então, escolhendo os pesos $\forall l = 1, 2 \dots L, v_l = \frac{w_l}{w(\cos \theta_l)}$, resultando em $\forall m = -N \dots N, \forall (i, j) \in \{1, 2\}^2, \forall p, q = \mu \dots N, (T_m^i W_L T_m^j)_{p,q} = \int_0^\pi t_{mn}^i(\theta) t_{mn'}^j(\theta) \sin \theta d\theta$. Em outras palavras, $\forall m = -N \dots N, \Theta_m^H W_m \Theta_m = I_{2(N-\mu+1)}$.

Mais uma vez tais resultados não são só somente bem condicionados, como também estatisticamente independentes dos coeficientes recuperados e a sua matriz associada é diagonal. Como consequência, os números de condicionamento são $c_{\Theta}(m, N) = 1$. Consequentemente, os coeficientes modais podem agora ser estimados a partir das equações de campo $\forall m = -N \cdots N$, os estimadores são:

$$\hat{Q}_m = \frac{\sqrt{\eta_0}}{k_0} D_m \Theta_m^H W_m F^m, \quad e \hat{F}^m = \frac{k_0}{\sqrt{\eta_0}} \Theta_m J_m \hat{Q}_m, \quad (3.33)$$

e $\hat{\epsilon}_{\theta}^m = F^m - \hat{F}^m$, onde:

$$D_m = J_m^{-1} (\Theta_m^H W_m \Theta_m)^{-1} = \left[\begin{array}{c|c} j_m^{-1} & 0_{N-\mu,1} \\ \hline 0_{N-\mu,1} & j_m^{-1} \end{array} \right], \quad (3.34)$$

na condição de:

$$\forall (p, q) = \mu \cdots N, (j_m^{-1})_{p,q} = h_{p,m}^{-1} j^{-p} \delta_{pq}, \quad (3.35)$$

$$\Omega_{\theta,m} = \mathbb{E} (\epsilon_{\theta}^m \otimes \epsilon_{\theta}^{m*}),$$

o campo reconstruído nas posições angulares amostradas é finalmente dado por:

$$\left[\hat{E}^R \right] = \frac{e^{-jk_0 r}}{k_0 r} \Phi \cdot \left(\hat{F}^{-N} \mid \dots \mid \hat{F}^N \right)^T. \quad (3.36)$$

De acordo com a relação 18.16.8 de [90], os zeros $\mathcal{C}_L^{(\alpha)}$, podem ser aproximados para $\forall l = 1 \cdots L$, $x_{L,l} = \cos \theta_{L,l}$, para $0 < \theta_{L,1} < \cdots < \theta_{L,L} < \pi$, com $\theta_{L,l} < \theta_{L+1,l} < \theta_{L+2,l}$, $\theta_{L,l} = \phi_l + \left((\alpha^2 - \frac{1}{4}) \frac{1 - \phi_l \cot \phi_l}{2\phi_l} \right) \frac{1}{\rho^2} + \phi_l^2 O\left(\frac{1}{\rho^3}\right)$, $\Phi_{L,l} = \frac{j_{L,l}}{\rho}$, $\rho = L + \frac{1}{2}(2\alpha + 1)$, $j_{\alpha,l}$ sendo o l -ésimo zero positivo da função de Bessel $J_{\alpha}(x)$. Essa distribuição ideal chamada de D_0 é custosa de ser executada experimentalmente e impacta diretamente no número de condicionamento. Dessa maneira ela deve ser aproximada por outras famílias de funções mais convenientes. De fato, elas possuem a mesma distribuição para todas as ordens modais m , com custo de algum erro no resultado obtido. Essas distribuições alternativas são, $\forall l = 1, 2 \cdots L$:

$$\begin{array}{l} D_1 : \theta_{L,l} = \frac{l\pi}{L+1} \\ D_2 : \theta_{L,l} = \frac{(2l-1)\pi}{2L} \\ D_3 : \theta_{L,l} = \frac{(l-1)\pi}{L-1} \end{array} \left| \begin{array}{l} D_4 : x_{L,l} = \frac{l}{L+1} \\ D_5 : x_{L,l} = \frac{2l-1}{2L} \\ D_6 : x_{L,l} = \frac{l-1}{L-1} \end{array} \right.,$$

$$D_7 : (x_{L,l})_{l=1,2,\dots,L} \text{ zeros ordenados de } \mathcal{C}_L^{(\mu-\frac{1}{2})}, \mu = 1$$

$$D_8 : (x_{L,l})_{l=1,2,\dots,L} \text{ zeros ordenados de } \mathcal{C}_L^{(\mu-\frac{1}{2})}, \mu = 2$$

$D_9 : (x_{L,l})_{l=1,2,\dots,L}$ zeros ordenados de $C_L^{(\mu-\frac{1}{2})}$, $\mu = L$

$D_{10} : (x_{L,l})_{l=1,2,\dots,L}$ zeros ordenados de $C_L^{(\mu-\frac{1}{2})}$, $\mu = \lfloor \frac{L}{2} \rfloor$ (3.37)

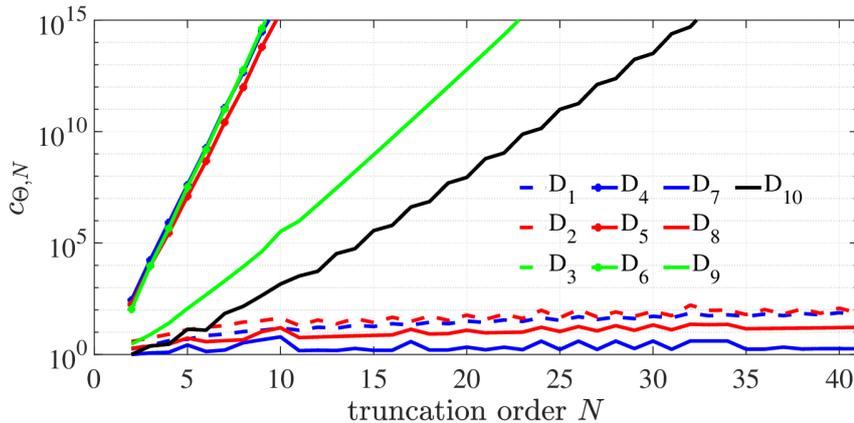
$[x]$ denotando o inteiro parte de $x \in \mathbb{R}$.

A distribuição $D_{1,2,3}$ são equiangulares e extremamente densas nos polos. Para contornar isso, as distribuições $D_{4,5,6}$ são distribuídas de maneira igual em no eixo (O, \hat{z}) . Já $D_{7,8,9,10}$ são os polinômios não nulos com diferentes argumentos $\mu - \frac{1}{2}$ e aplicados para todos os modos. A métrica para selecionar a melhor distribuição é o menor número de condicionamento máximo, definido como:

$$\forall N \in \mathbb{N}^*, c_{\Theta,N} = \max_{0 \leq m \leq N} c_{\Theta}(m, N), \quad (3.38)$$

que garante melhores inversões de sistemas lineares. Existem algoritmos que possuem a capacidade de inverter matrizes mal condicionadas em que o número de condicionamento é da ordem de 10^{10} , mas o resultado é bastante sensível a menor variação de erro. Números de condicionamento de até 100 são preferíveis para inverter sistemas de maneira mais segura. De acordo com a figura 3.1 a distribuição D_7 é a mais adequada pois produz número de condicionamento $c_{\Theta,N}$ máximo de 6,15. Já a distribuição D_8 , com número de condicionamento máximo de 22,62, também é adequada em θ , mas não é tão boa quando comparada a D_7 . A próxima distribuição que pode ser usada é a D_1 em que o número de condicionamento máximo é de 74,75. Outro ponto importante é o comportamento assintótico das distribuições: todas elas divergem, exceto D_7 que converge quando N aumenta tendendo para unidade.

Figura 3.1: Número de condicionamento $c_{\Theta,N}$ para o ângulo θ . Essa figura mostra que D_7 é a melhor distribuição com valor máximo de 6,15 e convergindo para 1,00 a medida que a ordem de truncamento N aumenta.



Dessa maneira, a escolha de D_7 resulta em matrizes não mais diagonais. Contudo, como as funções são todas contínuas na vizinhança dos ângulos originais da distribuição D_0 , as matrizes resultantes com a distribuição aproximada D_7 ainda são principalmente diagonais, e podem, então serem facilmente invertidas de maneira estável a partir do método de elimi-

nação de Gauss. Isso é confirmado pelo número de condicionamento c_Θ que converge para 1 a medida que a ordem de truncamento aumenta. Estatisticamente, as colunas da matriz das amostras não são mais independentes, mas com uma correlação extremamente fraca.

Outra coisa importante a ser ressaltada desse resultado: a escolha θ é extremamente importante para executar a expansão modal esférica. Dessa maneira, a distribuição aproximada do ângulo θ deve ser realizada somente para as amostras em posições angulares. As funções ponderadas w_l , são críticas na sensibilidade dos regularizadores, e devem ser avaliadas nos zeros exatos nos polinômios de Gegenbauer. Caso não, os números de condicionamento $c_{\Theta,N}$ começa a divergir exponencialmente quando N chega a 45.

3.3.3 Número total de Amostras

Para a técnica reconstrução proposta, a quantidade de amostras é $N_s = M.L$, em que $M \geq 2N + 1, L > N$. Consequentemente no caso limite $N_s = (N + 1)(2N + 1) = (x_s k_0 l + 1)(2x_s k_0 l + 1)$. Em [4] é proposto um esquema que o número de amostras é $N_H = (N+1)(2N+1)$. Em outras palavras no mínimo $N_s = N_H$. Entretanto, [4] necessita estender de maneira artificial de $[0, \pi]$ para $[0, 2\pi]$ a dependência em θ , espelhando as amostras com o objetivo de aplicar a transformada rápida de Fourier, *Fast Fourier Transform* (FFT), no intervalo artificial periódico em 2π no eixo θ e assim recuperar os coeficientes em situações em que o setor angular de amostragem é limitado.

Adicionalmente, uma comparação com a quantidade total de amostras mostrado por Bucci também foi executada. De acordo com ele [92], os graus de liberdade N_B do campo amostrado é dada por $N_B = \frac{4}{\pi} (k_0 l)^2 x_1$, em que x_1 é um fator de sobreamostragem levemente maior que a unidade, variando geralmente entre $x_1 = 1.15$. Tal relação pode ser considerada como a versão eletromagnética do teorema de Shannon-Nyquist. Logo, a relação representa o número mínimo de amostras para reconstruir o campo de maneira perfeita. Contudo essa técnica utiliza funções bem custosas computacionalmente para reconstruir o campo, além de necessitar de posições angulares específicas de amostragem. Dessa maneira, a razão:

$$\frac{N_S}{N_B} = \frac{\pi x_s}{2x_1} \left(1 + \frac{1}{x_s k_0 l}\right) \left(1 + \frac{1}{2x_s k_0 l}\right), \quad (3.39)$$

mostra que apesar do número de amostras ser ligeiramente maior que o número ótimo dado por [92], a técnica proposta define um mapa de amostragem retangular que é fácil de ser implementada em equipamentos de caracterização e uma reconstrução perfeita do campo eletromagnético amostrado, fornecendo o acesso ao campo em qualquer posição espacial que esteja fora da esfera de convergência de Wilcox. Dessa maneira a técnica proposta resulta em um problema bem condicionado, melhorando a independência estatística dos coeficientes resultantes já que suas matrizes praticamente são diagonais.

3.4 Análise Estatística

O ruído associado às medições pode ter as mais diversas origens, como interferência de outras fontes, efeito de multicaminho no ambiente de caracterização, desalinhamento da AUT, e o ruído causado pelo equipamento de caracterização.

Como a natureza é completamente desconhecida, o ruído presente nas medições ϵ é um ruído randômico centrado, independente e identicamente distribuído, de variância σ^2 a ser adicionado a cada polarização medida. É importante notar que o ruído é também representativo nos dígitos significativos presentes no campo reconstruído.

3.4.1 Propagação do Ruído

A matriz que estima o ruído presente medido possui distribuição $[\hat{\epsilon}] = [\hat{E}^R] - [E^s] \sim \mathcal{RV}(0_{M,2L}, \Omega_\phi)$, where $\Omega_\phi = \mathbb{E}([\hat{\epsilon}] \otimes [\hat{\epsilon}]^*) = \sigma^2 \mathbb{I}_{(M \times 2L)^2}$, com tensor unitário $\mathbb{I}_{(M \times 2L)^2} = \delta_{il} \delta^{jk} e_{jk}^{il}$ em \sum que segue a notação de Einstein para a base canônica e . Como $\mathbb{E} \|\epsilon\|_F^2 = 2LM\sigma^2$, em que $\|A\|_F$ está de acordo com a norma de Frobenius da matriz A , a variância natural do estimador é $\hat{\sigma}^2 = \mathbb{E} \|\hat{\epsilon}\|_F^2 / 2LM$. O ruído destruído estimado é $[\hat{\epsilon}] = M_\Phi[\epsilon] \sim M_\Phi \mathcal{RV}(0_{M,2L}, \Omega_\phi)$.

O erro de propagação então leva para transformada de Fourier intermediária $[F] = e^{jk_0 r} k_0 r \left((\Phi^H V_\Phi \Phi)^{-1} \Phi^H V_\Phi [E^s] \right)^T$, em que $F^m = e^{jk_0 r} k_0 r \left(C_m \left((\Phi^H V_\Phi \Phi)^{-1} \right) \Phi^H V_\Phi [E^s] \right)^T$, para $C_m(A)$ na m -ésima coluna da matriz A . Em outras palavras, $F^m \sim \mathcal{RV}(0_{2L}, \Omega_{\theta,m})$, em que a covariância é $\Omega_{\theta,m} = (k_0 r)^2 \mathbb{E}([E^s]^H A_{\Phi,m} [E^s])$, com matriz intermediária $A_{\Phi,m} = V_\Phi \Phi C_m \left((\Phi^H V_\Phi \Phi)^{-1} \right) C_m \left((\Phi^H V_\Phi \Phi)^{-1} \right)^H \Phi^H V_\Phi$ dessa maneira $A_{\Phi,m} = A_{\Phi,m}^H \succeq 0$. Então, a covariância reduz-se para $\Omega_{\theta,m} = \sigma^2 \text{Tr}(A_{\Phi,m}) I_{2L \times 2L} \succeq 0$, em que Tr denota o operador de traço. Dessa maneira os coeficientes modas são $\hat{Q}_m = \frac{\sqrt{\eta_0}}{k_0} D_m \Theta_m^H W_m F^m$. Em algum grau, $\hat{Q}_m \sim \mathcal{RV}(Q_m, \sigma^2 \Gamma_{\theta,m})$, em que $\Gamma_{\theta,m} = \text{Tr}(A_{\Phi,m}) \frac{\eta_0}{k_0^2} D_m \Theta_m^H W_m W_m \Theta_m D_m^H \succeq 0$.

3.4.2 Diagnóstico Residual

Em ordem de verificar a coerência no proposto modelo de regressão utilizando nos dados experimentais, é necessário analisar o erro residual estatístico ϵ , através de algumas propriedades como independência, homoscedasticidade e normalidade.

3.4.2.1 Independência

A função de autocorrelação, *Autocorrelation Function* (ACF), mede a similaridade entre as observações como uma função da posição angular e o atraso associado. Consequen-

temente, uma matriz predominantemente diagonal e de correlação normalizada da versão vetorizada do erro residual ϵ mostra um alto grau de independência estatística de ϵ .

3.4.2.2 Homocedasticidade

O ruído residual deve ter um alto grau de similaridade com o ruído gaussiano branco, e isso é checado através da homoscedasticidade. Para fazer isso uma série de testes foram desenvolvidos [93, 94, 95, 96, 97, 98]. Todos eles verificam a variância dos erros a partir de uma regressão linear que é dependente dos valores das variáveis independentes. Neste caso, existe a heterocedasticidade e no caso contrário, a homoscedasticidade. Os testes mais comuns para verificação de homoscedasticidade são o de Breush-Pagan com modificação de Koenker, ou *Breush-Pagan with Koenker Modification* (BPK), e o teste de White (W). Um valor baixo de p demonstra a independência da variância (homoscedasticidade), e em outras palavras, o ruído associado é do tipo branco.

3.4.2.3 Normalidade

De acordo com o teorema do limite central aplicado nas inúmeras fontes aleatórias, complexas, independentes e identicamente distribuída envolvidas no processo de medição, o ruído adicionado nas medições ϵ é esperado ser bem próximo ao ruído gaussiano complexo com distribuição aleatória no fim de todo o processo de reconstrução. Para checar o quão perto desse comportamento o ruído se aproxima, a literatura possui um grande acervo de testes de normalidade [99, 100, 101]. Logo, o conjunto de dados são submetidos ao teste de amostra única t -teste, *One-Sample t-test*, [102, 103, 104, 105], teste de Shapiro-Wilk [85] e o teste de Kolmogorov-Smirnov (KS) [102, 103, 104, 105]. Nestes testes de normalidade, a hipótese nula H_0 é que dada uma distribuição ela é normal sem média e variância especificadas. O valor de p representa a probabilidade de o resultado obtido confirmar ou não a hipótese H_0 . Com a normalidade checada, é possível considerar se o ruído associado é uma variável complexa de distribuição $[\epsilon] \sim \mathcal{CN}(0_{M,2L}, \Omega_\phi)$, independente e normalmente distribuída.

3.4.3 Filtragem de Coeficientes Esféricos

Como o ruído $[\epsilon] \sim \mathcal{CN}(0_{M,2L}, \sigma^2 \mathbb{I}_{(M \times 2L)^2})$ é uma variável aleatória e $V = \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_k^2$ é uma distribuição Chi-2 com $k = 2LM - 1$ graus de liberdade, $\forall i = 1 \dots 2(N - \mu + 1)$, $(\Gamma_{\theta,m})_{i,i} \neq 0 \Rightarrow Z_{m,i} = \frac{(\hat{Q}_m - Q_m)_i}{\sqrt{\sigma^2 (\Gamma_{\theta,m})_{i,i}}} \sim \mathcal{CN}(0, 1)$. Da mesma maneira a variável

aleatória $\forall i = 1 \dots 2(N - \mu + 1)$, $(\Gamma_{\theta,m})_{i,i} \neq 0 \Rightarrow Z_{m,i} = \frac{(\hat{Q}_m - Q_m)_i}{\sqrt{\sigma^2 (\Gamma_{\theta,m})_{i,i}}} \sim \mathcal{CN}(0, 1)$.

Quando $(\Gamma_{\theta,m})_{i,i} = 0$, não há necessidade de ir além já $(\hat{Q}_m)_i$ para ser considerada ser

determinística.

Do contrário, como $\mathbb{E}(Z_{m,i}V) = 0$, $Z_{m,i}$ e V sendo independente de acordo com o teorema Cochran [106] e $T_{m,i} = Z_{m,i}\sqrt{\frac{k}{V}} \sim \mathcal{C}t_k(0, 1)$ uma distribuição *t-student*. Tal distribuição, que uma é quantidade fundamental e conhecida como a estatística auxiliar, não depende dos valores dos parâmetros não observáveis (média, variância por exemplo), o que torna robusto para as amostras com o custo de serem dependentes do modelo como a normalidade de algumas distribuições. Como o grau de liberdade k facilmente alcança valores na casa das centenas, e usando a propriedade de $\mathcal{C}t_k(0, 1) \xrightarrow[k \rightarrow \infty]{\text{dist}} \mathcal{CN}(0, 1)$ que começa centrada na distribuição de t de Student tende para a normal no sentido limítrofe para altos graus de liberdade, então $T_{m,i} \sim \mathcal{CN}(0, 1)$. Em outras palavras, $|T_{m,i}| \sim \text{Rayleigh}(1)$.

Agora o teste de hipótese é aplicado com a hipótese nula H_0 de que não há contribuição do sinal no modo esférico para determinar a região de confiança ou significância estatística para cada coeficiente esférico, o que expressa nada mais do que a probabilidade do erro do tipo I. Para fazer isso, o valor de p , ou probabilidade que a variável aleatória $|T_{m,i}|$, com função de densidade de probabilidade $\rho_{|t_{m,i}|}(\tau) = \tau e^{-\tau^2/2}, \tau \geq 0$, possui seu valor acima de certo limiar τ_p , também referido como o nível de significância é dado por $1 - p = \mathbb{P}(|T_{m,i}| \leq \tau_p | H_0) = \int_0^{\tau_p} \rho_{|t_{m,i}|}(x) dx = 1 - e^{-\tau_p^2/2}$. Consequentemente, $\tau_p = \sqrt{-2 \ln p}$.

Dessa maneira, $(|T_{m,i}| \leq \tau_p) \iff (|\hat{Q}_m - Q_m| \leq \Delta_{p,m,i})$, em que a região de confiança é $\Delta_{p,m,i} = \tau_p \hat{\sigma} \sqrt{(\Gamma_{\theta,m})_{i,i} / k}$. Tal região representa a fração do ruído que foi inserido no modo estimado $(\hat{Q}_m)_i$ para o valor selecionado de p . Uma razão de sinal ruído modal é definida:

$$r_{p,m,i} = \left| \frac{(\hat{Q}_m)_i}{\Delta_{p,m,i}} \right|^2. \quad (3.40)$$

Logo, o teste estatístico é então equivalente a aceitar apenas modos em que a relação sinal ruído modal é estritamente maior que 0 dB.

Capítulo 4

Reconstrução de Campo com Inteligência Artificial

4.1 Construção da Base de Dados

Como explicado em 2.3.3, a base de dados é um elemento importante para que os algoritmos de aprendizagem de máquina possam aprender e executar a tarefa para o qual foram treinados. Essa seção irá tratar como foi construída a base de dados para o treinamento dos algoritmos LR, DT, RF, KNN, SVM e NN para que os mesmos sejam capazes de compreender e consigam resolver o problema de recuperação de fase mostrado no item 2.2.

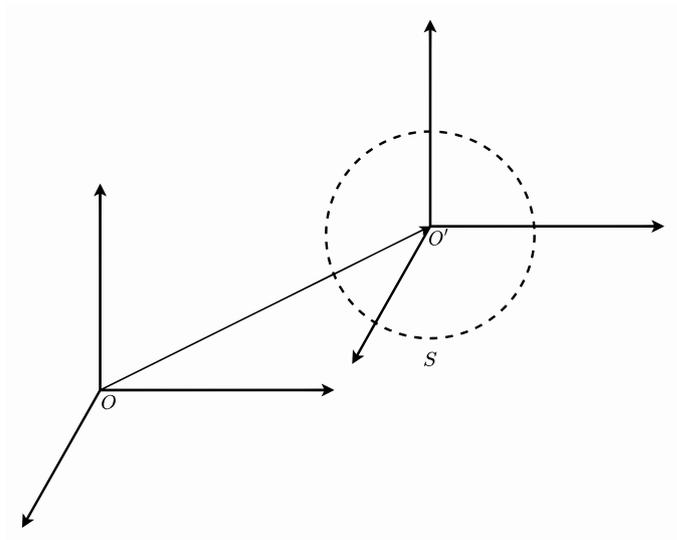
4.1.1 Translação da Superfície de Aquisição

Como foi abordado no item 2.3.3, os algoritmos de aprendizagem de máquina supervisionada necessitam que um agente externo construa a relação entre as variáveis independentes e a dependente, e essa relação é obtida por meio da base de dados. A partir dessa premissa, foi-se pensado uma maneira de obter uma base de dados robusta e fisicamente coerente para que os algoritmos de aprendizagem de máquina possam gerar modelos de inteligência artificial capazes de prever a fase de campo elétrico E a partir de dados de amplitude.

Basicamente, a estratégia busca causar variações na distribuição de fase do campo observado na esfera de aquisição de raio r_e . Essa ideia foi baseada nos algoritmos de dupla varredura, que fornecem diferentes pontos de vista do campo eletromagnético para que os algoritmos consigam recuperar a fase. No âmbito de caracterização de antenas, essas diferentes pontos de vista são obtidos de duas formas principais: a partir de duas superfícies alinhadas [16] ou a partir de translação da AUT em relação ao eixo global de referência em duas posições distintas utilizando apenas uma superfície [107, 42]. Em [4, 3] é mostrado o fato de que a translação da AUT têm um efeito similar à de realizar aquisição em duas superfícies não alinhadas, enquanto a AUT está no centro do sistema de coordenadas O como

mostrado na figura 4.1. Tal conceito foi utilizado para a construção da base de dados. O

Figura 4.1: Esquemática do posicionamento entre os dois sistemas de coordenadas O e O' .



entendimento disso vem a partir da consideração de dois sistemas de coordenada O , onde está localizada a AUT, e O' , onde está localizada a esfera de medição. Considerando que o ponto de aquisição p na esfera de aquisição S com raio de aquisição r_e possui coordenadas (x', y', z') e que \mathbf{l} é o vetor associado a localização da origem de O' em O com coordenadas $(\Delta x, \Delta y, \Delta z)$, a translação de p para O com coordenadas (x, y, z) é feita da seguinte maneira:

$$\begin{aligned} x &= x' + \Delta x, \\ y &= y' + \Delta y, \\ z &= z' + \Delta z. \end{aligned} \quad (4.1)$$

As coordenadas (x', y', z') podem ser expressas com dependência angular da seguinte maneira:

$$\begin{aligned} x' &= r_e \sin \theta' \cos \phi', \\ y' &= r_e \sin \theta' \sin \phi', \\ z' &= r_e \cos \theta'. \end{aligned} \quad (4.2)$$

Logo, é possível substituir as relações mostradas na equação 4.2 em 4.1, obtendo as coordenadas de p (x, y, z) em função das suas coordenadas esféricas (r', θ', ϕ') :

$$\begin{aligned} x &= r_e \sin \theta' \cos \phi' + \Delta x, \\ y &= r_e \sin \theta' \sin \phi' + \Delta y, \\ z &= r_e \cos \theta' + \Delta z, \end{aligned} \quad (4.3)$$

em que $r' = r_e$. Seguindo o mesmo raciocínio, as coordenadas esféricas de p em O são:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2}, \\ \theta &= \arctan \frac{\sqrt{x^2 + y^2}}{z}, \end{aligned}$$

$$\phi = \arctan \frac{x}{y}. \quad (4.4)$$

No caso de $\Delta x = \Delta y = \Delta z = 0$ $O = O'$ o que significa que a superfície de aquisição está centrada em O . Como discutido no item 5.5 e amplamente pela literatura, a distribuição de fase está intimamente ligada à distância de observação. Dessa maneira caso o ponto p seja transladado para O o raio de aquisição r_n em tal sistema de coordenadas é definido como:

$$r_n = \sqrt{(r_e \sin \theta' \cos \phi' + \Delta x)^2 + (r_e \sin \theta' \sin \phi' + \Delta y)^2 + (r_e \cos \phi' + \Delta z)^2}. \quad (4.5)$$

A partir disso fica bastante claro que o processo de projeção de p em O causa flutuações em r_n . Dessa maneira o campo elétrico \mathbf{E} em O pode ser escrito como [2]:

$$\mathbf{E} = \frac{jk\eta e^{-jkR}}{4\pi R} (\hat{\mathbf{r}} \times \mathbf{F} \times \hat{\mathbf{r}}), \quad (4.6)$$

em que $R = \sqrt{(x - x_f)^2 + (y - y_f)^2 + (z - z_f)^2}$ é a distância de observação. Considerando que a fonte está na origem do sistemas de coordenadas, isto é, $x_f = y_f = z_f = 0$, resulta em $R = r_n$. Analisando o termo e^{-jkR} da equação 4.5 fica claro que o movimento de translação a esfera de aquisição de O' para O causará uma perturbação na fase vista em tal sistema de coordenada. Já no diz respeito às amplitudes, as relações entre as componentes do campo nos dois sistemas de coordenadas são expressas da seguinte maneira [108]:

$$\begin{aligned} E'_x &= Ex, & E'_y &= \gamma(E_y - vB_z), & Ez' &= \gamma(E_z + vB_y) \\ B'_x &= Bx, & B'_y &= \gamma\left(B_y + \frac{v}{c_0}B_z\right), & Bz' &= \gamma\left(B_z - \frac{v}{c_0}B_y\right), \end{aligned} \quad (4.7)$$

em que $\gamma = 1/\sqrt{1 - v/c_0^2}$ e c_0 é o termo de contração de Lorentz e c_0 é velocidade da luz no vácuo. Como a superfície de aquisição não está em movimento, pode-se considerar que v é praticamente zero, o que faz $\gamma = 1$ e conseqüentemente:

$$\begin{aligned} E'_x &= Ex, & E'_y &= E_y, & Ez' &= E_z, \\ B'_x &= Bx, & B'_y &= B_y, & Bz' &= B_z. \end{aligned} \quad (4.8)$$

Utilizando a relação de conversão entre as coordenadas retangulares e esféricas, os campos E_r, E_θ, E_ϕ em O podem ser obtidos como [2, 25]:

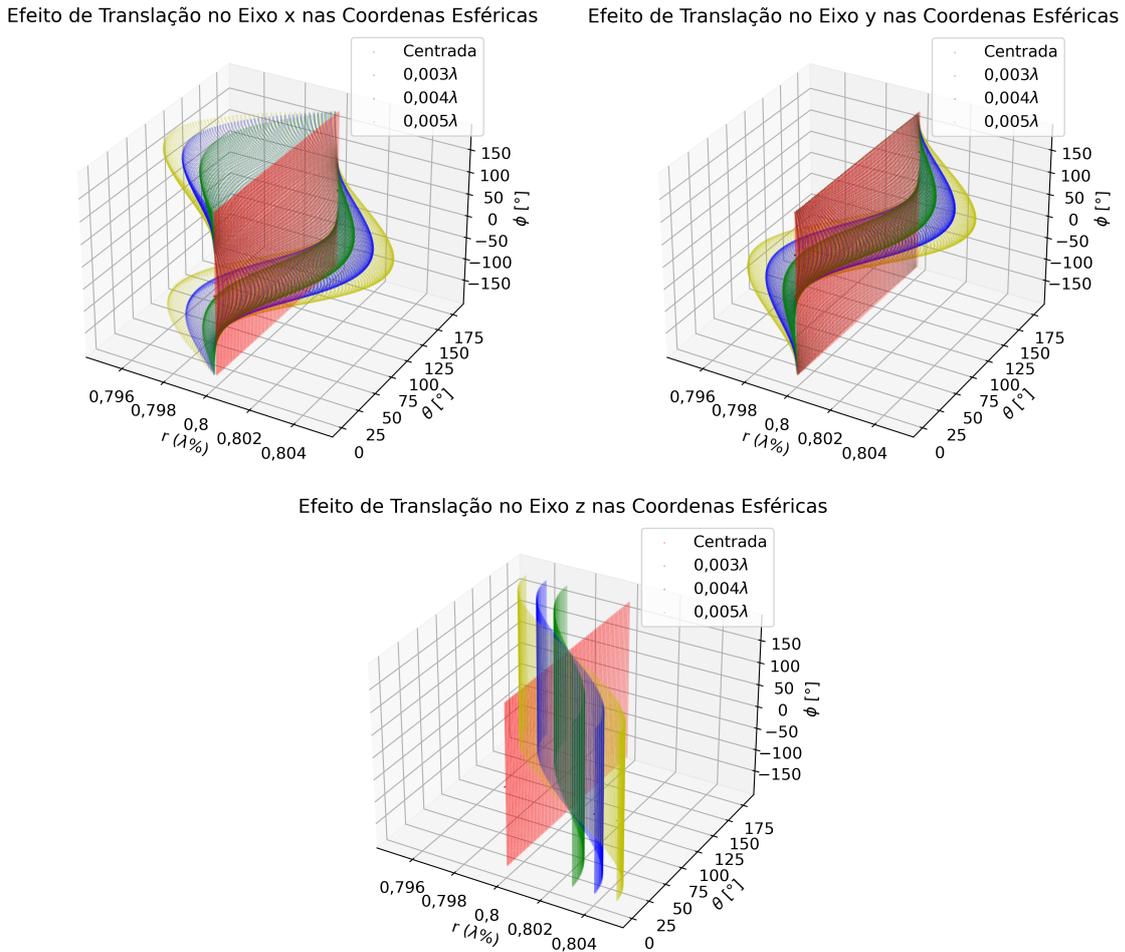
$$\begin{bmatrix} E_r \\ E_\theta \\ E_\phi \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}. \quad (4.9)$$

A partir dessas definições fica claro que para uma superfície de aquisição com P pontos experimentais $1 < p < P$ em O' , a translação irá causar variações na distribuição de fase do campo quando o mesmo é observado em O . Os efeitos do movimento de translação entre os sistemas de coordenadas pode ser visto na figura 4.2 para translações em torno de x, y e

z de O . Matematicamente, para a translação feita no eixo z [42], o raio de aquisição, e por consequência a distância de observação, assume a forma:

$$r_n = \sqrt{(r_e \sin \theta' \cos \phi')^2 + (r_e \sin \theta' \sin \phi')^2 + (r_e \cos \phi' + \Delta z)^2}. \quad (4.10)$$

Figura 4.2: Efeito de translação nas coordenadas de uma superfície S quando transladada de O' para O . Nota-se que o raio é influenciado pelas coordenadas esféricas dos pontos de S em O' . A figura a mostra o efeito caso a translação for no eixo x , a figura b caso for feita no eixo y e a figura c mostra o caso da translação no eixo z .



4.1.2 Sequência de Aprendizagem

Os algoritmos de aprendizagem de máquina supervisionada usados e disponíveis em [69] para resolver problemas de regressão, estão no fundo resolvendo um sistema de equações do tipo quando fazem as predições:

$$\mathbf{y} = \mathbf{Ax}, \quad (4.11)$$

em que $\mathbf{y} \in \mathbb{R}^P$, $\mathbf{x} \in \mathbb{R}^M$ e $\mathbf{A} \in \mathbb{R}^{P \times M}$ em que M denota o número de elementos preditores, ou variáveis independentes para $1 \leq m \leq M$ e P o número de amostras. Mas diferentemente das estratégias mostradas nos itens 2.1 e 2.2 em que \mathbf{y} e \mathbf{A} são conhecidos, na aprendizagem de máquina supervisionada os vetores \mathbf{y} e \mathbf{x} são conhecidos, enquanto \mathbf{A} não. Dessa maneira os algoritmos SL devem resolver o seguinte problema:

$$\begin{aligned} &\text{Encontre } \mathbf{A}, \\ &\text{tal que } \mathbf{y} = \mathbf{A}\mathbf{x}. \end{aligned} \quad (4.12)$$

Para o problema de recuperação de fase, os modelos devem ser capazes de obter a fase do campo eletromagnético a partir das entradas de amplitude, logo deve-se encontrar a melhor base de dados para resolver esse problema. Muitas investigações foram empregadas, em que a primeira delas foi baseada na formulação apresentada por Gerchberg-Saxton [10], em que somente as amplitudes medidas do campo elétrico de uma antena [2, 25], e a distribuição de fase foram fornecidas para o treinamento dos algoritmos. Primeiramente uma quantidade $T = K \cdot P$ de amostras de amplitude para um conjunto de superfícies de medição $\mathbf{S} = (S_1, S_2, \dots, S_K)$, para $1 \leq k \leq K$, de raios de aquisição r_e foram coletados, em que a S_K esferas estão transladadas por distâncias d_1, d_2, \dots, d_K em relação ao eixo z de O , gerando superfícies semiesferas de raios $r_{d_1}, r_{d_2}, \dots, r_{d_K}$ em que r_{d_K} segue a equação 4.10. O fato dessas superfícies resultantes serem semiesféricas advém do fato do raio resultante r_{d_K} ser "modulado" por r_e com dependência radial.

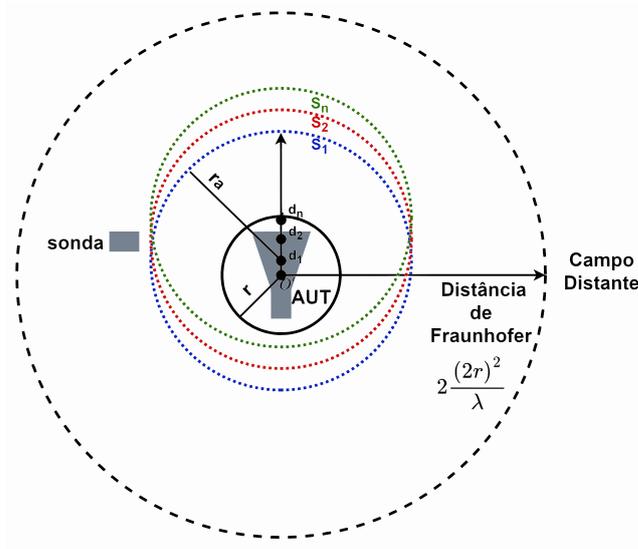
A escolha de tal geometria de aquisição é direta: elas possuem a capacidade de fornecer uma visão do completa do campo em r_e , ou seja, em todo 4π esterradiano da esfera S_K dependendo da estratégia de medição utilizada (equiangular, iglu etc), e são mais bem documentadas devido a suas utilizações em reconstrução de campo em SWE. Além disso são mais flexíveis no que diz respeito a investigações de validação [3]. Dessa maneira, a sequência de aprendizagem preliminar dada aos algoritmos de SL foi:

$$\mathbf{x} = \begin{bmatrix} \Lambda_{d1}(r_{d1}, \boldsymbol{\theta}_{d1}, \boldsymbol{\phi}_{d1}) \\ \Lambda_{dK}(r_{dK}, \boldsymbol{\theta}_{dK}, \boldsymbol{\phi}_{dK}) \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \mathbf{f}_{d1}(r_1, \boldsymbol{\theta}_{d1}, \boldsymbol{\phi}_{d1}) \\ \mathbf{f}_{dK}(r_K, \boldsymbol{\theta}_{dK}, \boldsymbol{\phi}_{dK}) \end{bmatrix}, \quad (4.13)$$

em que $\Lambda_{d_K} = [\Lambda_1(r_e, \boldsymbol{\theta}_K, \boldsymbol{\phi}_K), \dots, \Lambda_P(r_e, \boldsymbol{\theta}_K, \boldsymbol{\phi}_K)]$ simboliza o bloco de amostras de amplitude na esfera S_K , em que $\boldsymbol{\theta}_K = (\theta_1, \dots, \theta_P)$ e $\boldsymbol{\phi}_k = (\phi_1, \dots, \phi_P)$ o bloco das coordenadas dos pontos P de S_K transladados para O , em que a esfera possui distância de translação d_K . As mesmas considerações podem ser feitas para \mathbf{f}_{d_K} , que obviamente simboliza os dados de fase obtidos. Para verificar se os algoritmos de SL utilizados (LR, DT, RF, KNN e SVM) conseguem encontrar uma matriz \mathbf{A} capaz de estabelecer uma relação satisfatória, o base de dados preliminar foi separada como mostrado na figura 2.27 em uma proporção de 70% de \mathbf{y} e \mathbf{x} foram separados para treinamento, enquanto 30% de ambos foram usados para verificar o rendimento do modelo obtido.

Esse rendimento preliminar foi verificado por intermédio da métrica acurácia, que apesar

Figura 4.3: Ilustração do de aquisição de dados de amplitude para o treinamento os algoritmos de aprendizado para resolver o problema de recuperação de fase. A utilização de uma esfera permite uma visão completa do campo eletromagnético, além do fato de as estratégias de medição de campo em tal superfície é bem documentada e validada [3].



de ser mais comum em aplicações de classificação, nos estágios iniciais de investigação é muito útil para verificar o quanto o vetor de predições \hat{x} está próximo de x . Para essa primeira abordagem, todos os algoritmos citados tiveram uma acurácia abaixo de 50%. Logo, eles não foram capazes de fazer boas estimativas da distribuição da fase.

As discussões feitas pela literatura sobre reconstruções de campo e recuperação de fase apresentadas nos itens 2.1 e 2.2 mostram que é possível utilizar alguma informação conhecida previamente sobre o campo, ou seja, alguma informação obtida *a-priori*. Muitas configurações de informações *a-priori* sobre o campo irradiado pela AUT usada foram investigadas, mas uma delas se destacou e a mais promissora. Ela utilizada as posições angulares θ_{d_K} e ϕ_{d_K} para compor um vetor x , da seguinte forma:

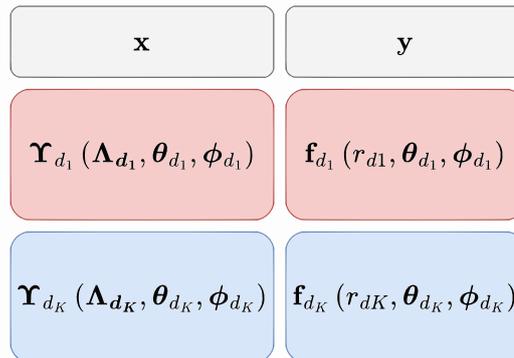
$$\mathbf{x} = \begin{bmatrix} \Upsilon_{d_1}(\Lambda_{d_1}, \theta_{d_1}, \phi_{d_1}) \\ \Upsilon_{d_K}(\Lambda_{d_K}, \theta_{d_K}, \phi_{d_K}) \end{bmatrix}. \quad (4.14)$$

em que Λ_{d_K} continua sendo as amplitudes do campo amostradas na esfera S_K e transladadas em um distância d_K em relação a O no eixo z . Devido ao alto rendimento, as análises preliminares apontaram para um acurácia maior que 79% para os algoritmos utilizados, tal sequência de entrada x foi adotada. Logo a nova sequência de aprendizagem adotada foi:

$$\mathbf{x} = \begin{bmatrix} \Upsilon_{d_1}(\Lambda_{d_1}, \theta_{d_1}, \phi_{d_1}) \\ \Upsilon_{d_K}(\Lambda_{d_K}, \theta_{d_K}, \phi_{d_K}) \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \mathbf{f}_{d_1}(r_{d_1}, \theta_{d_1}, \phi_{d_1}) \\ \mathbf{f}_{d_K}(r_{d_K}, \theta_{d_K}, \phi_{d_K}) \end{bmatrix}. \quad (4.15)$$

A figura 4.4 ilustra a sequência de aprendizado esquematizada na equação 4.15 e utilizada ao longo deste trabalho. Já a figura 4.3 mostra a estratégia de aquisição de dados nas superfícies esféricas S_K .

Figura 4.4: Visualização da base dados utilizada, em que $\Upsilon_{d_K}(\Lambda_{d_K}, \theta_{d_K}, \phi_{d_K})$ simboliza os dados de amplitude e posições angulares, e $\mathbf{f}_{d_K}(r_{d_K}, \theta_{d_K}, \phi_{d_K})$ os dados de fase. Os dados são adquiridos a partir do esquema de medição mostrada na figura 4.3.



4.2 Processo de Otimização, Validação e Obtenção dos Modelos

4.2.1 Otimização por *Grid Search*

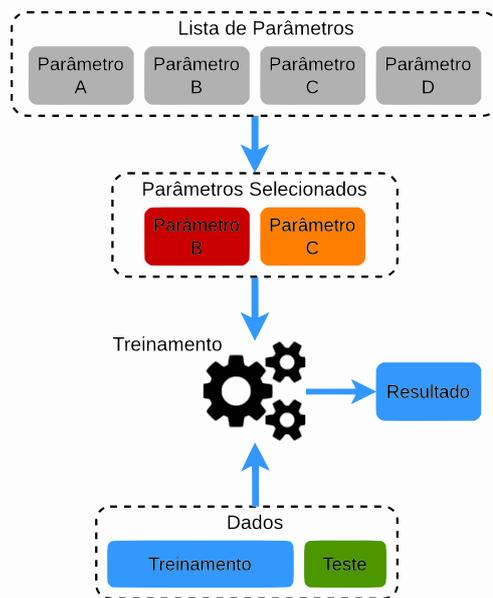
A etapa descrita anteriormente de separar a base de dados composta por x e y em uma fração de treinamento e outra para teste é uma etapa intermediária e seu uso é recomendado somente para análises preliminares. Além disso os algoritmos presentes na ferramenta *Scikit-Learn* possuem parâmetros padrão, que podem levar os algoritmos a não performarem de maneira ótima. Logo, a técnica de *Grid Search* é utilizada para contornar este problema.

Tal técnica possui como objetivo testar todas as combinações possíveis dos parâmetros associados a um algoritmo de inteligência artificial para encontrar a melhor combinação entre eles e encontrar o modelo. Essa tarefa também funciona de maneira interativa e é bastante similar ao processo de validação cruzada, em que um grupo de parâmetros são utilizados para treinar um determinado algoritmo e o seu rendimento é testado para uma determinada métrica (acurácia, R^2 , etc), determinando assim a melhor combinação e que vai obter o melhor modelo possível.

4.2.2 Validação Estatística dos Algoritmos

O processo de validação de cruzada CV é bastante importante no processo de aprendizado de máquina supervisionado como explicado no item 2.3.4. O modelo deve ao mesmo tempo fazer previsões \hat{x} precisas e ser generalista a tal ponto que possa fazer previsões de maneira satisfatória quando a inteligência artificial são inseridas é exposta a entradas não vistas antes. Contudo avaliar somente uma métrica de rendimento, por exemplo a acurácia, apesar de ser um processo importante não é interessante para a aplicação em discussão. O processo de PR é NP-difícil, e dessa maneira para validar a capacidade dos modelos gerados para tal tarefa

Figura 4.5: Esquemática do processo de *grid search*. Uma certa configuração dos parâmetros de um algoritmo é selecionada e inserida no mesmo. Logo em seguida ele é treinado, testado e o resultado é armazenado para futuras comparações e análises.



deve ser mais profunda.

Neste trabalho, todos os algoritmos LR, DT, RF, SVM, KNN e NN foram submetidos a quatro validações cruzadas. Duas para as métricas de verificação adequação dos modelos, que são as métricas acurácia e R^2 e duas para as métricas erro, que serão o erro absoluto médio MAE e erro quadrático médio SME. O intuito dessas quatro validações cruzadas é verificar o quão preciso e adequado são algoritmos e verificar o quão distante as predições podem estar dos valores reais. Para que um algoritmo seja selecionado para criar um modelo o mesmo deve ser aprovado em pelo menos uma métrica de adequação (acurácia e R^2) e uma métrica de erro (MAE e MSE) em relação aos concorrentes nos testes de Shapiro-Wilk, ANOVA e Tukey.

Esse critério foi estabelecido pois além de mostrar que os algoritmos são robustos o suficiente para fazer previsões, no processo de validação cruzada pode ocorrer o fato e que os resultados intermediários obtidos possuam uma distribuição normal em uma métrica e não em outra, sendo então assim reprovado no teste de normalidade. No cenário em que somente uma métrica seja submetida a CV para avaliar um determinado algoritmo e os resultados não estejam em uma distribuição normal, ele não poderá progredir para a etapa que são os testes ANOVA e Tukey, que possuem como intuito selecionar aqueles algoritmos cuja os resultados possuam distribuições normais distintas entre si, e caso esse algoritmo fosse promissor ele seria descartado de maneira prematura.

Todas as quatro validações cruzadas são feitas utilizando 10 grupos e 30 testes por iteração [69], totalizando de 300 testes. Dessa maneira todos os algoritmos serão submetidos a um total de 1200 validações, indicando que se um algoritmo alcançou um bom resultado em todas essas avaliações ele irá conseguir prever a fase a partir de novos dados de entrada.

O critério utilizado é que os algoritmos devem obter valores de acurácia e de coeficiente de determinação entre 78% e 96% e que apresente o menor erro entre as correntes. Algo a ser notado é que todos os testes de hipótese têm um nível de confiabilidade de 95%, ou de maneira complementar, um nível de significância de 5%.

4.2.3 Obtenção dos Modelos

Após a validação cruzada, os algoritmos que foram aprovados no teste de normalidade SW são submetidos ao teste ANOVA, que mostra que as distribuições obtidas são distintas entre si e o teste de Tukey que procura diferenças entre os pares de distribuições. O algoritmo será selecionado caso ela atenda ao critério estabelecido no item 4.2.2 e ele poderá prosseguir para treinamento. Como a base de dados foi submetida a um exaustivo processo de validação estatística, o processo de obtenção dos modelos utiliza a mesma em sua totalidade para realizar treinamento final de todos os algoritmos selecionados. Ao final desse processo os modelos obtidos são exportados para uso em futuras caracterizações.

Figura 4.6: Esquematização de como os algoritmos são selecionados para a geração de modelos. Somente aqueles que passam nos testes ANOVA e Tukey para uma métrica de adequação e uma de erro são selecionados para gerar modelos.

	Acurácia	R^2	MAE	SME	
Algoritmo 1	✓	✓	✓	✓	Aprovado
Algoritmo 2	✗	✓	✓	✗	Aprovado
Algoritmo 3	✗	✗	✓	✓	Reprovado

4.2.4 Validação Eletromagnética

Após a obtenção dos modelos, eles são submetidos em um cenário que simula uma reconstrução de campo utilizando uma nova esfera de aquisição. As predições de fase geradas são utilizadas para reconstruir o campo elétrico estimado $\hat{\mathbf{E}}$. Utilizando o campo real \mathbf{E} , é possível estimar o quão próximo os dois estão através da métrica global ENL, apresentada no item 2.1.3.2:

$$\text{ENL} = 20 \log_{10} \left(\text{mean} \left[\frac{\|\mathbf{E}(\theta, \phi) - \hat{\mathbf{E}}(\theta, \phi)\|}{\|\mathbf{E}(\theta, \phi)_{\max}\|} \right] \right). \quad (4.16)$$

Além disso, uma análise da influência do ruído é feita para verificar o quão robusto os modelos são para possíveis cenários de caracterização no mundo real, ou seja, levando em consideração o ruído.

Capítulo 5

Aplicações e Resultados

5.1 Reconstrução do Campo Eletromagnético para Dados Faltantes e Ruidosos

A técnica apresentada no item 3 foi implementada e testada em três modelos de antenas que operam em bandas distintas e possuem diretividades e padrões de radiações diferentes entre si. Elas foram: uma lente de Luneburg plana, do inglês *Luneburg Lens* (LL), operando a 12 GHz, um *Reflectarray* (RA) operando a 12 GHz e uma antena corneta, do inglês *Standard Gain Horn Antenna* (SGH), operando a 26 GHz. O processo de caracterização foi conduzido em campo distante em uma câmara anecoica utilizando uma estrutura CATR no Instituto de Eletrônica de Telecomunicações de Rennes (IETR), França. Os padrões tridimensionais dos elementos foram obtidos tanto em um setor angular limitado quanto em uma esfera completa. Vale ressaltar que as antenas também foram construídas no IETR.

Os parâmetros de amostragem estão dispostos na tabela 5.1. Já os resultados eletrônicos para cada antena são mostrados na tabela 5.2. A figura 5.1 mostra para todas as antenas estudadas, a relação entre o fator de regularização α e a variância do erro residual σ^2 , a potência de ruído N_{σ^2} , a relação sinal ruído SNR e a potência total.

Tabela 5.1: Resumo dos parâmetros de amostragem utilizados para as antenas caracterizadas em um mapa angular completo.

AUT	f [GHz]	l [m]	$k_0 l$	N	L	M	N_S	N_B
SGH	26	0,04	21,8	21	22	43	946	695
LL	12	0,26	65,3	65	66	131	8,646	6,252
RA	12	0,20	50,3	50	51	101	5,151	3,700

Nota 1: $x_1 = 1.15$ and $x_s = 1.00$.

Nota 2: No caso da SGH e RA, metade das posições do mapa foram anuladas para completar a esfera.

Tabela 5.2: Resumo dos parâmetros eletromagnéticos diretividade D , potência radiada em campo distante P_R^∞ , potência de ruído e relação sinal ruído SNR para as antenas SGH, LL e RA.

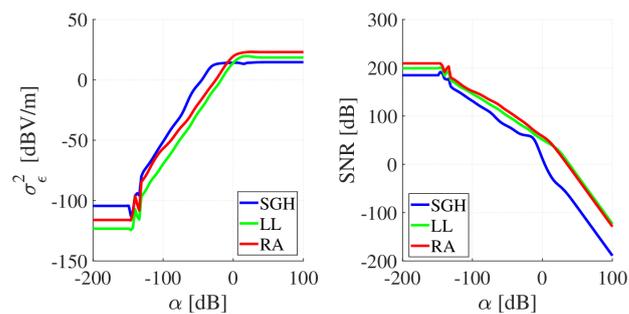
AUT	D [dB]	$\langle P_R^\infty \rangle_t$ [dBm]	N_{σ^2} [dBm]	SNR [dB]
SGH	18,7	8,77	-56,6	65,3
LL	21,3	-34,0	-101	67,5
RA	23,8	-32,8	-103	70,2

Tabela 5.3: Resultados da análise do ruído residual para as antenas SGH operando a 26 GHz, LL operando a 12 GHz e RA também a 12 GHz.

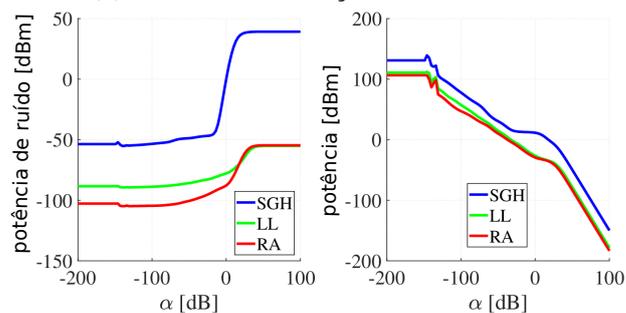
AUT	Momentos Estatísticos		Normalidade teste SW valor de p	Homocedasticidade teste BPK valor de p
	desvio padrão σ [V/m]	média μ [V/m]		
SGH	$1,59 \cdot 10^{-3}$	$2,36 \cdot 10^{-18}$	0	$3,18 \cdot 10^{-8}$
LL	$3,26 \cdot 10^{-5}$	$-9,70 \cdot 10^{-20}$	0	$2,28 \cdot 10^{-1}$
RA	$1,35 \cdot 10^{-5}$	$-2,29 \cdot 10^{-8}$	0	$3,09 \cdot 10^{-13}$

Testes estatísticos realizados com nível de significância de 5%.

Figura 5.1: Impacto do fator de regularização α na caracterização experimental executada para as antenas SGH, LL e RA. A figura (a) mostra tanto o quanto a variância do ruído residual σ^2 aumenta à medida que α aumenta, mantendo-se mais ou menos constante para $\alpha < -100$ dB. Já a relação sinal ruído decresce à medida que α aumenta. Já a figura (b) mostra como a potência de ruído só começa a aumentar para $\alpha > 0$ dB, enquanto potência possui um comportamento muito similar à SNR.



(a) Variância e Relação Sinal Ruído



(b) Potência de Ruído e Potência Total

5.1.1 Lente Plana de Luneburg

Uma antena de lente plana de Luneburg, *Luneburg Lens* (LL), operando na banda-X foi projetada na estrutura MERLIN, que é um laboratório compartilhado entre halès Alenia Space e o IETR. Detalhes sobre o comportamento de tal estrutura radiante são apresentados em [109]. Ambas as componentes E_θ e E_ϕ foram obtidas na esfera completa na frequência de 12 GHz. Os mapas bidimensionais de amplitude e fase de cada componente do campo medido e reconstruído são mostrados na figura 5.4. Nota-se que não há diferenças nítidas entre a amplitude e fase do campo medido e do reconstruído utilizando a técnica proposta em ambas as componentes. Já a figura 5.5-(a) fornece as visões em corte do padrão de radiação de E_θ e E_ϕ em campo distante nos setores angulares $\phi = 0$ e $\phi = \pi/2$. Essas figuras destacam a proximidade entre o padrão medido e o reconstruído, reforçando que a técnica proposta possui um rendimento satisfatório.

5.1.2 Antena Corneta

Uma antena corneta padrão, *Standard Gain Horn Antenna* (SGH), operando na banda Ka foi estudada na frequência de 26 GHz no CATR do IETR. Ambas as componentes E_θ e E_ϕ foram medidas em um setor circular limitado. Os mapas bidimensionais de amplitude e fase de cada componente para os campos medido e reconstruído são mostrados na figura 5.2. Aqui também fica nítido que no setor de interesse não é possível detectar diferenças significativas entre os campos medido e reconstruído.

A figura 5.5-(b) mostra o padrão de campo distante para ambas as componentes E_θ e E_ϕ nos setores angulares $\phi = 0$ e $\phi = \pi/2$ dos campos medidos e reconstruído, reforçando a similaridade entre eles. Para aplicar as expressões de diretividade e da potência radiada, é obrigatório possuir a expansão modal do campo válida para a esfera completa, sendo necessário completar o mapa parcial do campo amostrado por zeros no momento da obtenção dos coeficientes esféricos. Como consequência existe um efeito de Gibbs na metade inferior do plano devido a descontinuidade em $\theta = \pi/2$, que corresponde a região azul dos mapas de amplitude do campo reconstruído da figura 5.2, onde o campo é zerado para completar a esfera de varredura.

5.1.3 Reflectarray

Uma antena *reflectarray* (RA) composta de 193 células e projetada pela Thalès Alenia Space na estrutura do projeto R3MEMS foi caracterizado na banda X. Ambas as componentes E_θ e E_ϕ do campo eletromagnético foram medidas a partir de um setor angular limitado na frequência de 12 GHz. Os mesmos comentários feitos para a antena SGH no item 5.1.2 para o efeito de Gibbs, diretividade e potência podem ser feitos aos resultados obtidos para o *reflectarray*. Os mapas bidimensionais de amplitude e fase de cada componente dos campos

Figura 5.2: Mapas bidimensionais de amplitude e fase das componentes do campo distante E_θ e E_ϕ para comparação de campo elétrico mensurado e reconstruído para os parâmetros mostrados na tabela 5.1. A figura mostra os resultados para a antena corneta operando na banda Ka a 26 GHz no caracterizador CATR da IETR. Nota-se efeito de Gibbs no ponto de descontinuidade em $\theta = \pi/2$ onde o campo elétrico foi zerado devido a ausência de medições.

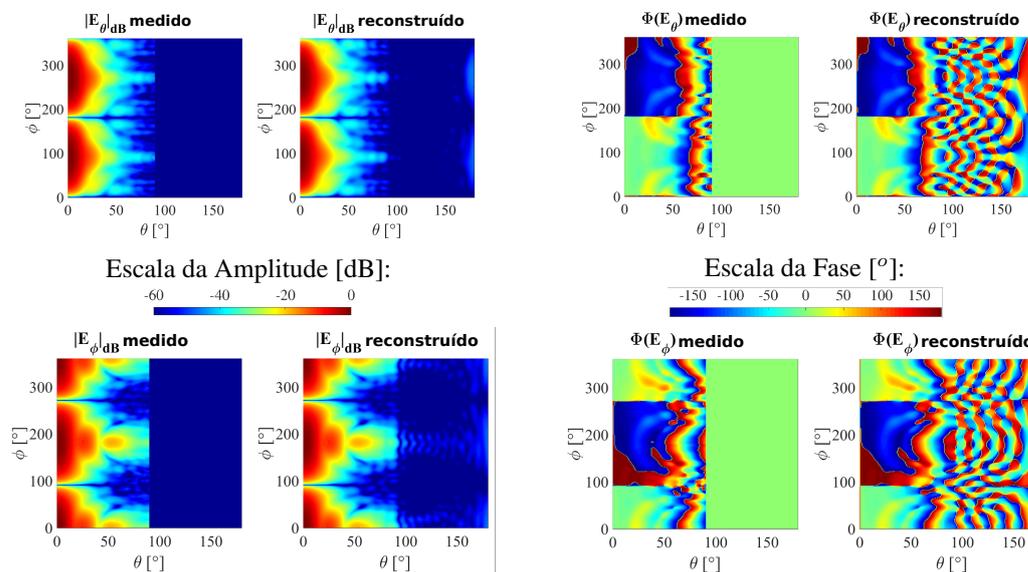
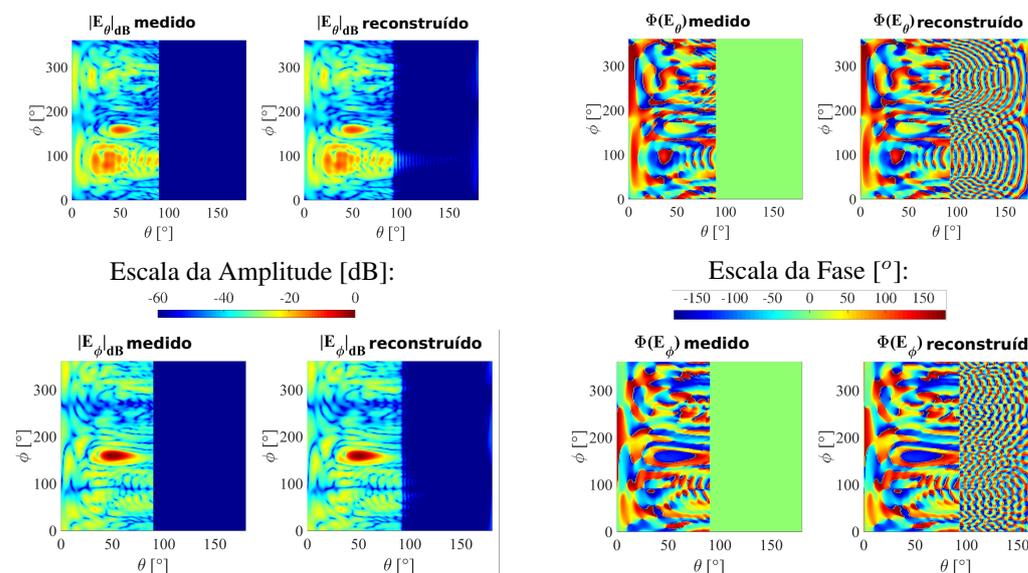


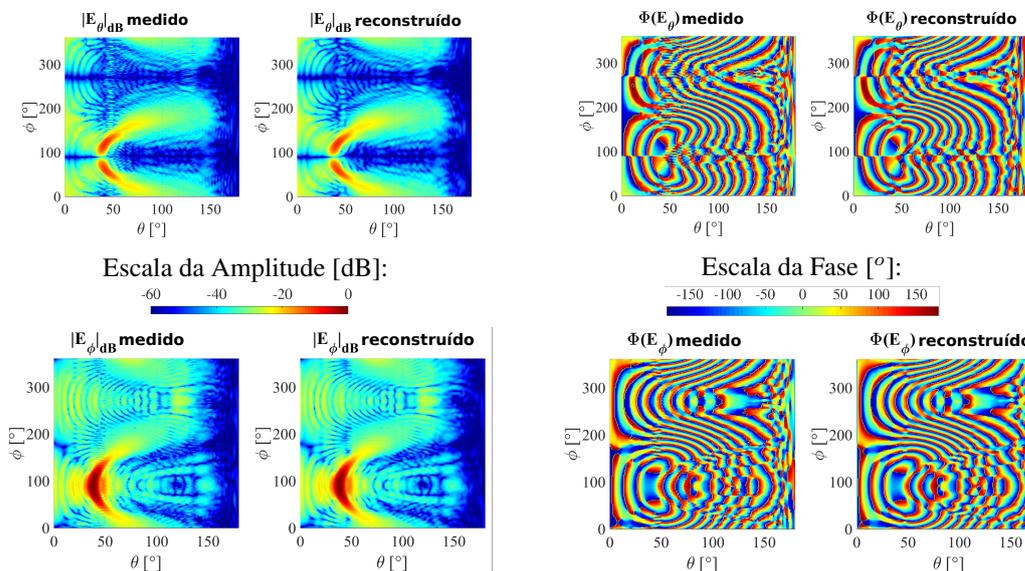
Figura 5.3: Mapas bidimensionais de amplitude e fase das componentes do campo distante E_θ e E_ϕ que compara o campo elétrico mensurado e o reconstruído para os parâmetros mostrados na tabela 5.1. Os resultados foram obtidos para o *reflectarray* de 193 células projetado pela Thalès Alenia Space no projeto R3MEMS. Nota-se efeito de Gibbs no ponto de descontinuidade em $\theta = \pi/2$ e isso ocorre que pois os dados campo elétrico são zerados devido a ausência de medições.



medido e reconstruído são mostrados na 5.3 e reforçam a capacidade da técnica proposta em reconstruir o campo com alto grau de fidelidade. Já a figura 5.5-(c) mostra o padrão de radiação para E_θ e E_ϕ em campo distante para os setores angulares $\phi = 0$ e $\phi = \pi/2$ e também

reforçam a capacidade da técnica proposta de reconstruir o campo de maneira satisfatória.

Figura 5.4: Mapas bidirecionais das componentes E_θ and E_ϕ em campo distante comparando amplitude e fase reconstruída e medida utilizando os parâmetros de amostragem da tabela 5.1 para a antena de lente plana de Luneburg operando à 12 GHz. Essa antena foi projetada na estrutura MERLIN, um laboratório conjunto entre Thalès Alenia Space e o IETR.



5.1.4 Análise do Ruído Residual

Para que a técnica proposta implementada seja considerada realmente satisfatória, os requisitos de que o erro seja uma variável centrada, com comportamento próximo do ruído branco e que não impacte a independência das matrizes associadas no processo de reconstrução e que ao final desse processo tenha uma distribuição normal como explicado no item 3. Primeiramente, como apresentado na tabela 5.2, a relação sinal ruído é alta para todas as antenas e isso está intimamente ligado ao fato do erro ser uma variável centrada. Isso pode ser visto na tabela 5.3 que mostra que a média e o desvio padrão do erro residual do processo de reconstrução. Para as famílias de antenas investigadas, o valor da média μ é de $2,36 \cdot 10^{-18}$ para a antena SGH, $9,70 \cdot 10^{-20}$ para a antena LL e de $9,70 \cdot 10^{-20}$ para a antena RA e dessa maneira o ruído pode ser considerado centrado.

O próximo ponto a ser analisado é a homocedasticidade do ruído residual. Como mencionado em 3, quanto menor o valor de p para os testes de hipóteses de homocedasticidade o ruído associado é mais semelhante ao branco. Para um nível de significância de 5%, a tabela 5.3 mostra os resultados do teste BPK ([110, 94, 111]), com valores de p da ordem de 10^{-13} para RA, 10^{-8} para SGH e 10^{-1} para a LL. Tais resultados confirmam que a homocedasticidade do ruído residual associado ao processo de reconstrução do campo para as antenas investigadas.

Já a independência das amostras é verificada através da matriz de autocorrelação, que deve ser idealmente diagonal ou predominantemente diagonal, a figura 5.7 mostra os resul-

tados para todas as antenas. Fica bem nítido que as antenas SGH e RA possuem matrizes de correlação praticamente diagonais, ao passo que para a antena LL alguns elementos da matriz que estão próximos à diagonal principal possuem valores significativos, mas ainda sim o seu comportamento é predominantemente diagonal. Tais resultados mostram que as amostras são estatísticas independentes e que a estimação proposta é muito precisa e evidenciam ainda mais como erro residual possui comportamento bem similar ao ruído gaussiano branco.

Para todas as antenas o teste de normalidade de Shapiro-Wilk [85] retornou um valor de p nulo. De fato isso indica que o ruído não segue uma distribuição gaussiana típica ao final do processo de reconstrução. Contudo ao fazer uma análise mais profunda das distribuições geradas presentes na figura 5.6, as distribuições possuem um pequeno gargalo quando comparada com a distribuição gaussiana padrão, que é contrabalançado com um pico maior na distribuição, tornado elas distribuições hiper gaussianas. A explicação mais provável desse comportamento advém do nível de significância requerido para a ser de 5%. Conseqüentemente, os dados obtidos estão um pouco mais próximos mas ainda sim formam uma distribuição normalmente distribuída, além de produzir um pequeno limiar nos filtros de SNR modais devido a alta SNR obtida é mostrada na tabela 5.2.

5.1.5 Comparação Entre as Técnicas Proposta e a de Hansen

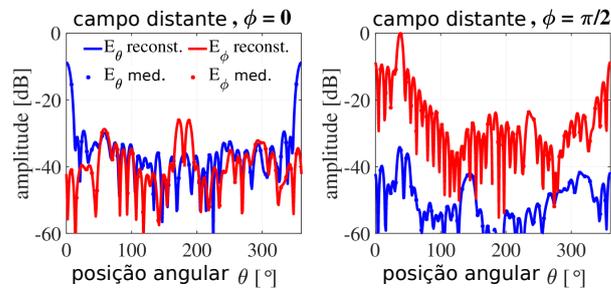
A última etapa da implementação da técnica de Tikhonov-Philips é reconstruir o campo total e isso foi feito para o *reflectarray*. O resultado é mostrado na figura 5.8, que também mostra o campo total normalizado medido e campo reconstruído com a técnica de Hansen. Vale ressaltar que tal antena foi caracterizada em um setor angular limitado.

A partir da análise 5.8 fica nítido que a técnica proposta consegue reconstruir o campo com alta fidelidade e isso fica mais evidenciado pelo baixo erro residual, que é mostrado na figura 5.8-d. Isso é explicado pelo fato da técnica proposta não necessitar que o campo fora de região de interesse seja zerado de maneira artificial. Já a técnica de Hansen por outro lado utiliza da transformada de Fourier para poder recuperar os coeficientes modais esféricos e necessita de tal manobra para situações similares de caracterização. Isso causa o já citado efeito Gibbs nos limites da janela de amostragem, O aparecimento de tal efeito aumenta o erro residual associado ao processo de reconstrução, e isso é ressaltado na figura 5.8-e.

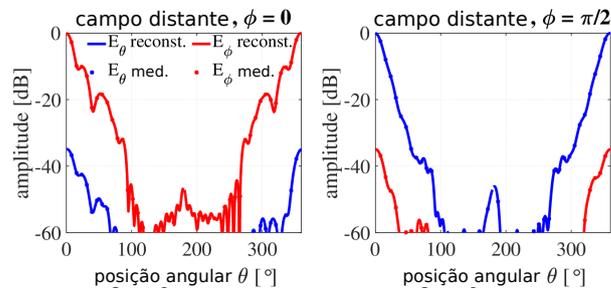
5.2 Reconstrução de Campo com Inteligência Artificial

Quatro geometrias de antenas foram consideradas para validar SL para resolver o problema de recuperação de fase por medições em amplitude: uma antena dipolo operando a 2,4 GHz (DP), uma antena corneta operando a 10 GHz (SGH), uma guia de onda (WG) operando a 10 GHz e uma antena *patch* operando a 10 GHz. Os dados de amplitude de distribuição

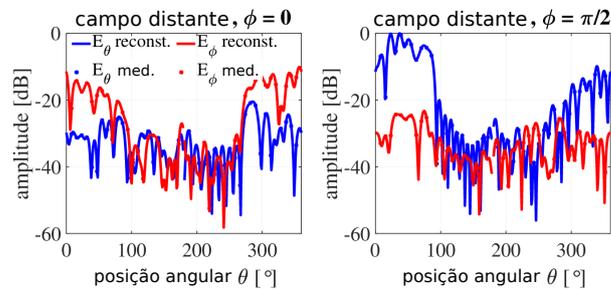
Figura 5.5: Padrão de campo distante para ambas as planos componentes E_θ e E_ϕ do campo elétrico obtido baseado nos parâmetros mostrados na tabela 5.1, para (a) para antena de lente metálica de Luneburg projetada na estrutura de MERLIN para operar a 12 GHz, (b) mostra os resultados para uma antena corneta operando na banda Ka em 26 GHz projetada no IETR e (c) um *reflectarray* composto de 193 células, projetado para operar a 12 GHz pela Thalès Alenia Space na estrutura do projeto R3MEMS. Ambas as antenas atreladas às figuras (b) e (c) foram caracterizadas em um setor angular limitado.



(a) Lente Plana de Luneburg



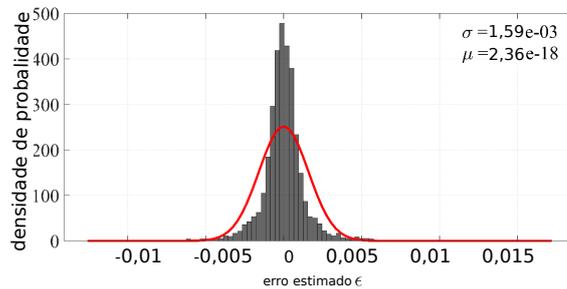
(b) Antena Corneta



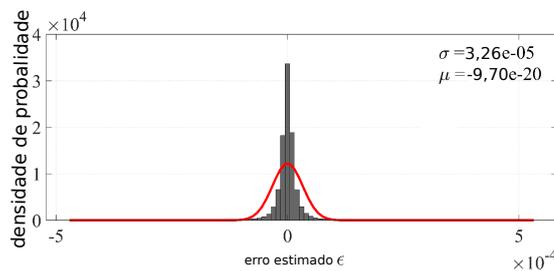
(c) *Reflectarray*

de fase para todas as antenas foram obtidos a partir do *software* Ansys Electronics. Como explicado na parte teórica a superfície de medição escolhida foi a esférica. Adicionalmente foi utilizada a estratégia equiangular explicada no item 2.1.2.1, pois fornece uma quantidade de dados substancial, apesar da sobreamostragem presentes nos polos, mas tal característica a torna mais adequada para a abordagem de solução do PR utilizando modelos de inteligência artificial.

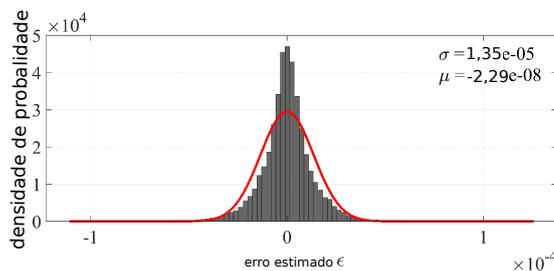
Figura 5.6: Densidade de probabilidade de erro para as três antenas caracterizadas: a figura (a) para antena SGH, a figura (b) para a antena LL e a figura (c) para antena RA. Tais resultados mostram que a interpolação proposta é eficiente, e que o ruído associado já uma distribuição do tipo hiper gaussiana.



(a) Antena Corneta

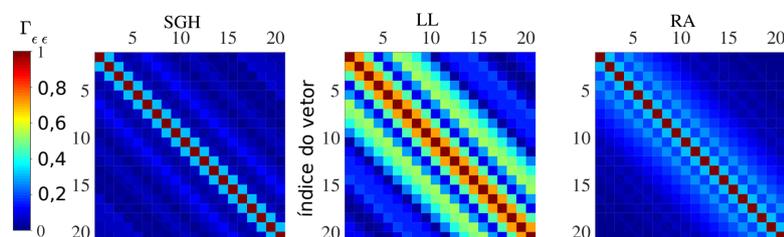


(b) Lente Plana de Luneburg



(c) Reflectarray

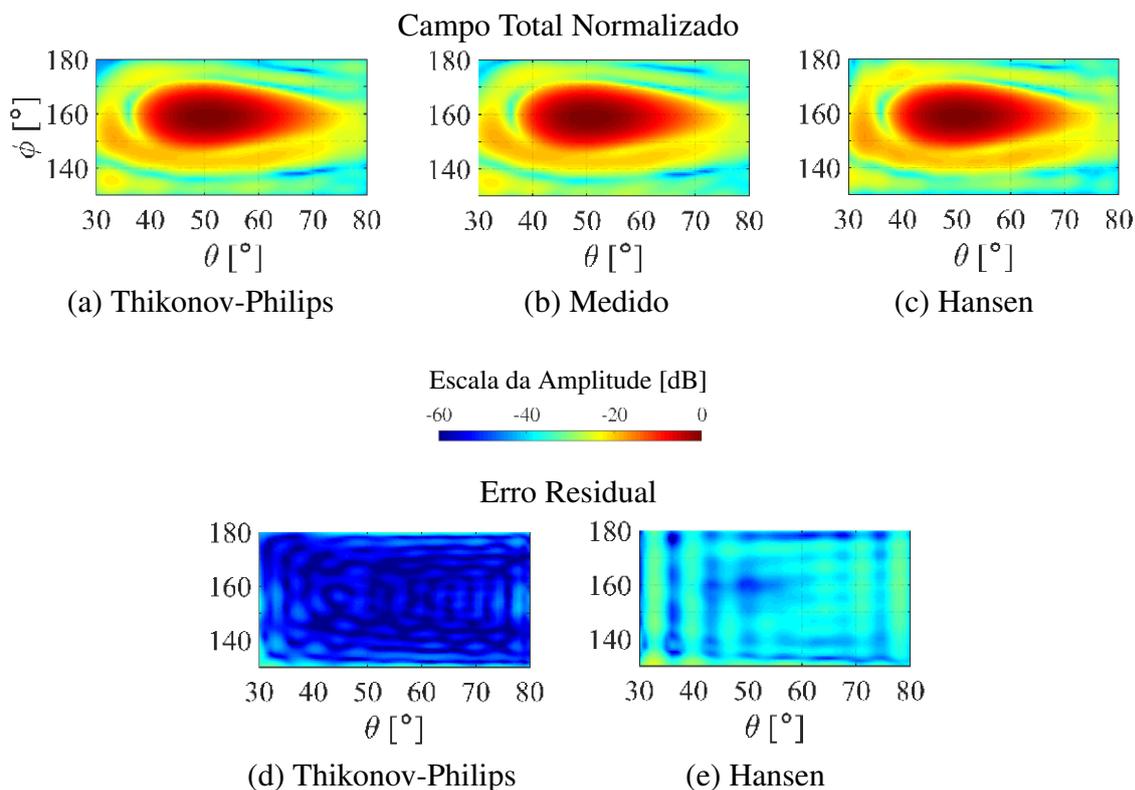
Figura 5.7: Estimação da magnitude de auto correlação, seguindo o cenário estabelecido na tabela 5.2 para as três antenas caracterizadas: a figura (a) para a SGH, a figura (b) para a antena LL e a figura (c) para antena RA. A matriz $\Gamma_{\epsilon\epsilon}$ é predominantemente diagonal para todas as antenas, justificando a independência estatística das amostras do ruído residual.



5.2.1 Antena Guia de Onda

Para essa fonte, a distância de campo distante é de $R_{FF} = 2,6$ m com esfera de raio mínimo de $r = 50$ mm e comprimento de onda $\lambda = 30$ mm. Três esferas de aquisição foram

Figura 5.8: Campo total normalizado para: (a) reconstruído com a técnica proposta, (b) medido e (c) reconstruído com a técnica de Hansen. Tais resultados são para o *reflectarray* projetado na estrutura Thalès Alenia Space no projeto R3MEMS e caracterizado em um setor angular limitado. Devido ao fato da técnica de Hansen ser baseada na transformada de Fourier e necessitar que o campo fora de região interesse seja zerado para obter os coeficientes modais, ela sofre do efeito Gibbs nos limites da janela de amostragem e consequentemente tem um aumento do erro residual. A técnica proposta porém não necessita de tal manobra. Tal comportamento fica evidenciado pelas figuras (d) e (e) que correspondem ao erro residual das técnicas propostas e de Hansen respectivamente.



usadas para construir a base de dados mostrada na figura 4.4, com distâncias de translação em torno do eixo z de $d_1 = 15\%$, $d_2 = 17,5\%$ de $d_3 = 20\%$ de λ . Todas as esferas possuem raio de aquisição r_e de 20% da distância de campo distante. No que diz respeito a amostragem, para cada esfera de aquisição foram produzidas 2701 amostras com taxa de aquisição em θ' de $\Delta\theta' = 5^\circ$ e para ϕ' a taxa foi de $\Delta\phi' = 10^\circ$. Ambas as componentes E_θ e E_ϕ foram validadas.

A validação cruzada foi executada com uma divisão de 10 grupos e 30 por iteração, totalizando um total de 300 testes para cada métrica utilizada para validação que foram: a acurácia, a R^2 , a MAE e a MSE. Os resultados são mostrados na tabela 5.4. Os algoritmos de árvore de decisão, floresta randômica e KNN demonstraram os melhores resultados nas métricas de avaliação usadas. Sobre os testes de hipóteses, a tabela 5.5 mostra os resultados do teste de Shapiro-Wilk com nível de significância de 5% . Os algoritmos de melhor rendimento (DT, RF e KNN) apresentaram distribuições predominantemente normais e prosseguiram para o restante do processo de validação.

Tabela 5.4: Resumo dos resultados do processo CV para as métricas investigadas utilizando 10 grupos e 30 testes por iteração para a antena guia de onda.

Componente	Modelo	Acurácia [%]	R ² [%]	MAE °	MSE [°] ²
E_θ	LR	-0,15	-0,15	90,15	1,12 10 ⁴
	DT	93,65	93,69	5,07	7,13 10 ²
	RF	94,68	94,68	8,30	5,97 10 ²
	KNN	79,84	79,84	16,05	2,26 10 ³
	SVM	-0,45	-0,45	90,03	1,12 10 ⁴
E_ϕ	LR	-0,05	-0,05	90,50	11,30 10 ³
	DT	84,37	84,51	9,42	1,77 10 ³
	RF	89,73	89,74	14,10	1,15 10 ³
	KNN	81,92	81,92	14,90	2,04 10 ³
	SVM	-0,27	-0,27	90,28	11,3 10 ³

Tabela 5.5: Resultados de Shapiro-Wilk para as métricas investigadas para as componentes E_θ and E_ϕ . Nível de significância de 5%.

Componente	Modelo	Acurácia	R ²	MAE	MSE
E_θ	LR	0,16	0,16	0,10	0,44
	DT	0,74	0,68	0,58	0,45
	RF	0,32	0,32	0,20	0,29
	KNN	0,38	0,38	0,19	0,40
	SVM	0,36	0,36	0,004	0,013
E_ϕ	LR	0,97	0,97	0,95	0,95
	DT	0,074	0,15	0,26	0,25
	RF	0,42	0,28	0,84	0,22
	KNN	0,25	0,24	0,26	0,29
	SVM	0,26	0,26	0,73	0,049

Já a tabela 5.6 mostra os resultados do teste de Tukey para os algoritmos de melhor rendimento (DT,RF e KNN). Neste contexto, verdadeiro significa que existe diferença significativa entre o par investigado, e por consequência falso significa o contrário. Como resultado, modelos de inteligência artificial foram gerados para ambas as componentes E_θ e E_ϕ utilizando os três melhores algoritmos .

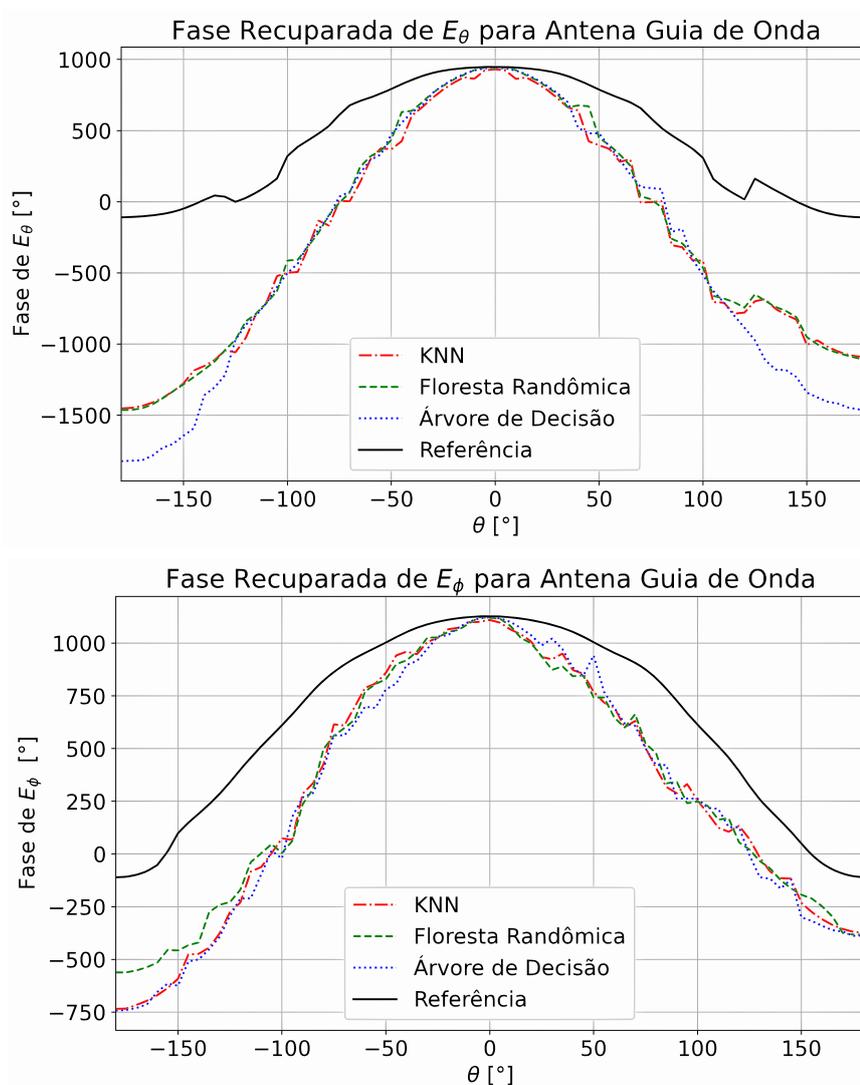
A validação eletromagnética foi executada para uma esfera de aquisição de dados centrada, ou seja, $r_e = r_n$ e que não foi utilizada para treinamento, onde a distribuição de fase para ambas as polarizações E_θ e E_ϕ foram recuperadas para reconstrução do campo elétrico. O algoritmo de árvore de decisão apresentou um nível equivalente de ruído ENL = -12,38 dB, o algoritmo de floresta randômica apresentou um resultado de ENL = -13,35 dB e o algoritmo KNN apresentou ENL = -13,83 dB. A combinação cruzada, ou seja utilizar um modelo gerado para prever a componente E_θ e outro para a E_ϕ , não foi utilizada e sim somente modelos do mesmo algoritmo. A figura 5.9 mostra a fase recuperada para ambas as componentes. Para executar as predições o tempo de predição foi de 2 ms para a árvore de decisão, 130 ms para a floresta randômica e 1,13 s para o KNN. A título de comparação, o algoritmo de de dupla varredura de Gerchberg-Saxton foi implementado usando também a esfera de validação eletromagnética, levando em torno de seis horas

para recuperar a distribuição de fase.

Tabela 5.6: Resumo dos resultados do teste de hipótese de Tukey para os melhores algoritmos para ambas as componentes E_θ e E_ϕ . Nível de significância de 5%.

Componente Par	Acurácia	R ²	MAE	MSE	
E_θ	DT-RF	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro
	DT-KNN	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro
	RF-KNN	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro
E_ϕ	DT-RF	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro
	DT-KNN	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro
	RF-KNN	Verdadeiro	Verdadeiro	Verdadeiro	Verdadeiro

Figura 5.9: Distribuição de fase para esfera centrada com raio de aquisição de 20% de λ para a antena guia de onda operando a 10 GHz. Ambas as componentes E_θ e E_ϕ são mostradas no setor circular em $\phi = 120^\circ$. O algoritmo DT obteve uma de ENL = -12,38 dB, RF de ENL = -13,35 dB e KNN de ENL = -13,83 dB.



5.2.2 Antena Corneta

Para essa fonte, a distância de campo distante é $R_{FF} = 1.16$ m, com raio de esfera mínima $r = 33.02$ mm, e comprimento de onda $\lambda = 30$ mm. Como feito para antena guia de onda, três esferas de aquisição foram usadas com distâncias de translação em relação ao eixo z $d_1 = 15\%$ de λ , $d_2 = 17,5\%$ de λ e $d_3 = 20\%$ de λ . Todas as esferas possuem um raio de aquisição de 20% da distância de campo distante. O número de amostras em cada esfera foi de 2701. A taxa de amostragem no eixo θ' foi de $\Delta\theta' = 5^\circ$, e no eixo ϕ' a taxa foi de $\Delta\phi' = 10^\circ$. Ambas as polarizações E_θ e E_ϕ foram investigadas. Os resultados da validação cruzada nas métricas investigadas (acurácia, R^2 , MAE e MSE) estão dispostos na tabela 5.7 para todos os algoritmos. Para a antena corneta os melhores algoritmos foram a árvore de decisão, floresta randômica e o KNN. A tabela 5.8 mostra os resultados para o teste

Tabela 5.7: Resumo dos resultados para a validação cruzada para as métricas investigadas para ambas as componentes E_θ e E_ϕ . Foram utilizadas 10 divisões e 30 testes por iteração.

Componente	Modelo	Acurácia [%]	R^2 [%]	MAE °	MSE [°] ²
E_θ	LR	3,39	3,39	90,06	$1.03 \cdot 10^4$
	DT	80,80	80,93	12,61	$2,02 \cdot 10^3$
	RF	89,41	89,41	12,69	$1,13 \cdot 10^3$
	KNN	87,50	87,51	11,83	$1,34 \cdot 10^3$
	SVM	12,17	11,80	81,12	$9,46 \cdot 10^3$
E_ϕ	LR	8,16	8,16	85,82	$1,00 \cdot 10^4$
	DT	76,44	76,88	14,88	$2,52 \cdot 10^3$
	RF	83,90	84,10	19,50	$1,73 \cdot 10^3$
	KNN	86,85	86,85	12,01	$1,43 \cdot 10^3$
	SVM	10,37	10,37	80,23	$9,77 \cdot 10^3$

de normalidade de Shapiro-Wilk para as métricas investigadas com nível de significância de 5%. Para a antena corneta, todos os algoritmos apresentaram uma distribuição normal.

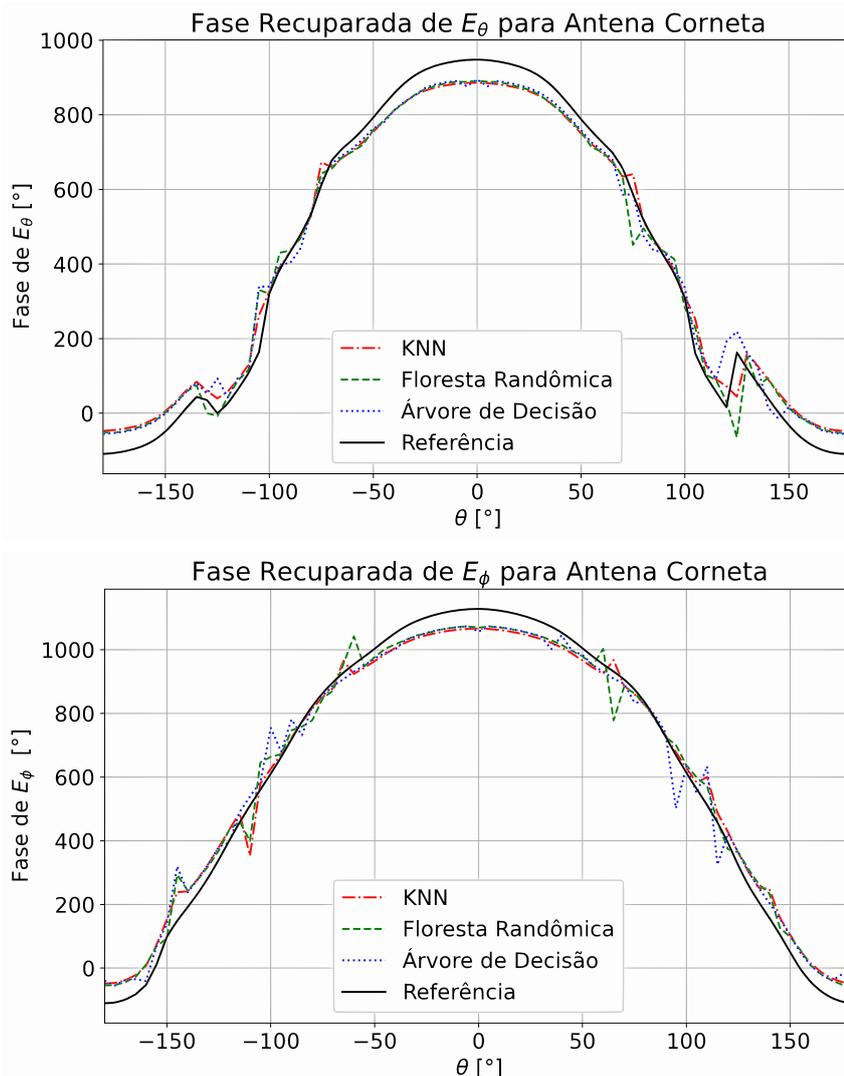
Tabela 5.8: Resultados para o teste de Shapiro-Wilk para as métricas investigadas para ambas as componentes E_θ and E_ϕ . Com nível de significância de 5%.

Componente	Model	Score	R^2	MAE	MSE
E_θ	LR	0,047	0,047	0,97	0,62
	DT	0,54	0,98	0,16	0,06
	RF	0,27	0,27	0,46	0,32
	KNN	0,21	0,21	0,15	0,19
	SVM	0,34	0,34	0,81	0,72
E_ϕ	LR	0,98	0,98	0,14	0,012
	DT	0,17	0,041	0,52	0,50
	RF	0,92	0,95	0,20	0,92
	KNN	0,52	0,56	0,40	0,54
	SVM	0,75	0,75	0,86	0,32

Já a tabela ?? mostra os resultados para o teste de Tukey para os algoritmos de melhor

rendimento, DT, RF e KNN, para ambas as polarizações E_θ e E_ϕ . Nota-se que para o par de comparação DT-RF o teste de Tukey evidenciou que não existe uma diferença significativa para a métrica MSE. Contudo, seguindo o processo e validação estabelecido no item 4.2.3 esses algoritmos não foram descartados. Para a validação eletromagnética, a fase foi recuperada para ambas as componentes E_θ e E_ϕ e o campo reconstruído para uma antena centrada, ou seja, $r_e = r_n$ com raio de aquisição de 20% da distância de campo distante utilizando os modelos gerados a partir dos melhores algoritmos. O algoritmo DT obteve $ENL = -13,76$ dB no processo de reconstrução, o algoritmo RF obteve $ENL = -13,31$ dB e o KNN alcançou o valor $ENL = -13,14$ dB, com tempos de predição de 5 ms para o algoritmo DT, 220 ms para o RF e 1.12 s. A título de comparação o algoritmo de Gerchberg-Saxton também foi implementado para essa antena, levando em torno de seis horas para recuperar a distribuição de fase.

Figura 5.10: Fase recuperada no setor circular $\phi = 120^\circ$ para ambas as componentes E_θ e E_ϕ para a antena corneta operando a 10 GHz com esfera de aquisição centrada com raio de caracterização de 20% da distância de campo distante. O modelo de árvore de decisão alcançou $ENL = -13,76$ dB, o algoritmo de floresta randômica alcançou $ENL = -13,31$ dB e o algoritmo KNN obteve $ENL = -13,14$ dB.

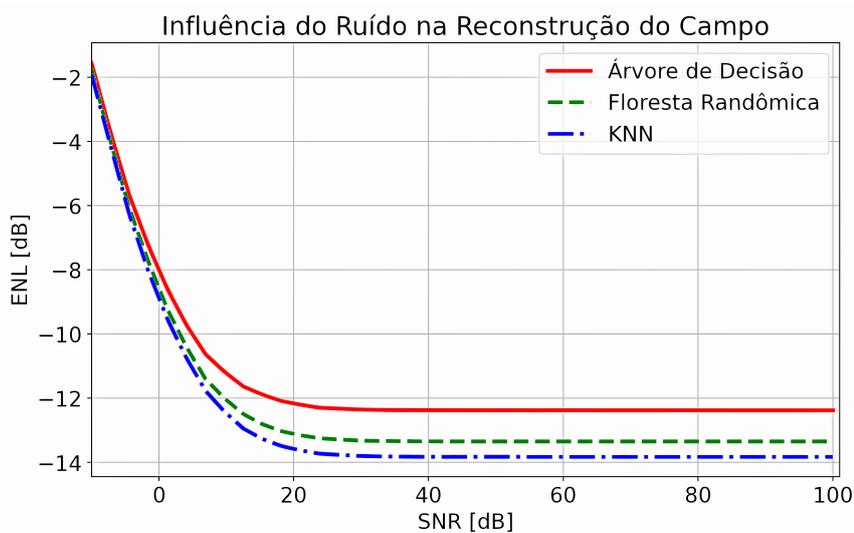


5.2.3 Influência do Ruído

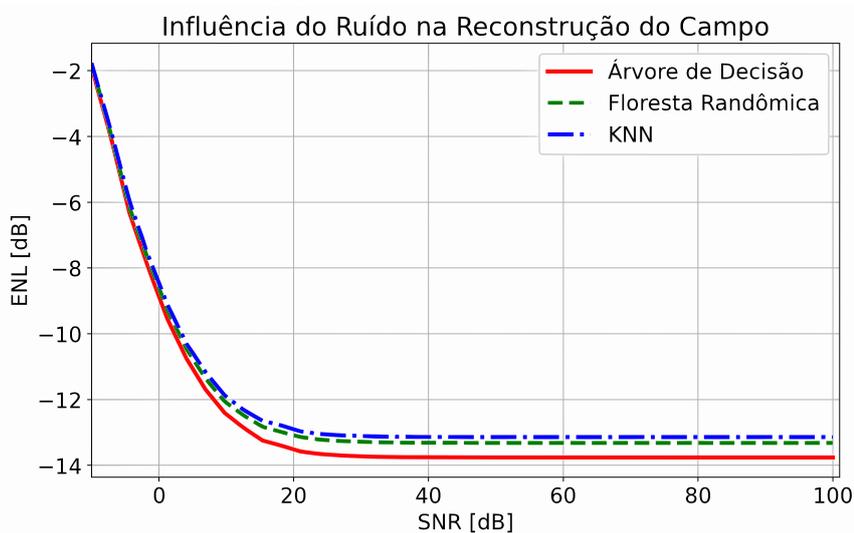
Como discutido no item 2.1 as medições experimentais estão corrompidas com ruído. Dessa maneira, como é feito amplamente na literatura, os modelos obtidos para a previsão da fases das componentes E_θ e E_ϕ foram testados para verificar o impacto do ruído na reconstrução do campo. Dessa maneira, a esfera centrada usada para testar os algoritmos teve a suas medições em amplitude corrompidas com ruído branco. Para executar essa análise, a relação sinal ruído dos dados de amplitude foi variada entre -10 dB a 100 dB. Logo, para todas as antenas estudadas em que os algoritmos foram testados em um cenário em que há uma corrupção considerável dos dados pelo ruído, até aquele próximo do ideal. Para as antenas guia de onda e corneta, o seu desempenho na recuperação de fase do campo cai de maneira vertiginosa para uma SNR abaixo de 0 dB.

Isso é mostrado pela métrica ENL que fornece uma estimativa da qualidade do processo de reconstrução. Para uma SNR de -10 dB a ENL apresenta valores próximos de -2 dB para os algoritmos DT, RF e KNN. Os resultados estão dispostos na figura 5.11. De forma complementar a partir de uma SNR de 5 dB os algoritmos conseguem reconstruir o campo relativamente bem apresentando uma ENL de em torno de -10 dB. Já para $SNR \geq 20$ dB a influência do ruído praticamente não é mais sentida no processo de reconstrução usando os modelos a partir dos algoritmos DT, RF e KNN, convergindo para aos valores apresentados nos itens 5.2.1 e 5.2.2.

Figura 5.11: Influência do ruído na recuperação de fase efetuada pelos algoritmos DT, RF e KNN para antena guia de onda, figura (a), e corneta figura (b) ambas operando a 10 GHz. Os algoritmos se perdem bastante para SNR abaixo de 0 dB, ao passo que para uma SNR acima de 5 dB a reconstrução torna-se satisfatória, aproximando-se dos resultados ideais à medida que ela aumenta.



(a) Antena Guia de Onda



(b) Antena Corneta

Conclusão

Este trabalho apresentou novas abordagens para reconstrução do campo eletromagnético a partir de obtenção de dados experimentais do campo eletromagnético irradiado. Primeiramente, um novo esquema de interpolação baseada na técnica de Tikhonov-Phillips baseada em expansões de ondas esféricas foi apresentado e implementado. Nela, o número de condicionamento foi a métrica utilizada para obter a melhor malha retangular de amostragem.

Análises foram efetuadas para garantir um processo de inversão matricial rápido, estável e estatisticamente robusto para qualquer ordem de truncamento da série modal esférica em que um filtro estatístico que mantém apenas os coeficientes modais esféricos relevantes também foi utilizado. Os resultados obtidos para as famílias de antenas testadas mostram que tal abordagem pode ser aplicada tanto para amostragens parciais, ou seja, em um setor angular limitado, tanto para amostragens em esferas completas reconstruindo o campo eletromagnético de maneira satisfatória. Isso é evidenciado pelo fato de não existir uma diferença significativa entre o campo medido e o reconstruído pela técnica proposta. No que se refere a amostragens utilizando setores angulares limitados, a estratégia proposta mostrou-se eficiente em evitar qualquer efeito de Gibbs indesejados nos limites da janela de interesse, que aparecem quando os dados experimentais de campo fora do setor de interesse são zerados. Esses resultados foram validados a partir da comparação com o campo medido e com o reconstruído com a técnica de Hansen que é bastante suscetível a tal efeito. Logo, a técnica proposta é computacionalmente eficiente, possui um grande impacto e permite conduzir caracterização de antenas com dados faltantes e contaminados de maneira significativa com ruído.

Em segundo lugar, uma nova técnica de reconstrução de campo baseada no problema de recuperação de fase utilizando algoritmos de aprendizado de máquina supervisionada é proposta. Diferente das abordagens convencionais baseadas em dupla varredura e relaxação convexa, a técnica apresentada permite estimar a distribuição de fase do campo eletromagnético para diversas famílias de antenas aplicando algoritmos de aprendizagem de máquina supervisionada.

A base de dados foi obtida a partir de uma nova metodologia de aquisição, em que a esfera de medição é translada em relação ao sistema de coordenadas da antena, sendo composta pelos os dados de amplitude e das posições angulares de amostragem como elementos preditores, enquanto a variável alvo é a distribuição de fase. Os modelos gerados foram

validados a partir de uma análise estatística robusta de validação cruzada, que garante a eficiência dos modelos em prever a fase a partir de novas entradas. Adicionalmente, foram utilizados testes ANOVA e Tukey para obter os algoritmos que geram resultados comprovadamente diferentes entre si para executar futuras previsões. O nível equivalente de ruído ENL do campo reconstruído com a utilização da fase recuperada para as famílias de antenas testadas é comparável com as técnicas clássicas. Os modelos também apresentam um rendimento satisfatório para uma SNR nas medições acima de 5 dB. Adicionalmente, os modelos obtidos possuem a vantagem de serem de fácil uso, portáteis e demandando menor consumo de recursos computacionais para prever a distribuição de fase, aumentando a velocidade de reconstrução em torno de 10^6 após o processo de treinamento, e esse resultado é obtido quando a técnica proposta é comparada com o método de dupla varredura.

Vale destacar que as técnicas propostas abordam a reconstrução de campo de maneira distinta. A primeira técnica é de fato mais robusta e possui resultados muito superiores quando comparada com a segunda, pois consegue reconstruir o campo com um alto grau de similaridade em cenários hostis com dados faltantes e poluídos com ruído, ao passo que a segunda, apesar de apresentar resultados satisfatórios, não consegue atingir tal nível de fidelidade além de ser bem suscetível ao ruído. Logo a técnica de baseada na regularização é recomendada para caracterizações onde é necessário reconstruir o campo com alto grau de fidelidade em ambientes não controlados seja para uma esfera completa ou em um setor angular específico.

Já a técnica de reconstrução de campo utilizando inteligência artificial é recomendada para reconstruções onde somente dados de amplitudes estão disponíveis e onde os recursos computacionais são reduzidos. Vale salientar no entanto que apesar das limitações dessa técnica ela, até onde se sabe, foi a primeira abordagem de resolver o problema de recuperação de fase utilizando algoritmos de aprendizagem de máquina supervisionada.

5.3 Contribuições deste Trabalho

As contribuições deste trabalho foram:

- Implementação de uma técnica que reconstrução baseada na técnica de regularização de Thikonov-Philips aplicada a dados faltosos e ruidosos;
- Primeira abordagem de reconstrução do campo eletromagnético utilizando inteligência artificial;
- Desenvolvimento do *framework Phase Retrieval Model Generator - Artificial Intelligence*-(PRMG - AI), com registro de processo de propriedade intelectual N° BR512023003530-0, no Instituto Nacional de Propriedade Intelectual (INPI) ;
- Aceite na conferência *Workshop on Communication Networks and Power Systems*

(WCNPS) do artigo *Phase Retrieval Based on Supervised Machine Learning Regression Algorithms* no ano de 2023;

- Submissão na revista *IEEE Antennas and Wireless Propagation Letters (AWPL)* do artigo *Spherical Harmonic Expansion of Antenna Far Field in Presence of Noisy and Incomplete Data*, no ano de 2023 (Qualis A1).

5.4 Trabalhos Futuros

Alguns trabalhos futuros podem ser propostos e o mesmos são: (a) utilizar a técnica de interpolação em ambientes não controlados com grandes fontes de ruído e incertezas de medição, (b) obter modelos de inteligência artificial para recuperação dos coeficientes esféricos a partir de medições em amplitude, (c) obter modelos de inteligência artificial capazes de obter as fontes equivalentes e (d) a utilização dos modelos de reconstrução de fase juntamente com o algoritmo de dupla varredura de Gerchberg-Saxton, em que os modelos de inteligência artificial determinaria a distribuição de fase inicial necessária para iniciar o algoritmo. Adicionalmente, algumas atividades e melhorias podem ser implementadas, entre elas:

- Expansão da base de dados das antenas. Considerando a frequência de operação como uma nova informação *a-priori*;
- Investigação do modelos de inteligência artificial para arranjos e antenas de cavidade ressonante, validando a capacidade dos modelos quando expostos a campos com configurações mais complexas;
- Consideração do posicionamento da antena no sistema de coordenadas global e o seu impacto nas predições feitas;
- Utilização de algoritmos de inteligência artificial que conseguem fazer previsões múltiplas e eliminar a necessidade de criar modelos diferentes para as componentes E_θ e E_ϕ , uma abordagem promissora é a *Multi Task CNN* mostrado em [112].

Bibliografia

- [1] LIMA, E. L. *Espaços métricos*. [S.l.]: Instituto de Matemática Pura e Aplicada, CNPq Rio de Janeiro, 1983. v. 4.
- [2] ORFANIDIS, S. J. *Electromagnetic waves and antennas*. Rutgers University New Brunswick, NJ, 2016.
- [3] MÉZIÈRES, N. *Contributions to fast and accurate antenna characterization*. Tese (Doutorado) — Université Rennes 1, 2021.
- [4] HANSEN, J. E. *Spherical near-field antenna measurements*. [S.l.]: Iet, 1988. v. 26.
- [5] DONOHO, D. L. Compressed sensing. *IEEE Transactions on information theory*, IEEE, v. 52, n. 4, p. 1289–1306, 2006.
- [6] CANDÈS, E. J.; WAKIN, M. B. An introduction to compressive sampling. *IEEE signal processing magazine*, IEEE, v. 25, n. 2, p. 21–30, 2008.
- [7] GIORDANENGO, G. et al. Fast antenna testing with reduced near field sampling. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 62, n. 5, p. 2501–2513, 2014.
- [8] HANSEN, W. A new type of expansion in radiation problems. *Physical review*, APS, v. 47, n. 2, p. 139, 1935.
- [9] STRATTON, J. A. *Electromagnetic theory*. [S.l.]: John Wiley & Sons, 2007. v. 33.
- [10] GERCHBERG, R. W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, v. 35, p. 237–246, 1972.
- [11] FIENUP, J. R. Reconstruction of an object from the modulus of its fourier transform. *Optics letters*, Optical Society of America, v. 3, n. 1, p. 27–29, 1978.
- [12] FIENUP, J. R. Phase retrieval algorithms: a comparison. *Applied optics*, Optical Society of America, v. 21, n. 15, p. 2758–2769, 1982.
- [13] CANDES, E. J.; STROHMER, T.; VORONINSKI, V. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, Wiley Online Library, v. 66, n. 8, p. 1241–1274, 2013.

- [14] WALDSPURGER, I.; D'ASPREMONT, A.; MALLAT, S. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, Springer, v. 149, p. 47–81, 2015.
- [15] CHEN, Y. et al. Phaseless characterization of compact antenna test range via improved alternating projection algorithm. *Electronics*, Multidisciplinary Digital Publishing Institute, v. 10, n. 13, p. 1545, 2021.
- [16] FUCHS, B. et al. Fast antenna far-field characterization via sparse spherical harmonic expansion. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 65, n. 10, p. 5503–5510, 2017.
- [17] BOLDRINI, J. L. et al. *Álgebra linear*. [S.l.]: Harper & Row, 1980.
- [18] DONOHO, D. L.; ELAD, M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 100, n. 5, p. 2197–2202, 2003.
- [19] ELAD, M. *Sparse and redundant representations: from theory to applications in signal and image processing*. [S.l.]: Springer, 2010. v. 2.
- [20] FOUCART, S. et al. *An invitation to compressive sensing*. [S.l.]: Springer, 2013.
- [21] MALLAT, S. G.; ZHANG, Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, IEEE, v. 41, n. 12, p. 3397–3415, 1993.
- [22] DAVIS, G.; MALLAT, S.; ZHANG, Z. Adaptive time-frequency approximations with matching pursuits. In: *Wavelet Analysis and Its Applications*. [S.l.]: Elsevier, 1994. v. 5, p. 271–293.
- [23] BLUMENSATH, T.; DAVIES, M. E. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, Elsevier, v. 27, n. 3, p. 265–274, 2009.
- [24] DONOHO, D. L.; TANNER, J. Precise undersampling theorems. *Proceedings of the IEEE*, IEEE, v. 98, n. 6, p. 913–924, 2010.
- [25] BALANIS, C. A. *Antenna theory: analysis and design*. [S.l.]: John wiley & sons, 2016.
- [26] PETRE, P.; SARKAR, T. K. A planar near-field to far-field transformation using an equivalent magnetic current approach. In: IEEE. *IEEE Antennas and Propagation Society International Symposium 1992 Digest*. [S.l.], 1992. p. 1534–1537.
- [27] SADIKU, M. N. *Numerical techniques in electromagnetics*. [S.l.]: CRC press, 2000.
- [28] FUCHS, B. et al. Phaseless near-field antenna measurements from two surface scans—numerical and experimental investigations. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 68, n. 3, p. 2315–2322, 2019.

- [29] ÁLVAREZ, Y.; LAS-HERAS, F.; PINO, M. R. Reconstruction of equivalent currents distribution over arbitrary three-dimensional surfaces based on integral equation algorithms. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 55, n. 12, p. 3460–3468, 2007.
- [30] PARINI, C. et al. *Theory and practice of modern antenna range measurements*. [S.l.]: Institution of Engineering and Technology, 2014.
- [31] GREGSON, S.; MCCORMICK, J.; PARINI, C. *Principles of planar near-field antenna measurements*. [S.l.]: IET, 2007. v. 53.
- [32] PARINI, C.; GREGSON, S.; MCCORMICK, J. Principles of planar near-field antenna measurements. *London, UK: IET*, 2014.
- [33] BREINBJERG, O. Spherical near-field antenna measurements—the most accurate antenna measurement technique. In: IEEE. *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*. [S.l.], 2016. p. 1019–1020.
- [34] BUCCI, O. M.; GENNARELLI, C.; SAVARESE, C. Optimal interpolation of radiated fields over a sphere. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 39, n. 11, p. 1633–1643, 1991.
- [35] SAFF, E. B.; KUIJLAARS, A. B. Distributing many points on a sphere. *The mathematical intelligencer*, Springer US, v. 19, p. 5–11, 1997.
- [36] MINATTI, G. et al. A circularly-polarized isoflux antenna based on anisotropic metasurface. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 60, n. 11, p. 4998–5009, 2012.
- [37] SACCARDI, F. et al. Truncation error mitigation in free-space automotive partial spherical near field measurements. In: IEEE. *2017 Antenna Measurement Techniques Association Symposium (AMTA)*. [S.l.], 2017. p. 1–6.
- [38] HARRISON, R. W. Phase problem in crystallography. *JOSA a*, Optica Publishing Group, v. 10, n. 5, p. 1046–1055, 1993.
- [39] GRIFFIN, D.; LIM, J. Signal estimation from modified short-time fourier transform. *IEEE Transactions on acoustics, speech, and signal processing*, IEEE, v. 32, n. 2, p. 236–243, 1984.
- [40] FUCHS, B.; COQ, L. L. Excitation retrieval of microwave linear arrays from phaseless far-field data. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 63, n. 2, p. 748–754, 2014.
- [41] WALTHER, A. The question of phase retrieval in optics. *Optica Acta: International Journal of Optics*, Taylor & Francis, v. 10, n. 1, p. 41–49, 1963.

- [42] MEZIERES, N.; COQ, L. L.; FUCHS, B. Spherical phaseless antenna measurements experimental validation of a two-antenna-positions procedure. *IEEE Antennas and Wireless Propagation Letters*, IEEE, v. 21, n. 4, p. 813–817, 2022.
- [43] KNAPP, J. et al. Multifrequency phase retrieval for antenna measurements. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 69, n. 1, p. 488–501, 2020.
- [44] VARELA, F. R. et al. Numerical and experimental investigation of phaseless spherical near-field antenna measurements. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 69, n. 12, p. 8830–8841, 2021.
- [45] OGAWA, T.; HASEYAMA, M.; KITAJIMA, H. Reconstruction method of missing texture using error reduction algorithm. In: IEEE. *IEEE International Conference on Image Processing 2005*. [S.l.], 2005. v. 2, p. II–1026.
- [46] GOLDSTEIN, T.; STUDER, C. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, IEEE, v. 64, n. 4, p. 2675–2689, 2018.
- [47] CHEN, S.; DONOHO, D. Basis pursuit. In: IEEE. *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*. [S.l.], 1994. v. 1, p. 41–44.
- [48] CHEN, S. S.; DONOHO, D. L.; SAUNDERS, M. A. Atomic decomposition by basis pursuit. *SIAM review*, SIAM, v. 43, n. 1, p. 129–159, 2001.
- [49] NETRAPALLI, P.; JAIN, P.; SANGHAVI, S. Phase retrieval using alternating minimization. *Advances in Neural Information Processing Systems*, v. 26, 2013.
- [50] MONDELLI, M.; MONTANARI, A. Fundamental limits of weak recovery with applications to phase retrieval. In: PMLR. *Conference On Learning Theory*. [S.l.], 2018. p. 1445–1450.
- [51] LU, Y. M.; LI, G. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, Oxford University Press, v. 9, n. 3, p. 507–541, 2020.
- [52] CHEN, P.; FANNJIANG, A.; LIU, G.-R. Phase retrieval with one or two diffraction patterns by alternating projections with the null initialization. *Journal of Fourier Analysis and Applications*, Springer, v. 24, p. 719–758, 2018.
- [53] WANG, G.; GIANNAKIS, G. B.; ELDAR, Y. C. Solving systems of random quadratic equations via truncated amplitude flow. *IEEE Transactions on Information Theory*, IEEE, v. 64, n. 2, p. 773–794, 2017.
- [54] BENDORY, T.; ELDAR, Y. C.; BOUMAL, N. Non-convex phase retrieval from stft measurements. *IEEE Transactions on Information Theory*, IEEE, v. 64, n. 1, p. 467–484, 2017.

- [55] GUPTA, V. et al. An overview of supervised machine learning algorithm. In: IEEE. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*. [S.l.], 2022. p. 87–92.
- [56] HAHNE, F. et al. Unsupervised machine learning. *Bioconductor case studies*, Springer, p. 137–157, 2008.
- [57] AHMED, M.; SERAJ, R.; ISLAM, S. M. S. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, MDPI, v. 9, n. 8, p. 1295, 2020.
- [58] UPRETY, A.; RAWAT, D. B. Reinforcement learning for iot security: A comprehensive survey. *IEEE Internet of Things Journal*, IEEE, v. 8, n. 11, p. 8693–8706, 2020.
- [59] Kaelbling, L. P.; Littman, M. L.; Moore, A. W. Reinforcement learning: A survey. *Journal of artificial intelligence research*, v. 4, p. 237–285, 1996.
- [60] Vermorel, J.; Mohri, M. Multi-armed bandit algorithms and empirical evaluation. In: Springer. *European conference on machine learning*. [S.l.], 2005. p. 437–448.
- [61] NASTESKI, V. An overview of the supervised machine learning methods. *Horizons. b*, v. 4, p. 51–62, 2017.
- [62] WEISBERG, S. *Applied linear regression*. [S.l.]: John Wiley & Sons, 2005. v. 528.
- [63] Rokach, L.; Maimon, O. Top-down induction of decision trees classifiers—a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, IEEE, v. 35, n. 4, p. 476–487, 2005.
- [64] PAINSKY, A.; ROSSET, S. Cross-validated variable selection in tree-based methods improves predictive performance. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 11, p. 2142–2153, 2016.
- [65] HE, Q.; CHEN, J.-F. The inverse problem of support vector machines and its solution. In: IEEE. *2005 International Conference on Machine Learning and Cybernetics*. [S.l.], 2005. v. 7, p. 4322–4327.
- [66] JOHANNESSEN, N. J.; KOLHE, M. L.; GOODWIN, M. Evaluating anomaly detection algorithms through different grid scenarios using k-nearest neighbor, iforest and local outlier factor. In: IEEE. *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*. [S.l.], 2022. p. 1–6.
- [67] RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986.
- [68] JOSPIN, L. V. et al. Hands-on bayesian neural networks—a tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, IEEE, v. 17, n. 2, p. 29–48, 2022.

- [69] SCIKIT-LEARN. *Scikit-Learn: Machine Learning in Python*. 2011. Disponível em: <<https://www.scikit-learn.org>>.
- [70] PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- [71] EFRON, B. et al. Least angle regression. 2004.
- [72] CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random forests. *Ensemble machine learning: Methods and applications*, Springer, p. 157–175, 2012.
- [73] BREIMAN, L.; FRIEDMAN, J. Ra olshen und cj stone. *Classification and regression trees*, 1984.
- [74] HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2.
- [75] CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, Acm New York, NY, USA, v. 2, n. 3, p. 1–27, 2011.
- [76] BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*. [S.l.: s.n.]. p. 144–152.
- [77] JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. *Computer*, IEEE, v. 29, n. 3, p. 31–44, 1996.
- [78] KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.
- [79] CAMERON, A. C.; WINDMEIJER, F. A. An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of econometrics*, Elsevier, v. 77, n. 2, p. 329–342, 1997.
- [80] KARCH, J. Improving on adjusted r-squared. *Collabra: Psychology*, University of California Press, v. 6, n. 1, 2020.
- [81] TAYLOR, J. R.; THOMPSON, W. *An introduction to error analysis: the study of uncertainties in physical measurements*. [S.l.]: Springer, 1982. v. 2.
- [82] ISO, B. *Accuracy (trueness and precision) of measurement*. 1998. 1994 p.
- [83] ZHANG, D. A coefficient of determination for generalized linear models. *The American Statistician*, Taylor & Francis, v. 71, n. 4, p. 310–316, 2017.

- [84] DÁVILA, V. H. L. Teste de hipóteses. *Instituto de Matemática, Estatística e Computação, UNICAMP*. Disponível em: < https://www.ime.unicamp.br/~hlachos/Inferencia_Hipo1.pdf, p. 3, 2017.
- [85] SHAPIRO, S.; WILK, M. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3-4, p. 591–611, December 1965.
- [86] KIM, T. K. Understanding one-way anova using conceptual figures. *Korean journal of anesthesiology*, The Korean Society of Anesthesiologists, v. 70, n. 1, p. 22–26, 2017.
- [87] TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, JSTOR, p. 99–114, 1949.
- [88] HANUSZ, Z.; TARASIŃSKA, J. Simulation study on improved shapiro–wilk tests for normality. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 43, n. 9, p. 2093–2105, 2014.
- [89] GURVICH, V.; NAUMOVA, M. Logical contradictions in the one-way anova and tukey–kramer multiple comparisons tests with more than two groups of observations. *Symmetry*, MDPI, v. 13, n. 8, p. 1387, 2021.
- [90] NIST Handbook of Mathematical Functions. 2010. Disponível em: <<http://dlmf.nist.gov/>>.
- [91] RONDINEAU, S. *Modélisation de lentilles sphériques à gradient d'indice et sources conformes associées*. Tese (Doutorado) — Rennes 1, 2002.
- [92] BUCCI, O. M.; GENNARELLI, C.; SAVARESE, C. Representation of electromagnetic fields over arbitrary surfaces by a finite and nonredundant number of samples. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 46, n. 3, p. 351–359, 1998.
- [93] BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, JSTOR, p. 1287–1294, 1979.
- [94] COOK, R.; WEISBERG, S. Diagnostics for heteroskedasticity in regression. *Biometrika*, v. 70, n. 1, p. 1–10, 1983.
- [95] KOENKER, R. A note on studentizing a test for heteroscedasticity. *Journal of econometrics*, Elsevier, v. 17, n. 1, p. 107–112, 1981.
- [96] WOOLDRIDGE, J. M. *Introductory econometrics: A modern approach*. [S.l.]: Cengage learning, 2015.
- [97] GREENE, W. H. *Econometric analysis*. [S.l.]: Pearson Education India, 2003.
- [98] KENNEDY, P. *A guide to econometrics*. [S.l.]: John Wiley & Sons, 2008.

- [99] RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, v. 2, n. 1, p. 21–33, 2011.
- [100] COX, D. R.; SMALL, N. Testing multivariate normality. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 263–272, 1978.
- [101] FAN, Y. Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis*, Elsevier, v. 62, n. 1, p. 36–63, 1997.
- [102] JR, F. J. M. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, Taylor & Francis, v. 46, n. 253, p. 68–78, 1951.
- [103] MILLER, L. H. Table of percentage points of kolmogorov statistics. *Journal of the American Statistical Association*, Taylor & Francis, v. 51, n. 273, p. 111–121, 1956.
- [104] MARSAGLIA, G.; TSANG, W. W.; WANG, J. Evaluating kolmogorov's distribution. *Journal of statistical software*, v. 8, p. 1–4, 2003.
- [105] SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, Institute of Mathematical Statistics, v. 19, n. 2, p. 279–281, 1948.
- [106] COCHRAN, W. G. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.l.], 1934. v. 30, n. 2, p. 178–191.
- [107] ÁLVAREZ, J. F.; MATTES, M.; BREINBJERG, O. Phase retrieval for spherical near-field measurements using two antenna positions. In: IEEE. *2021 Antenna Measurement Techniques Association Symposium (AMTA)*. [S.l.], 2021. p. 1–6.
- [108] GRIFFITHS, D. J. *Introduction to electrodynamics*. [S.l.]: American Association of Physics Teachers, 2005.
- [109] DIALLO, C. et al. All-metal ku-band luneburg lens antenna based on variable parallel plate spacing fakir bed of nails. In: IEEE. *2017 11th European Conference on Antennas and Propagation (EUCAP)*. [S.l.], 2017. p. 1401–1404.
- [110] BREUSCH, T.; PAGAN, A. A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, v. 47, n. 5, p. 1287–1294, 1979.
- [111] R.KOENKER. A note on studentizing a test for heteroskedasticity. *J. of Economet.*, v. 17, n. 1, p. 107–112, 1981.
- [112] ESGARIO, J. G.; KROHLING, R. A.; VENTURA, J. A. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*, Elsevier, v. 169, p. 105162, 2020.

[113] JENSEN, F.; FRANDSEN, A. On the number of modes in spherical wave expansions.
Proc. 26th AMTA, v. 2, n. 1, p. 489–494, 2004.

Apêndices

5.5 Apêndice A: Eletromagnetismo para Elementos Radiantes

Antenas são elementos irradiam ondas eletromagnéticas a partir da aplicação de uma distribuição de correntes e cargas em si. Essas ondas podem propagar-se para longas distâncias [2], e por essa característica tais componentes são extremamente importantes aplicações de telecomunicações. Nesta seção serão abordados os conceitos de eletromagnetismo para tais elementos a partir de duas abordagens: a clássica baseada em campos gerados pelos potenciais elétrico e magnético e por meio expansões em modos esféricos.

5.5.1 Equações de Maxwell

As antenas são elementos que conseguem traduzir a aplicação de uma fonte de correntes e cargas em ondas eletromagnéticas, e por esse motivo elas são totalmente explicadas pelas equações de Maxwell [2, 9, 25]:

$$\begin{aligned}\nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \partial_t \mathbf{D} \\ \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0.\end{aligned}\tag{5.1}$$

Em que a primeira equação é a lei de Faraday da indução, a segunda a lei de Ampère, enquanto a terceira e a quarta são as leis de Gauss para os campos elétricos e magnéticos [2]. As grandezas \mathbf{E} e \mathbf{H} correspondem respectivamente aos campos elétrico e magnético, enquanto \mathbf{D} e \mathbf{B} aos fluxos elétrico e magnético. Já \mathbf{J} corresponde ao fluxo de corrente e ρ a densidade volumétrica de carga [2]. As equações de Maxwell são o ponto de partida para a compreensão do fenômeno físico da radiação de ondas eletromagnéticas tanto pelos potenciais elétricos e magnéticos e modos esféricos.

5.5.2 Campos Radiantes Através de Potenciais

Devido ao fato de os campos eletromagnéticos serem gerados a partir da distribuição de corrente contida na antena Isso permite, sem perda de generalidade, usar funções potenciais em vez do campo elétrico \mathbf{E} e magnético \mathbf{H} , isso é mostrado pela simetria das leis de Faraday e Gauss, que utilizando as funções potenciais podem ser escritas como:

$$\begin{aligned}\mathbf{E} &= -\nabla V - \partial_t \mathbf{A} \\ \mathbf{B} &= \nabla \times \mathbf{A},\end{aligned}\tag{5.2}$$

em que \mathbf{A} e V são os potenciais elétricos e magnéticos respectivamente. Os potenciais \mathbf{A} e V não possuem uma definição única e podem ser alterados. Para manter a invariância das equações de Maxwell é necessário utilizar uma padronização [2]. Considerando uma função escalar dependente da posição r e do tempo t , ou seja $f(\mathbf{r}, t)$, a:

$$\begin{aligned}V' &= V + \partial_t f \\ \mathbf{A}' &= \mathbf{A} \nabla f,\end{aligned}\tag{5.3}$$

garante a invariância necessária. A definição 5.3 permite definir relações convenientes entre os potenciais. Para a radiação o mais comum é utilizar a o calibre de Lorenz [2]:

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \partial_t V = 0,\tag{5.4}$$

em que c e a velocidade da luz no meio de propagação. Dessa maneira, caso os potenciais não satisfaçam a equação de calibre, as suas versões padronizadas \mathbf{A}' e V' podem ser construídas para contornar tal situação utilizando uma função f apropriada que satisfaça a equação de onda não homogênea:

$$\frac{1}{c^2} \frac{\partial^2 f}{\partial t^2} - \nabla^2 f = \nabla \cdot \mathbf{A} + \frac{1}{c^2} \partial_t V,\tag{5.5}$$

que substituída nas leis de Ampère e de Gauss para o magnetismo resulta na equação de onda equivalente para os potenciais [2]:

$$\begin{aligned}\frac{1}{c^2} \frac{\partial^2 V}{\partial t^2} - \nabla^2 V &= \frac{\rho}{\epsilon}, \text{ e} \\ \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} &= \mu \mathbf{J},\end{aligned}\tag{5.6}$$

em que ϵ e μ são respectivamente a permeabilidade elétrica e a permissividade magnética. Tais equações de onda são compatíveis com o calibre de Lorenz mostrada na equação 5.4 da seguinte maneira:

$$\nabla \cdot \mathbf{J} + \partial_t \rho = 0.\tag{5.7}$$

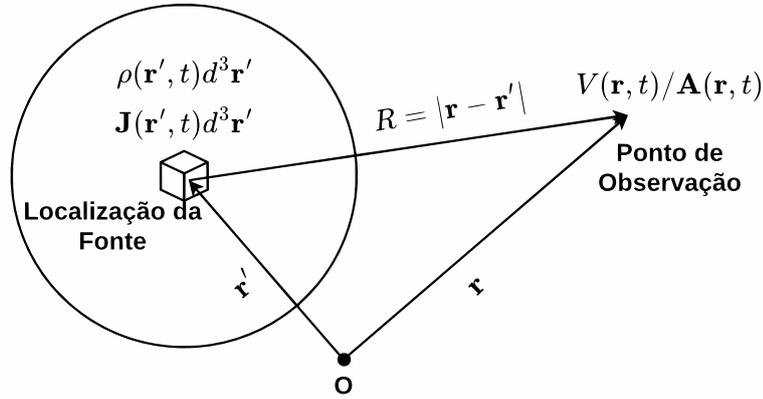
Logo a formulação pelos potências levam a conclusão que caso as densidade de carga ρ

e densidade de corrente \mathbf{J} são conhecidas as soluções causais das equações de onda são:

$$\begin{aligned} V(\mathbf{r}, t) &= \int_V \frac{\rho(\mathbf{r}', t - R/c)}{4\pi\epsilon R} d^3\mathbf{r}', \text{ e} \\ \mathbf{A}(\mathbf{r}, t) &= \int_V \mu \frac{\mathbf{J}(\mathbf{r}', t - R/c)}{4\pi R} d^3\mathbf{r}', \end{aligned} \quad (5.8)$$

em que $R = |\mathbf{r} - \mathbf{r}'|$, em \mathbf{r} denota a posição do ponto de observação e \mathbf{r}' o ponto de localização das fontes, isso é representado na figura 5.12. Logo, os potenciais são obtidos por sobreposição dos campos devido às fontes dentro do volume infinitesimal de volume $d^3\mathbf{r}'$ em um instante $t' = t - R/c$. Isso explica o fato do efeito causado pelas fontes no ponto \mathbf{r}' não ser instantâneo, mas sim sentido de maneira atrasada levando um tempo R/c para chegar até o ponto de observação, ou seja, o efeito de propagação é um efeito causal [2]. Geralmente, o

Figura 5.12: Representação do efeito não instantâneo da aplicação das fontes ρ e \mathbf{A} em uma determinada localização do espaço, demonstrando que a propagação das ondas eletromagnéticas é um efeito causal. Adaptado de [2]



maior interesse é em elementos que operam em uma única frequência. Logo, considerando regime harmônico as densidades de carga e de corrente são:

$$\begin{aligned} \rho(\mathbf{r}, t) &= \rho(\mathbf{r})e^{j\omega t}, \text{ e} \\ \mathbf{J}(\mathbf{r}, t) &= \mathbf{J}(\mathbf{r})e^{j\omega t}. \end{aligned} \quad (5.9)$$

Substituindo as definições feitas na equação 5.9 naquelas mostradas na equação 5.8 obtém-se:

$$\begin{aligned} V(\mathbf{r}, t) &= \int_V \frac{\rho(\mathbf{r}') e^{jkR}}{4\pi\epsilon R} d^3\mathbf{r}', \text{ e} \\ \mathbf{A}(\mathbf{r}, t) &= \int_V \mu \frac{\mathbf{J}(\mathbf{r}') e^{jkR}}{4\pi R} d^3\mathbf{r}', \end{aligned} \quad (5.10)$$

em que $k = \omega/c$. Considerando o calibre de Lorenz, o campo elétrico pode ser expresso

totalmente em função de \mathbf{A} [2]:

$$\mathbf{E} = \frac{1}{j\omega\mu\epsilon} [\nabla \times (\nabla \times \mathbf{A}) - \mu\mathbf{J}]. \quad (5.11)$$

Que é obtida pela resolução da lei de Ampère, considerando o espaço livre, a equação 5.11 assume a forma:

$$\mathbf{E} = \frac{1}{j\omega\mu\epsilon} \nabla \times (\nabla \times \mathbf{A}). \quad (5.12)$$

A partir dessas considerações é possível obter ambos os campos \mathbf{E} e \mathbf{H} em função da distribuição de corrente \mathbf{J} , em:

$$\begin{aligned} \mathbf{E} &= \frac{1}{j\omega\epsilon} \int_V [(\nabla' \mathbf{J}) \nabla' G - k^2 \mathbf{J} G] dV', \text{ e} \\ \mathbf{H} &= \int_V \mathbf{J} \times \nabla' G dV'. \end{aligned} \quad (5.13)$$

Em que $d^3\mathbf{r}$ foi substituído por dV' , e G é a função de Green dependente de $\mathbf{r} - \mathbf{r}'$, com definição [2]:

$$G(R) = \frac{e^{jkR}}{4\pi R}. \quad (5.14)$$

A partir disso, fica claro que os campos radiados podem ser obtidos a partir da transformação de Fourier da distribuição de corrente \mathbf{J} .

5.5.2.1 Aproximação de Campo Distante a Partir dos Potenciais

No estudo dos campos emanados por antenas o maior interesse de estudo é a onda observada a longas distâncias, geralmente denominado como campo distante [2, 25], como ilustrado na figura 5.13. Para atingir tal objetivo primeiramente é considerada uma aproximação para R do tipo:

$$R = r - r' \cos \psi - \frac{r'^2}{2r} \sin^2 \psi. \quad (5.15)$$

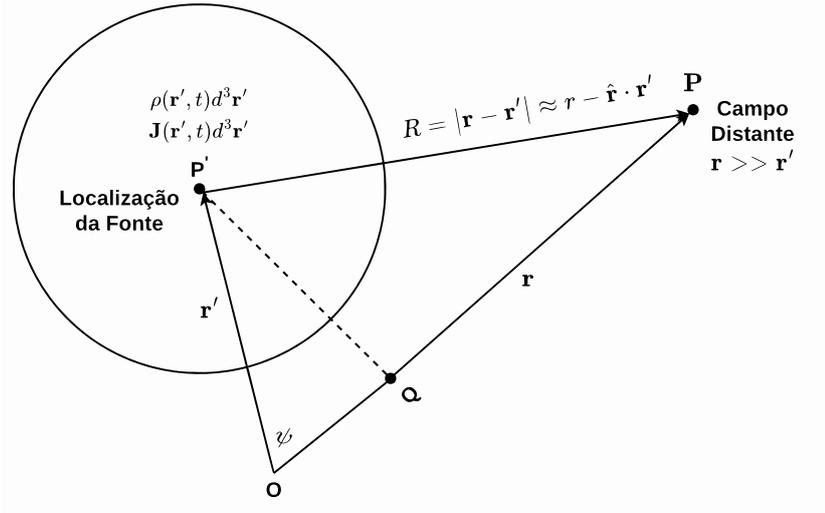
Para $r \gg r'$, ou seja, a distância de observação deve estar muito distante da posição da fonte r' . Tipicamente a distância de campo distante r segue a definição de Fraunhofer [2, 25, 4], que é ilustrada na figura 5.14:

$$r = \frac{2L^2}{\lambda}, \quad (5.16)$$

em que L é a maior dimensão da antena e λ o comprimento de onda. Considerando apenas os termos de primeira ordem $R = r - r' \cos \psi = r - \hat{\mathbf{r}} \cdot \mathbf{r}'$ e substituindo o resultado na equação 5.10 os potenciais tornam-se:

$$\begin{aligned} V(\mathbf{r}) &= \frac{e^{jkr}}{4\pi\epsilon r} \int_V \rho(\mathbf{r}') e^{j\mathbf{k}\mathbf{r}'} d^3\mathbf{r}', \text{ e} \\ \mathbf{A}(\mathbf{r}) &= \frac{\mu e^{jkr}}{4\pi r} \int_V \mathbf{J}(\mathbf{r}') e^{j\mathbf{k}\mathbf{r}'} d^3\mathbf{r}'. \end{aligned} \quad (5.17)$$

Figura 5.13: Aproximação em campo distante de \mathbf{E} e \mathbf{H} é descrita pelas funções potenciais V e \mathbf{A} . Geralmente somente o primeiro termo da expansão de R é suficiente para tal análise.



Considerando que [2]:

$$\begin{aligned} Q(\mathbf{k}) &= \int_V \rho(\mathbf{r}') e^{j\mathbf{k}\mathbf{r}'} d^3\mathbf{r}' \\ \mathbf{F}(\mathbf{k}) &= \int_V \mathbf{J}(\mathbf{r}') e^{j\mathbf{k}\mathbf{r}'} d^3\mathbf{r}' \end{aligned} \quad (5.18)$$

são os vetores de radiação, em que V e A são:

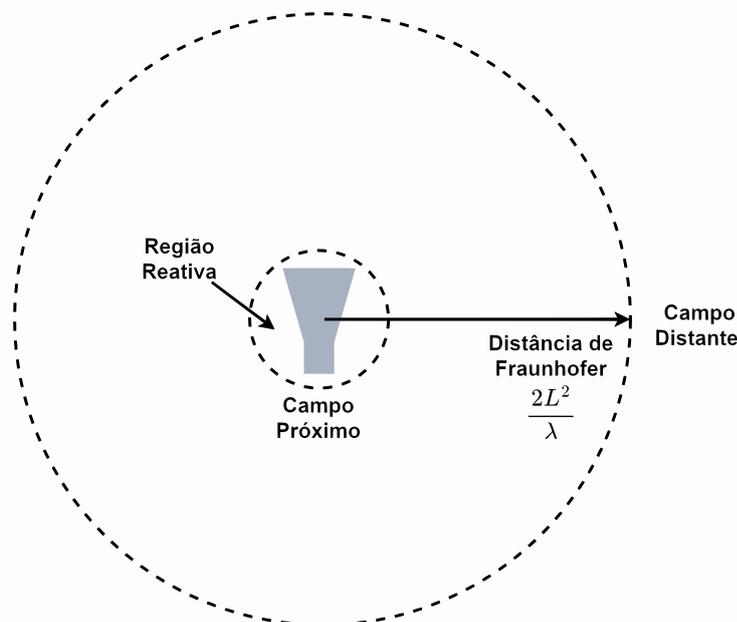
$$\begin{aligned} V(\mathbf{r}) &= \frac{e^{jkr}}{4\pi\epsilon r} Q(\theta, \phi), \text{ e} \\ \mathbf{A}(\mathbf{r}) &= \frac{\mu e^{jkr}}{4\pi r} \mathbf{F}(\theta, \phi). \end{aligned} \quad (5.19)$$

A substituição de \mathbf{k} na transição entre as equações 5.18 e 5.19 é devido ao fato de ω e k serem dependentes do vetor unitário $\hat{\mathbf{r}}$ que é completamente definido pelos ângulos das coordenadas θ e ϕ [2]. Substituindo tal resultado nas definições feitas na equação 5.2 e utilizando a relação mostrada na equação 5.11 os campos pela aproximação de campo distante podem ser escritos como:

$$\begin{aligned} \mathbf{E} &= jk\eta \frac{e^{jkr}}{4\pi r} (\hat{\mathbf{r}} \times \mathbf{F}) \times \hat{\mathbf{r}}, \text{ e} \\ \mathbf{H} &= -jk \frac{e^{jkr}}{4\pi r} \hat{\mathbf{r}} \times \mathbf{F}, \end{aligned} \quad (5.20)$$

em que $\eta = \sqrt{\mu/\epsilon}$ é a impedância do meio. Tais considerações deixam mais evidente que através do uso dos potenciais V e \mathbf{A} os campos radiados por uma antena pode ser obtido como a transformada de Fourier espacial das fontes de densidade de corrente \mathbf{J} e ρ , em que a determinação de tais fontes é de extrema importância para o conhecimento das características intrínsecas do elemento radiante como a potência emanada e seu diagrama de radiação.

Figura 5.14: A aplicação da aproximação de campo distante depende do ponto de observação, uma convenção muito utilizada é a distância de Fraunhofer em $r = 2L^2/\lambda$, em que L é a maior dimensão da antena.



5.5.3 Campos Radiantes Através de Expansões de Ondas Esféricas

Outra maneira de obtenção dos campos radiados é através da obtenção da expansão das ondas esféricas [4, 9], do inglês *Spherical Wave Expansions* (SWE). O processo da SWE também começa pelas equações de Maxwell mostradas em 5.1, mais precisamente as pelas leis de Faraday e Ampère em um meio homogêneo, isotrópico, linear e livre de fontes, ou seja, $\mathbf{J} = 0$ e $\rho = 0$. Utilizando a dependência harmônica, tais equações podem ser escritas como:

$$\begin{aligned}\nabla \times \mathbf{E} &= -j\omega\mu\mathbf{H}, \text{ e} \\ \nabla \times \mathbf{H} &= j\omega\epsilon\mathbf{E}.\end{aligned}\tag{5.21}$$

Considerando que a ausência de fontes leva a $\nabla \cdot \mathbf{D} = 0$, e aplicando o rotacional em ambos os lados das ambas equações mostradas em 5.21, elas assumem a forma de uma equação de onda de Helmholtz do tipo $\nabla^2 \mathbf{C} + k^2 \mathbf{C} = 0$, em que \mathbf{C} pode denotar tanto o campo elétrico \mathbf{E} quanto o campo magnético \mathbf{H} . Nesse sentido, as equações estão descritas em um sistema de coordenadas esférico [4].

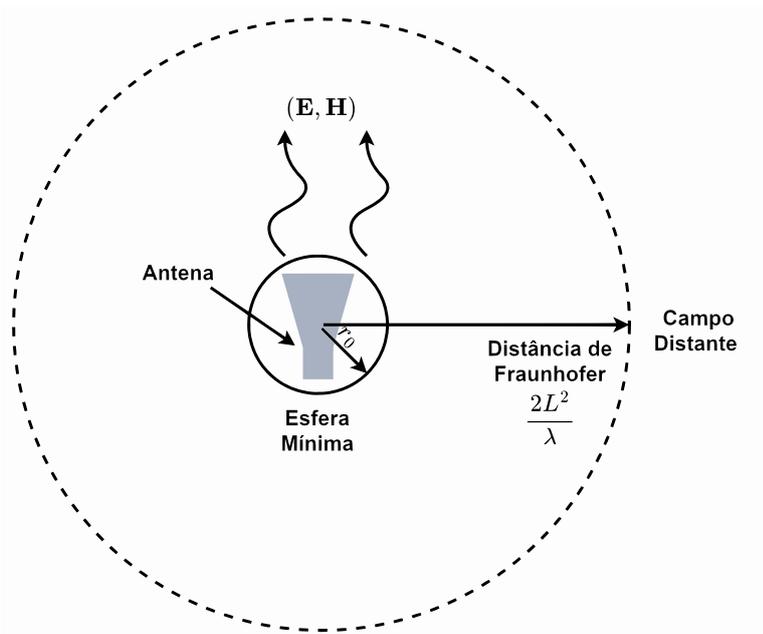
O próximo passo é definir uma função escalar $\Psi = \Psi(r, \theta, \phi)$ que seja a solução da equação de onda $\nabla^2 \Psi + k^2 \Psi = 0$ e que $r\hat{\mathbf{r}}$ é o raio da esfera de propagação. Em seguida, são definidas duas funções vetoriais $\mathbf{m} = \nabla \Psi \times \mathbf{r}$ and $\mathbf{n} = k^{-1} \nabla \times \mathbf{m}$, que satisfaçam a equação de Helmholtz e representem o campo eletromagnético no meio homogêneo. A função escalar $\Psi(r, \theta, \phi)$ pode ser então obtida pelo método de separação de variáveis [4, 91, 9], resultando

em:

$$\begin{aligned} \Psi(r, \theta, \phi)_{mn}^{(c)} &= Z_n^{(c)}(kr) P_n^m(\cos \theta) \cos m\phi, & \text{simetria par, e} \\ \Psi(r, \theta, \phi)_{mn}^{(c)} &= Z_n^{(c)}(kr) P_n^m(\cos \theta) \sin m\phi, & \text{simetria ímpar.} \end{aligned} \quad (5.22)$$

Em que $n = 1, 2, \dots$ and $m = 0, 1, \dots, n$, $P_n^m(\cos \theta)$ é a função de Legendre de ordem n e grau m e com dependência em θ . Já o termo $Z_n^{(c)}$ podem ser quatro funções: função esférica de Bessel, função esférica de Neumann, função de Hankel de primeiro tipo e função de Henkel de segundo tipo, em que o valor de c varia de um à quatro e selecionando o tipo de função a ser usada. A função escalar Ψ mostrada na equação 5.22 possui a propriedade de ser expressa com uma expansão infinita de funções trigonométricas proporcionais a potências inversas de k . Algo a ser enfatizado que o correto valor de c permite escolher um conjunto de funções ortogonais que podem ser usadas como função de base para todas as soluções com valores únicos e contínuos em três regiões de propagação: $c = 1, 2$ (funções de Bessel e Neumann) para ondas estacionárias, $c = 3$ (Hankel de primeiro tipo) para ondas emanadas pela antena e $c = 4$ (Hankel de segundo tipo) para ondas recebidas pela antena. Sendo assim, é possível

Figura 5.15: Os resultados da expansão das ondas esféricas implica que tanto campo elétrico \mathbf{E} quanto magnético \mathbf{H} são emanados a partir de uma esfera raio r_0 que engloba toda a antena independente de sua posição, tal esfera é chamada de esfera mínima [4].



obter as soluções da equação de onda vetorial, substituindo Ψ definida pela equação 5.22 nas

definições de \mathbf{m} e \mathbf{n} , resultando em:

$$\begin{aligned} \mathbf{m}_{mn}^{(c)}(r, \theta, \phi) &= \mp Z_n^{(c)}(kr) \frac{m P_n^m(\cos \theta)}{\sin \theta} \frac{\cos}{\sin} m\phi \hat{\boldsymbol{\theta}} - Z_n^{(c)}(kr) \frac{dP_n^m(\cos \theta)}{d\theta} \frac{\cos}{\sin} m\phi \hat{\boldsymbol{\phi}}, \text{ e} \\ \mathbf{n}_{mn}^{(c)} &= \frac{n(n+1)}{kr} Z_n^{(c)} P_n^m(\cos \theta) \frac{\cos}{\sin} m\phi \hat{\mathbf{r}} + \frac{1}{kr} \frac{d}{d(kr)} \{kr Z_n^{(c)}(kr)\} \frac{dP_n^m(\cos \theta)}{d\theta} \dots \\ &\dots \frac{\cos}{\sin} m\phi \hat{\boldsymbol{\theta}} \mp \frac{1}{kr} \frac{d}{d(kr)} \{kr Z_n^{(c)}(kr)\} \frac{m P_n^m(\cos \theta)}{\sin \theta} \frac{\cos}{\sin} m\phi \hat{\boldsymbol{\phi}}, \end{aligned} \quad (5.23)$$

em que os termos \cos / \sin denotam a possibilidade de simetria ímpar ou par [4]. Para as regiões $c = 3, 4$ é necessário o uso de formas assintóticas de $Z_n^{(c)}$, que são obtidas fazendo $kr \rightarrow \infty$ e são extremamente necessárias para conexão das definições de campo distante. Uma notação mais compacta e amplamente utilizada foi criada por [4], introduzindo o vetor \mathbf{F}_{smn} , em que s denota o vetor \mathbf{m} caso seja igual a um ou vetor \mathbf{n} caso seja igual a dois. Dessa maneira, o vetor \mathbf{F}_{1mn} pode ser escrito como:

$$\begin{aligned} \mathbf{F}_{1mn}^{(c)}(r, \theta, \phi) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m z_n^{(c)}(kr) \frac{j \bar{P}_n^{|m|}(\cos \theta)}{\sin \theta} e^{jm\theta} \hat{\boldsymbol{\theta}} - \dots \\ &\dots - \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m \frac{d\bar{P}(\cos \theta)}{d\theta} e^{jm\theta} \hat{\boldsymbol{\phi}}. \end{aligned} \quad (5.24)$$

Enquanto o vetor \mathbf{F}_{2mn} assume a forma:

$$\begin{aligned} \mathbf{F}_{2mn}^{(c)} &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m \frac{n(n+1)}{kr} z_n^{(c)}(kr) e^{jm\theta} \bar{P}_n^{|m|}(\cos \theta) e^{jm\theta} \hat{\mathbf{r}} + \dots \\ &\dots + \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m \frac{1}{kr} \frac{d}{d(kr)} [kr z_n^{(c)}(kr)] \frac{d\bar{P}_n^{|m|}(\cos \theta)}{d\theta} e^{jm\theta} \hat{\boldsymbol{\theta}} + \dots \\ &\dots + \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m \frac{1}{kr} \frac{d}{d(kr)} [kr z_n^{(c)}(kr)] \frac{j m \bar{P}_n^{|m|}(\cos \theta)}{\sin \theta} e^{jm\theta} \hat{\boldsymbol{\phi}}, \end{aligned} \quad (5.25)$$

em que $\bar{P}_n^{|m|}(\cos \theta)$ é a função de Legendre normalizada e a expressão $(-m/|m|)^m$ garante que a fase dos modos siga a fase dos harmônicos esféricos. Por definição, $(-m/|m|)^m = 1$ para $m = 0$ [4]. De forma complementar, [4] usa uma função escalar diferente:

$$F_{mn}^{(c)}(r, \theta, \phi) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|}\right)^m z_n^{(c)}(kr) \bar{P}_n^{|m|}(\cos \theta) e^{jm\theta}, \quad (5.26)$$

que é necessária para a notação compactada. Dessa maneira, considerando que os campos

elétrico e magnético podem ser escritos na forma [9]:

$$\begin{aligned}\mathbf{E} &= -\sum_n a_n \mathbf{m}_n + b_n \mathbf{n}_n, \text{ e} \\ \mathbf{H} &= \frac{-k}{j\omega\mu} \sum_n a_n \mathbf{n} + b_n \mathbf{m}.\end{aligned}\tag{5.27}$$

Usando a notação [4], os campos podem ser expressos como uma série ponderada infinita, representada como:

$$\begin{aligned}\mathbf{E}(r, \theta, \phi) &= \sqrt{\frac{k}{\eta}} \sum_{csmn} Q_{smn}^{(c)} \mathbf{F}_{smn}^{(c)}(r, \theta, \phi), \text{ e} \\ \mathbf{H}(r, \theta, \phi) &= -jk\sqrt{\eta} \sum_{csmn} Q_{smn}^{(c)} \mathbf{F}_{smn}^{(c)}(r, \theta, \phi),\end{aligned}\tag{5.28}$$

ou evidenciando o somatório quádruplo omitido na equação 5.28:

$$\begin{aligned}\mathbf{E}(r, \theta, \phi) &= \sqrt{\frac{k}{\eta}} \sum_{c=3}^4 \sum_{s=1}^2 \sum_{m=-n}^n \sum_{n=1}^{\infty} Q_{smn}^{(c)} \vec{\mathbf{F}}_{smn}^{(c)}(r, \theta, \phi), \text{ e} \\ \mathbf{H}(r, \theta, \phi) &= -jk\sqrt{\eta} \sum_{c=3}^4 \sum_{s=1}^2 \sum_{m=-n}^n \sum_{n=1}^{\infty} Q_{smn}^{(c)} \vec{\mathbf{F}}_{smn}^{(c)}(r, \theta, \phi).\end{aligned}\tag{5.29}$$

As considerações acima possuem uma interpretação física: a radiação de ondas eletromagnéticas pode ser visualizada como uma guia de onda esférica que se estende da menor esfera que engloba a antena com raio $r = r_0$ até o infinito, como mostrado na figura 5.15. Como toda guia de onda, a guia de onda esférica possui modos de propagação ortogonais evanescente, com regiões de corte e propagação [4, 9]. Esses modos propagam em formas de ondas esféricas através das seções da guia de onda que possuem área e espectro discreto. Os modos esféricos podem ser divididos em duas classes: modo transversal elétrico, do inglês *Transverse Electric* (TE) ligados aos modos \mathbf{F}_{1mn} e o transversal magnético, do inglês *Transverse Magnetic* (TM) que obviamente estão ligados ao vetor \mathbf{F}_{2mn} . Algo bastante importante a ser ressaltado é o papel dos coeficientes Q_{smn} ligados a cada modo F_{smn} . Devido ao fato do somatório quádruplo ser uma combinação linear o que caracteriza uma antena em particular é o valor de tais coeficientes, ou seja, fazendo um paralelo com os campos radiados a partir de funções potenciais os coeficientes modais Q_{smn} são os responsáveis por conter todas as informações importantes sobre um determinado elemento radiante.

5.5.3.1 Ordem de Truncamento

Os modos esféricos são por definição infinitos como mostrado nas equações 5.28 e 5.29. Essa característica resulta na existência de modos de amplitude iguais que podem ser atenuados para valores de $n > kr_0$. Para evitar tal problema, geralmente apenas os modos para um

número limite $N < kr_0$ são levados em conta, e por consequência as combinações lineares que resultam nos campos \mathbf{E} e \mathbf{H} são truncadas em $n = N$. Dessa maneira, a equação 5.29 assume a forma:

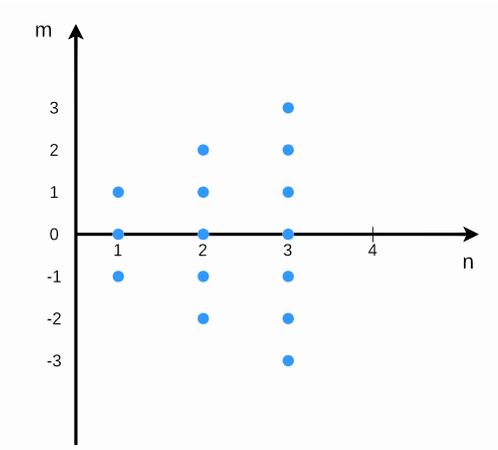
$$\begin{aligned}\mathbf{E}(r, \theta, \phi) &= \sqrt{\frac{k}{\eta}} \sum_{c=3}^4 \sum_{s=1}^2 \sum_{m=-n}^n \sum_{n=1}^N Q_{smn}^{(c)} \vec{\mathbf{F}}_{smn}^{(c)}(r, \theta, \phi), \text{ e} \\ \mathbf{H}(r, \theta, \phi) &= -jk\sqrt{\eta} \sum_{c=3}^4 \sum_{s=1}^2 \sum_{m=-n}^n \sum_{n=1}^N Q_{smn}^{(c)} \vec{\mathbf{F}}_{smn}^{(c)}(r, \theta, \phi).\end{aligned}\tag{5.30}$$

A figura 5.16 mostra a quantidade de modos para $N = 3$. Geralmente o valor máximo de N é definido como:

$$N = kr_0 + n_1,\tag{5.31}$$

em que n_1 é uma constante, em que na maioria das aplicações das expansões em ondas esféricas $n_1 = 10$ [113], possuindo o papel de mitigar possíveis erros gerados pelo truncamento do somatório.

Figura 5.16: Relação entre os índices m e n , em m pode ser considerado uma função n , do tipo $m = 2n + 1$. Nesta ilustração a ordem é truncada para $N = 3$. A quantidade de modos esféricos está intimamente ligada com o valor n , mais precisamente $Q_n = n(n + 2)$.



5.5.3.2 Aproximação de Campo Distante das Ondas Esféricas

As expressões de campo de distante baseado em expansão em ondas esféricas é obtido fazendo $kr \rightarrow \infty$ nos vetores $\mathbf{F}_{smn}^{(c)}$ para ambas polarizações, dessa maneira a influência da componente radial do campo é altamente atenuada, fazendo que ambos os campos elétrico e magnético sejam funções das componentes θ e ϕ . Fazendo essas considerações, os vetores assumem a forma:

$$\mathbf{K}_{smn}(\theta, \phi) = \lim_{kr \rightarrow \infty} \left[\frac{1}{\sqrt{\eta}} \frac{1}{\sqrt{4\pi}} \frac{e^{jkr}}{kr} \mathbf{F}_{smn}^{(3)}(r, \theta, \phi) \right].\tag{5.32}$$

E por consequência, os campos \mathbf{E} e \mathbf{H} assumem a forma [2]:

$$\begin{aligned}\mathbf{E}(\theta, \phi) &= \frac{1}{\sqrt{\eta}} \frac{1}{\sqrt{4\pi}} \frac{e^{jkr}}{kr} \sum_{smn} Q_{smn} \mathbf{K}_{smn}(\theta, \phi), \text{ e} \\ \mathbf{H}(\theta, \phi) &= \eta \hat{\mathbf{r}} \times \mathbf{E}(\theta, \phi).\end{aligned}\quad (5.33)$$

5.5.4 Potência Radiada

Com a representação de campo distante tanto para os campos descritos pelos potenciais ou pelos modos esféricos é possível definir alguns conceitos importantes como a potência radiada. Para a definição do campo por via das funções potenciais o fluxo de energia pode ser definido através do vetor de Poynting:

$$\mathbf{P} = \frac{1}{2} \text{Re} (\mathbf{E} \times \mathbf{H}^*) = \frac{\eta k^2}{32\pi r^2} |\mathbf{F}_\perp(\theta, \phi)|^2. \quad (5.34)$$

Considerando $d\Omega$ o ângulo sólido [2] a superfície de radiação, a intensidade pode ser definida como:

$$U(\theta, \phi) = \frac{d\mathbf{P}}{d\Omega} = \frac{\eta k^2}{32\pi} |\mathbf{F}_\perp(\theta, \phi)|^2. \quad (5.35)$$

O que permite calcular a potência radiada da seguinte maneira:

$$P_{rad} = \int_0^\pi \int_0^{2\pi} U(\theta, \phi) d\Omega. \quad (5.36)$$

Já para a definição para os modos esféricos essa expressão pode ser computada a partir dos coeficientes Q_{smn} [4]:

$$P_{rad} = \frac{1}{2} \sum_{smn} |Q_{smn}^{(3)}|. \quad (5.37)$$

5.5.5 Diretividade e Ganho

A diretividade e o ganho são conceitos importantes para estudos de antenas, utilizando a definição de intensidade o ganho diretivo pode ser definido como:

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{U_1}, \quad (5.38)$$

em que U_1 é a intensidade de uma antena isotrópica, ou seja,

$$U_1 = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} U(\theta, \phi) d\Omega. \quad (5.39)$$

A diretividade é região onde o ganho diretivo é máximo:

$$D_{max} = \frac{U_{max}}{U_1}, \quad (5.40)$$

em que U_{max} é valor máximo da intensidade [2]. Já para definição dos modos esféricos a diretividade é definida seguindo a definição:

$$D_{max} = \frac{\left| \sum_{smn} Q_{smn}^{(3)} K_{smn}(\theta, \phi) \right|^2}{\sum_{smn} \left| Q_{smn}^{(3)} \right|^2}. \quad (5.41)$$

Já o ganho pode ser definido tanto pela intensidade

$$G(\theta, \phi) = \frac{4\pi}{P_T} U(\theta, \phi), \quad (5.42)$$

em que P_T é a potência entregue nos terminais da antena [2], quanto pelos os modos esféricos, assumindo a forma [4]:

$$G(\theta, \phi) = \frac{\left| \sum_{smn} T_{smn} \vec{K}_{smn}(\theta, \phi) \right|^2}{\frac{P_{loss}}{P_{inc}} + \sum |T_{smn}|^2}, \quad (5.43)$$

em que T_{smn} são os coeficientes de transmissão definidos por [4] e P_{loss} e P_{inc} são as potências perdida e incidente na esfera mínima de raio r_0 .

5.5.6 Polarização

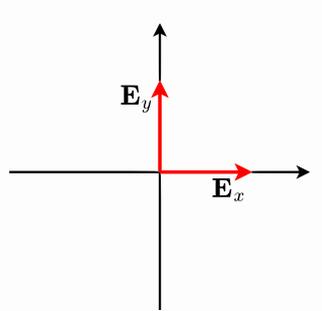
A polarização de uma onda eletromagnética pode ser definida com a curva obtida pela relação com campo eletromagnético variando no tempo a partir de um plano perpendicular a direção de propagação [2, 25], ou seja, a polarização de uma onda eletromagnética descreve como o vetor de oscilação se comporta no tempo para uma determinada posição. Existem três tipos de polarização, que são:

- Polarização linear: a onda eletromagnética varia em uma única direção ortogonal ao plano de propagação;
- Polarização circular: aqui o campo propaga-se de maneira isotrópica em todas as direções ortogais à direção de propagação;
- Polarização elíptica: o campo propaga-se com diferentes intensidades nas direções ortogonais à direção de propagação.

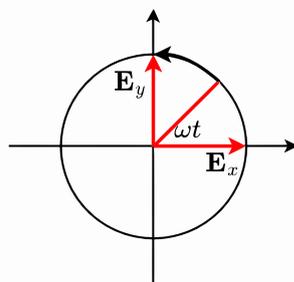
Para antenas, a polarização refere-se àquela em que o ganho é máximo [3], exceto dito o contrário. Nas polarizações circulares e elípticas o campo gira, e a direção dessa rotação

pode ser horária, *Right Handed Circular Polarization* (RHCP), ou anti-horária, *Left Handed Circular Polarization* (LHCP).

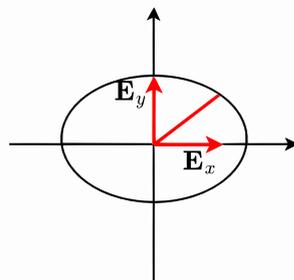
Figura 5.17: Polarizações da onda eletromagnética. Figura (a) ilustra a polarização linear, a figura (b) a polarização circular, que pode ser do tipo LHCP e RHCP, e a figura (c) a polarização elíptica.



(a) Polarização Linear



(b) Polarização Circular

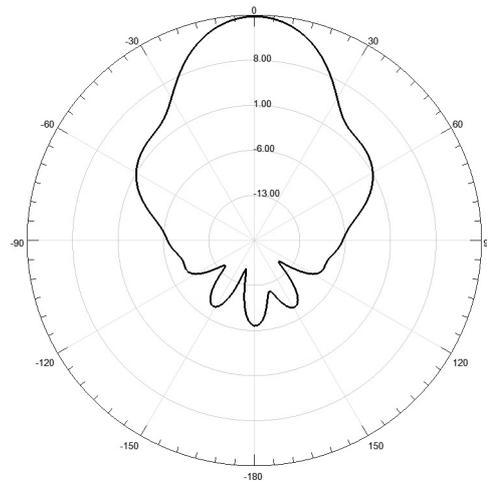


(c) Polarização Elíptica

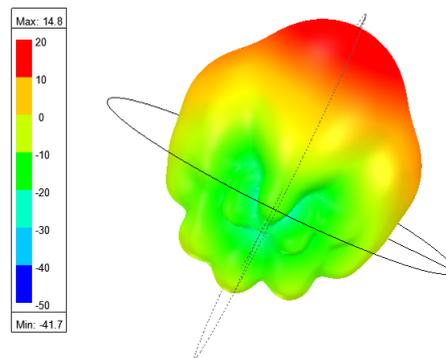
5.5.7 Padrão de Radiação

O padrão de radiação pode ser definido como a métrica que informa a distribuição espacial do campo gerado por elementos irradiantes [2, 25]. Geralmente expresso em coordenadas esféricas (θ, ϕ) , o padrão de radiação é uma função da posição angular do campo para uma dada distância de observação em campo distante. Em termos gerais o diagrama de radiação expressa a configuração espacial do campo elétrico, ao menos que seja expresso o contrário. Exemplos de padrões de radiação são mostrados na 5.18.

Figura 5.18: Exemplos de diagramas de radiação obtidos via *software* Ansys Electronics para uma antena corneta operando a 10 GHz. A figura (a) mostra a disposição do campo no eixo θ enquanto (b) mostra sua disposição tridimensional.



(a) Padrão de Radiação no Eixo θ



(b) Padrão de Radiação, Representação Tridimensional

5.6 Apêndice B: Sistema Linear Para Expansão em Ondas Esféricas

Como mostrado no capítulo 5.5, o campo elétrico pode ser expresso por meio de uma série modal esférica, que é mostrada na equação 4.2 e que pode ser truncada em uma ordem máxima N . Dessa maneira o campo elétrico pode ser escrito de uma maneira expandida da

seguinte maneira:

$$\begin{aligned} \mathbf{E}(r, \theta, \phi) = & \sqrt{\frac{k}{\eta}} \left[\left(\sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^r(r, \theta, \phi) + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^r(r, \theta, \phi) \right) \hat{\mathbf{r}} + \dots \right. \\ & \dots \left(\sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\theta(r, \theta, \phi) + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\theta(r, \theta, \phi) \right) \hat{\boldsymbol{\theta}} \dots \\ & \left. \dots \left(\sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\phi(r, \theta, \phi) + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\phi(r, \theta, \phi) \right) \hat{\boldsymbol{\phi}} \right], \end{aligned} \quad (5.44)$$

em que F_{smn}^α é componente α de \mathbf{F}_{smn} para $s = 1, 2$ e $\alpha = r, \theta, \phi$. Fazendo $F_{smn}^\alpha(r, \theta, \phi) = F_{smn}^\alpha$. Analisando o lado esquerdo da equação 5.44, sabe-se que:

$$\mathbf{E}(r, \theta, \phi) = E^r(r, \theta, \phi) \hat{\mathbf{r}} + E^\theta(r, \theta, \phi) \hat{\boldsymbol{\theta}} + E^\phi(r, \theta, \phi) \hat{\boldsymbol{\phi}}. \quad (5.45)$$

Ocultando a dependencia das posições angulares na equação 5.45 e igualando-a com a equação 5.44, obtém-se o seguinte resultado:

$$\begin{aligned} E^r &= \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^r + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^r, \\ E^\theta &= \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\theta + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\theta, \text{ e} \\ E^\phi &= \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\phi + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\phi. \end{aligned} \quad (5.46)$$

A equação 5.46 já mostra que os modos esféricos podem ser escritos em um sistema de equações lineares, contudo essa não é a sua forma final. Em sistemas de caracterização esférica a componente E^r não é amostrada, visto que só as componentes perpendiculares a direção de propagação, que é radial, são consideradas. Dessa maneira:

$$\begin{aligned} E^\theta &= \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\theta + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\theta, \text{ e} \\ E^\phi &= \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\phi + \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\phi. \end{aligned} \quad (5.47)$$

Considerando $N = 2$, os somatórios em m e n para F_{1mn}^α e F_{2mn}^α podem ser expandidas da seguinte maneira:

$$\begin{aligned} \sum_{m=-n}^n \sum_{n=1}^2 Q_{1mn} F_{1mn}^\alpha &= Q_{1-11} F_{1-11}^\alpha + Q_{101} F_{101}^\alpha + Q_{111} F_{111}^\alpha + \dots \\ &\dots Q_{1-22} F_{1-22}^\alpha + Q_{1-12} F_{1-12}^\alpha + Q_{102} F_{102}^\alpha + Q_{112} F_{112}^\alpha + Q_{122} F_{122}^\alpha, \text{ e} \end{aligned} \quad (5.48)$$

$$\begin{aligned} \sum_{m=-n}^n \sum_{n=1}^2 Q_{2mn} F_{2mn}^\alpha &= Q_{2-11} F_{2-11}^\alpha + Q_{201} F_{201}^\alpha + Q_{211} F_{211}^\alpha + \dots \\ &\dots Q_{2-22} F_{2-22}^\alpha + Q_{2-12} F_{2-12}^\alpha + Q_{202} F_{202}^\alpha + Q_{212} F_{212}^\alpha + Q_{222} F_{222}^\alpha. \end{aligned}$$

De maneira expandida a equação 5.47 assume a forma:

$$\begin{aligned} \sum_{m=-n}^n \sum_{n=1}^N Q_{1mn} F_{1mn}^\alpha &= Q_{1-11} F_{1-11}^\alpha + \dots + Q_{1NN} F_{1NN}^\alpha, \text{ e} \\ \sum_{m=-n}^n \sum_{n=1}^N Q_{2mn} F_{2mn}^\alpha &= Q_{2-11} F_{2-11}^\alpha + \dots + Q_{2NN} F_{2NN}^\alpha. \end{aligned} \quad (5.49)$$

Substituindo a equação 5.49 na 5.47:

$$\begin{aligned} E^\theta &= Q_{1-11} F_{1-11}^\theta + \dots + Q_{1NN} F_{1NN}^\theta + Q_{2-11} F_{2-11}^\theta + \dots + Q_{2NN} F_{2NN}^\theta, \text{ e} \\ E^\phi &= Q_{1-11} F_{1-11}^\phi + \dots + Q_{1NN} F_{1NN}^\phi + Q_{2-11} F_{2-11}^\phi + \dots + Q_{2NN} F_{2NN}^\phi. \end{aligned} \quad (5.50)$$

Dessa maneira a equação 5.50 pode ser escrito com um sistema linear:

$$\begin{bmatrix} E^\theta(r, \theta, \phi) \\ E^\phi(r, \theta, \phi) \end{bmatrix} = \sqrt{\frac{k}{\eta}} \begin{bmatrix} F_{1-11}^\theta & \dots & F_{1NN}^\theta & F_{2-11}^\theta & \dots & F_{2NN}^\theta \\ F_{1-11}^\phi & \dots & F_{1NN}^\phi & F_{2-11}^\phi & \dots & F_{2NN}^\phi \end{bmatrix} \begin{bmatrix} Q_{1-11} \\ \vdots \\ Q_{1NN} \\ Q_{2-11} \\ \vdots \\ Q_{2NN} \end{bmatrix}. \quad (5.51)$$

Utilizando um sistema de blocos:

$$\begin{bmatrix} E^\theta(r, \theta, \phi) \\ E^\phi(r, \theta, \phi) \end{bmatrix} = \sqrt{\frac{k}{\eta}} \begin{bmatrix} \mathbf{F}_{1mn}^\theta(r, \theta, \phi) & \mathbf{F}_{2mn}^\theta(r, \theta, \phi) \\ \mathbf{F}_{1mn}^\phi(r, \theta, \phi) & \mathbf{F}_{2mn}^\phi(r, \theta, \phi) \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{1mn} \\ \mathbf{Q}_{2mn} \end{bmatrix}. \quad (5.52)$$

Para P pontos experimentais:

$$\begin{bmatrix} E^\theta(r, \theta_1, \phi_1) \\ E^\phi(r, \theta_1, \phi_1) \\ \vdots \\ E^\theta(r, \theta_P, \phi_P) \\ E^\phi(r, \theta_P, \phi_P) \end{bmatrix} = \sqrt{\frac{k}{\eta}} \begin{bmatrix} \mathbf{F}_{1mn}^\theta(r, \theta_1, \phi_1) & \mathbf{F}_{2mn}^\theta(r, \theta_1, \phi_1) \\ \mathbf{F}_{1mn}^\phi(r, \theta_1, \phi_1) & \mathbf{F}_{2mn}^\phi(r, \theta_1, \phi_1) \\ \vdots & \vdots \\ \mathbf{F}_{1mn}^\theta(r, \theta_P, \phi_P) & \mathbf{F}_{2mn}^\theta(r, \theta_P, \phi_P) \\ \mathbf{F}_{1mn}^\phi(r, \theta_P, \phi_P) & \mathbf{F}_{2mn}^\phi(r, \theta_P, \phi_P) \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{1mn} \\ \mathbf{Q}_{2mn} \end{bmatrix}. \quad (5.53)$$

Formando um sistema linear do tipo $\mathbf{y} = \mathbf{A}\mathbf{x}$, onde $\mathbf{x} \in \mathbb{C}^{N_c \times 1}$, $\mathbf{A} \in \mathbb{C}^{P \times P \times N_c}$ e $\mathbf{y} \in \mathbb{W}^{P \times 1}$, em que \mathbb{W} simboliza o fato de \mathbf{y} poder pertencer à \mathbb{R} ou \mathbb{C} dependendo de condições da aquisição dos pontos experimentais e $N_c = 2[N(N+2)]$.

5.7 Apêndice C: Propriedades de Campo Distante das Funções de Base Esféricas

As funções de Legendre podem ser definidas como:

$$\begin{aligned} P_n^m(x) &= (1-x^2)^{m/2} \partial_{x^m} P_n(x), \\ \text{onde } P_n(x) &= \frac{1}{2^n n!} \partial_{x^n} (x^2-1)^n \end{aligned} \quad (5.54)$$

A partir das relações de ortogonalidade associadas as funções de Legendre fornecida por [90], segue: $\forall (n, n') \in \mathbb{N}^2, (n, n') \neq (0, 0)$,

$$\begin{aligned} \int_{-1}^1 \partial_t P_n^m(t) \partial_t P_{n'}^m(t) \frac{dt}{1-t^2} &= \left(\frac{2n(n+1)}{2n+1} - m \right) \frac{(n+m)!}{(n-m)!} \delta_{n,n'} \\ \int_{-1}^1 P_n^m(t) P_{n'}^m(t) \frac{dt}{1-t^2} &= \frac{1}{m} \frac{(n+m)!}{(n-m)!} \delta_{n,n'}. \end{aligned} \quad (5.55)$$

Como consequência direta:

$$\begin{aligned} \int_0^\pi (t_{mn}^1 t_{mn'}^1 + t_{mn}^2 t_{mn'}^2)_{(\theta)} \sin \theta \, d\theta &= \delta_{n,n'}, \text{ e} \\ \int_0^\pi (t_{mn}^1 t_{mn'}^2 + t_{mn}^2 t_{mn'}^1)_{(\theta)} \sin \theta \, d\theta &= 0. \end{aligned} \quad (5.56)$$