

Cleison Daniel Silva

**Processamento de Sinais de EEG para  
Classificação de Tarefas Motoras em Sistemas  
de Interface Cérebro-Máquina**

Florianópolis(SC) - Brasil

09 de maio de 2017



Cleison Daniel Silva

**Processamento de Sinais de EEG para Classificação de  
Tarefas Motoras em Sistemas de Interface  
Cérebro-Máquina**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Automação e Sistemas da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Engenharia de Automação de Sistemas.

Universidade Federal de Santa Catarina – UFSC

Departamento de Automação e Sistemas

Programa de Pós-Graduação em Engenharia de Automação e Sistemas

Orientador: Prof. Dr. Alexandre Trofino Neto

Coorientador: Prof. Dr. Jefferson Luiz Brum Marques

Florianópolis(SC) - Brasil

09 de maio de 2017

---

Cleison Daniel Silva

Processamento de Sinais de EEG para Classificação de Tarefas Motoras em Sistemas de Interface Cérebro-Máquina/ Cleison Daniel Silva. – Florianópolis(SC) - Brasil, 09 de maio de 2017-

189 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Alexandre Trofino Neto

Tese – Universidade Federal de Santa Catarina – UFSC

Departamento de Automação e Sistemas

Programa de Pós-Graduação em Engenharia de Automação e Sistemas, 09 de maio de 2017.

1. Interface Cérebro-Máquina. 2. Desigualdade Matricial Linear. I. Trofino, Alexandre. II. Universidade Federal de Santa Catarina. III. Programa de Pós-Graduação em Engenharia de Automação e Sistemas. IV. Título

CDU 02:141:005.7

---

Cleison Daniel Silva

**Processamento de Sinais de EEG para Classificação de  
Tarefas Motoras em Sistemas de Interface  
Cérebro-Máquina**

Esta Tese foi julgada aprovada para a obtenção do Título de Doutor em Engenharia de Automação e Sistemas, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Automação e Sistemas da Universidade Federal de Santa Catarina

Florianópolis(SC) - Brasil, 24 de maio de 2017.

---

**Prof. Daniel Ferreira Coutinho, Dr.**

Coordenador do Programa de Pós-Graduação em Engenharia de Automação e Sistemas

---

**Prof. Alexandre Trofino Neto, Dr.**

Orientador - DAS/UFSC

---

**Prof. Jefferson Luiz Brum Marques, Dr.**

Co-orientador - IEB/UFSC

**Banca Examinadora:**

---

**Prof. Alexandre Trofino Neto, Dr.**  
Presidente

---

**Prof. Carlos Julio Tierra Criollo, Dr.**  
Membro da Banca - PEB/UFRJ

---

**Prof. Sidney Nascimento Junior Givigi, Dr.**  
Membro da Banca - DECE/RMC

---

**Prof. Antonio Pereira Junior, Dr.**  
Membro da Banca - IC/UFRN

---

**Prof. Marcio Holsbach Costa, Dr.**  
Membro da Banca - DEEL/UFSC

---

**Prof. Hector Bessa Silveira, Dr.**  
Membro da Banca - DAS/UFSC



# Resumo

Nesta tese, é proposta uma nova abordagem para o processamento do sinal de eletroencefalograma (EEG) aplicado à interface cérebro-máquina (ICM) baseado em imagética motora (IM). A nova abordagem consiste em introduzir graus de liberdade aos classificadores baseados em Padrões Espaciais Comuns (CSP - *Common Spatial Pattern*) mais Análise Discriminante Linear (LDA - *Linear Discriminant Analysis*) e um classificador baseado na distância de Riemann. Os graus de liberdade correspondem a uma transformação para os sinais que é expressa como uma parametrização das matrizes de covariância. Esses parâmetros são determinados resolvendo um problema de otimização convexa com restrições de Desigualdades Matriciais Lineares (LMI - *Linear Matrix Inequalities*). Estes problemas são formulados para destacar propriedades que melhoram a discriminação de sinais de diferentes classes. A fim de reduzir a demanda computacional dos problemas de otimização, uma representação compacta dos sinais EEG é usada para expressar a parametrização da matriz de covariância e reduzir a dimensão do problema. O conteúdo espectral dos sinais é explicitamente encontrado nos coeficientes da representação compacta que são obtidos eficientemente projetando os sinais numa base sinusoidal definida em uma faixa de frequências de interesse. Esta propriedade permite emular a estratégia de filtro de passa-faixa a partir dos coeficientes da nova representação com redução no custo computacional. Os resultados, em termos de acurácia de classificação, foram ilustrados através de um conjunto de dados de sinal EEG registrado durante a execução de tarefas cognitivas de imagética motora. Estes conjuntos de dados estão disponíveis online gratuitamente e foram gerados por duas competições internacionais, promovidos para testar e validar algoritmos aplicados a sistemas ICM. Embora os resultados obtidos possam ser melhorados, a introdução e ajuste dos novos graus de liberdade propostos nesta tese provaram ser uma alternativa viável para obter melhores taxas de acurácia na classificação do sinal quando comparados com os métodos convencionais, via Riemann e via CSP -LDA. Outro aspecto interessante é que a representação compacta permite reduzir drasticamente o custo computacional de métodos baseados em CSP e filtros de frequência sub-banda sem comprometer o desempenho da ICM.

**Palavras-chaves:** Linear Matrix Inequalities (LMI), interface cérebro-máquina (ICM), Eletroencefalograma (EEG).



# Abstract

In this thesis, a new approach is proposed for the electroencephalogram (EEG) signal processing applied to brain-computer interface (BCI) based on motor imagery (IM). The new approach consists of introducing degrees of freedom to the classifiers based on Common Spatial Patterns (CSP) plus Linear Discriminant Analysis (LDA) and a classifier based on Riemann distance. The degrees of freedom correspond to a transformation to the signals which is expressed as a parameterization of the covariance matrices. These parameters are determined by solving a convex optimization problem with Linear Matrix Inequalities (LMI) constraints. These problems are formulated to highlight properties that improve discrimination of signals from different classes. In order to reduce the computational demand of the optimization problems, a compact representation of the EEG signals is used to express the parameterization of the covariance matrix and reduce the dimension of the problem. The spectral content of the signals is explicitly found in the coefficients of the compact representation which are obtained efficiently by projecting the signals on a sinusoidal basis defined in a range of frequencies of interest. This property allows to emulate the pass-band filter strategy from the coefficients of the new representation. The results, in terms of classification accuracy, were illustrated through an EEG signal dataset recorded during the execution of cognitive tasks of motor imagery. These datasets are available online for free and they were generated by two international competitions, promoted to test and validate algorithms applied to BCI systems. Although the results obtained can be improved, the introduction and adjustment of the new degrees of freedom proposed in this thesis have proved to be a viable alternative to obtain better accuracy rates in the signal classification when compared with the conventional methods, via Riemann and via CSP-LDA, specially when the classical methods offer poor results. Another interesting aspect is that the compact representation allows to drastically reduce the a computational cost of methods based on CSP and frequency sub-band filters without compromising BCI performance.

**Key-words:** Linear Matrix Inequalities (LMI), Brain-Computer Interface (BCI), Electroencephalogram (EEG).



# Lista de ilustrações

Figura 1 – O córtex cerebral e suas regiões, retirado de (SILVERTHORN, 2010) . . . . .	39
Figura 2 – O Homunculus de Penfield. . . . .	40
Figura 3 – Estrutura de um neurônio. . . . .	42
Figura 4 – Sistema Internacional 10-20 de Posicionamento dos Sensores para EEG (SANEI; CHAMBERS, 2007; WOLPAW; WOLPAW, 2012). . . . .	47
Figura 5 – Densidade espectral de potência do sensor na posição $O_2$ . As frequências do estímulo utilizadas foram: $9Hz$ e $10Hz$ . Destacam-se os picos referentes as frequências fundamentais e a segunda harmônica. Retirado de (LI; ANG; GUAN, 2010). . . . .	50
Figura 6 – (Curva contínua) Potencial P300 obtido a partir da média de realizações de experimentos de um único sensor (localizado na posição $P_z$ ). (Curva pontilhada) Ausência de potencial P300. Adaptado de (RIVET et al., 2009) . . . . .	50
Figura 7 – (Curva azul) Redução de energia (ERD) no espectro do EEG na faixa de frequência em torno de $10Hz$ . (Curva vermelha) Espectro do EEG sem dessincronização. Adaptado de (LI; ANG; GUAN, 2010) . . . . .	51
Figura 8 – Diagrama com as principais etapas de processamento de sinal de EEG em sistemas ICM. . . . .	53
Figura 9 – Cronologia indicando um exemplo de uma seqüência de instruções que pode ser apresentada aos indivíduos sob coleta de sinais de EEG com a finalidade de se obter sinais para ajuste dos algoritmos. . . . .	56

Figura 10	– Exemplo de $p = 22$ autovalores da matriz $S_p$ da Eq. (3.12) em ordem crescente. Observe que os valores estão limitados ao intervalo entre 0 e 1. Em geral, os autovetores associados aos pares com maior e menor valores, são escolhidos para compor a transformação definida em Eq. (3.13) . . . . .	63
Figura 11	– Exemplo de dois padrões espaciais comuns obtidos para o caso em que as classes $\mathcal{A}$ e $\mathcal{B}$ são: imaginação do movimento da mão esquerda e direita, respectivamente. . . . .	64
Figura 12	– Esboço do hiperplano de separação linear entre duas classes via LDA. Adaptado de (LOTTE et al., 2007b)	70
Figura 13	– Esboço de uma Variedade de Riemann e espaços tangentes associados. Observa-se que para dois planos tangentes tem-se mapeamentos diferentes. Adaptado de (YGER, 2013) . . . . .	72
Figura 14	– Exemplo de espectro de amplitude dos sinais $Z$ (bruto), $\hat{X}$ (projecção) e $\tilde{X}$ (filtrado, filtro não-ideal). A base de sinais senoidais foi definida com 45 frequências diferentes, igualmente espaçadas ( $\delta f = \frac{1}{2}$ ) entre as frequências de corte inferior e superior, 8 e 30Hz, respectivamente. Nota-se que os sinais $Z$ (marcador círculo e cor azul) e $\hat{X}$ (marcador asterisco e cor vermelha) apresentam o mesmo conteúdo frequencial para as frequências da base de sinais $B$ . O sinal $\tilde{X}$ (marcador quadrado e cor preta) foi obtido na saída de um filtro passa-faixa realizável, portanto não-ideal. Por essa razão, o espectro de $\tilde{X}$ é diferente do espectro de $Z$ e $\hat{X}$ , principalmente, para as frequências próximas as frequências de corte. Especificadamente, para as frequências abaixo de 8,5 Hz e acima de 28 Hz. . . . .	86

Figura 15 – a) O sinal  $X$  é a versão idealmente filtrada do sinal  $Z$  na faixa de frequência de interesse  $\Delta f = (f_l, f_u)$ . b) O sinal  $\hat{X}$  é uma projeção do sinal  $Z$  na base de sinais  $B$  definida na faixa de frequência de interesse. 87

Figura 16 – Um segmento de um canal dos sinais filtrado e projetado é mostrado em vermelho (com marcador estrela) e em preto (com marcador círculo), respectivamente. O erro quadrático médio entre esses sinais é:  $MSE = 1.6114$  que corresponde à 7,34% da energia do sinal filtrado. . . . . 88

Figura 17 – a) Sinal de EEG bruto obtido do sensor localizado em  $C_z$  com  $q = 500$  amostras ( $f_s = 250Hz$ ). b) Esboço dos termos senoidais da base de sinais senoidais ( $m = 15$ ), no intervalo entre as frequências  $f_1 = 8$  e  $f_m = 15$  Hz com  $\delta f = 0,5$ . c) Em vermelho e marcador círculo, a projeção de do sinal de EEG na base de sinais. Em preto e marcador contínuo, sinal de EEG filtrado por um filtro passa-faixa com frequências de corte  $f_l = 8Hz$  e  $f_u = 15Hz$ , respectivamente. d) Coeficientes obtidos da projeção do sinal em a) na base de sinais em b). As cores (vermelho e azul) diferenciam os coeficientes que correspondem aos termos cossenoidais e senoidais, respectivamente, da base de sinais. . . . . 89

Figura 18	– Exemplo de duas Sub-bandas de frequência: Sub-banda A e Sub-banda B, ambas definidas a partir dos coeficientes de um sensor resultantes da projeção dos sinais $Z$ na base $B$ . Tem-se os coeficientes associados aos sinais senoidais (círculo na cor azul) e aos sinais cossenoidais (asterisco na cor vermelha) da base de sinais. Para a sub-banda A tem-se $\omega_b = 13$ , enquanto que para a sub-banda B tem-se $\omega_b = 21$ . A escolha das sub-bandas A ou B, pôde ser realizada definindo-se a os elementos da matriz $H_0$ na Eq. (5.8). A energia do sinal na sub-banda A corresponde a 59,51%, enquanto a energia na sub-banda B corresponde a 29,90% da energia total na faixa entre 8 a 30 $Hz$ . . . . .	101
Figura 19	– Cadeia de processamento dos sinais de EEG utilizando as ferramentas de Riemann. Nas colunas na lateral do fluxo principal, tem-se uma breve descrição das diferenças entre a abordagem convencional e a abordagem proposta nessa tese. . . . .	113
Figura 20	– Esboço da função logarítmica ao quadrado utilizada para o cálculo da distância de Riemann. No eixo das abcissas estão os autovalores ( $\lambda$ ). No eixo das ordenadas estão os valores da função ( $f(\lambda) = \log^2(\lambda)$ ). . . . .	116
Figura 21	– Cronologia da seqüência de instruções apresentadas aos indivíduos sob coleta de sinais de EEG. Destaque para o intervalo de tempo onde os segmentos de sinais de EEG foram extraídos para ajuste e teste dos algoritmos. . . . .	123
Figura 22	– Estrutura de um método baseado em Sub-bandas adaptado de (NOVI et al., 2007a). . . . .	140
Figura 23	– Representação gráfica da projeção ortogonal da matriz $C_i$ na reta que une as matrizes médias $C_a$ e $C_b$ . . . . .	174

Figura 24 – Mapeamento do espaço de entrada $\mathcal{X}$ para o espaço de <i>features</i> $\mathcal{H}$ . A superfície de decisão correspondente a um hiperplano em $\mathcal{H}$ não necessariamente é um hiperplano em $\mathcal{X}$ . . . . .	182
Figura 25 – Forma canônica do hiperplano de separação. Destaca-se a margem de separação, linha pontilhada. . . . .	184

## Lista de tabelas

Tabela 1 – Principais ritmos da atividade cerebral . . . . .	44
Tabela 2 – Resultados de classificação dos sinais de EEG em termos de acurácia utilizando as ferramentas de Riemann. $\lambda$ é o parâmetro de regularização utilizado para cada sujeito. São mostrados os resultados com a transformação definida em Eq. (5.16) e sem a transformação, $H_0 = I$ . . . . .	105
Tabela 3 – Resultados obtidos para os dados IIa da Competição IV de sistemas ICM. São mostrados a acurácia de classificação para três abordagens. [ $H_0 = I$ ] Para a matriz $H_0$ igual a matriz identidade (equivale a substituir a filtragem convencional pela projeção dos sinais na base senoidal). [ $H_0$ opt ] Para a matriz $H_0$ obtida a partir do problema de otimização e [Refer] para os resultados extraídos da referência (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013) . . . . .	125

Tabela 4	– Resultados em termos da acurácia de classificação obtidos para os dados Iia da competição IV de sistemas ICM. [τ] Indica o pior caso de dispersão para as classes LH e RH. [proposto] Resultados obtidos para a matriz $H_0$ é resultante da solução o problema de otimização. [clássico] Resultados para a matriz $H_0$ igual a matriz identidade. . . . .	136
Tabela 5	– Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao utilizar um algoritmo de referência baseado no CSP-LDA e no sistema equivalente mas utilizando uma abordagem de projeção, descrito no Capítulo 4. . . . .	148
Tabela 6	– Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao restringir o sistema a utilizar uma banda total entre 8 e 30 Hertz e ajustar o número de sub-bandas. . . . .	148
Tabela 7	– Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao restringir o sistema a utilizar uma banda total entre 0 e 50 Hertz com 24 sub-bandas. . . . .	149
Tabela 8	– Resultados obtidos para os dados Iva da competição III de ICM ao ajustar automaticamente a sua banda total e o número de sub-bandas. . . . .	150
Tabela 9	– Resultados obtidos para o conjunto de dados 2a da Competição IV de ICM. A banda total e o número de filtros foram encontrados por um algoritmo de otimização de hiper-parâmetros no caso SBCSP. A coluna "tempo de computação" refere-se ao tempo de treinamento e teste de um sujeito em uma combinação de tarefas mentais. . . . .	152
Tabela 10	– Porcentagem de acurácia por sujeito (indicado pela primeira linha) e tarefa (indicada pela primeira coluna) do algoritmo CSP-LDA clássico. . . . .	153

Tabela 11 – Percentual de acurácia por sujeito (indicado pela primeira linha) e tarefa (indicada pela primeira coluna) do algoritmo SBCSP com as alterações propostas. . 153



# Lista de abreviaturas e siglas

ICM	Interface Cérebro-Máquina
EEG	Eletroencefalograma
LMI	Linear Matrix Inequalities
CSP	Common Spatial Pattern
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
SNC	Sistema Nervoso Central
SNP	Sistema Nervoso Periférico
ECoG	Eletrocorticograma
MEG	Magnetoencefalografia
fMRI	Imageamento por ressonância magnética funcional
NIRS	Espectroscopia do infravermelho próximo
ERP	Event Related Potencial
ERS	Event Related Synchronization
ERD	Event Related Desynchronization
SSVEP	Steady State Visually Evoked Potential
SNR	Signal-Noise Ratio
EOG	Eletro-oculograma
EMG	Eletromiograma
HRV	Heart Rate Variability
FIR	Finite Impulse Response
IIR	Impulse Infinite Response
CSP	Common Spatial Pattern
LDA	Linear Discriminant Analysis
SPD	Symmetric Positive Definite
SDP	Semidefinite Programming
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
BCI	Brain-computer interface



# Lista de símbolos

$\mathbb{R}^{p \times q}$	Conjunto de matrizes reais de dimensão $p \times q$ ;
$\mathbb{R}^p$	Espaço euclidiano de dimensão $p$ ;
$med\{X\}$	Vetor de dimensão $p$ , resultante da média entre as colunas da matriz $X$ de dimensões $p \times q$ ;
$cov\{X\}$	Matriz de covariância do segmento de sinal $X$ de dimensão $p \times q$ ;
$tr\{A\}$	Traço da matriz $A$ ;
$diag\{A\}$	Vetor contendo os elementos da diagonal da matriz $A$ ;
$\mathcal{M}^n$	Variedade de Riemann de dimensão $n$ ;
$S^n$	Espaço de matrizes simétricas de ordem $n$ ;
$\mathcal{P}^n$	Espaço de matrizes simétricas positivas definidas;
$T_\Omega \mathcal{M}$	Plano tangente à $\Omega$ , onde $\Omega \in \mathcal{M}^n$ ;
$vec\{S_i\}$	Vetor obtido a partir das colunas da matriz $S_i$ concatenadas;
$\exp(A)$	Matriz exponencial da matriz $A$ ;
$\ln(A)$	Matriz logaritmo da matriz $A$ ;
$\Theta_\beta(\Phi, \Omega)$	Geodésica entre as matrizes $\Phi$ e $\Omega$ , com $\beta \in [0, 1]$ ;
$\ A\ _F$	Norma Frobenius da matriz $A$ ;
$I_{2m}$	Matriz Identidade de ordem $2m$ ;
$H \succeq 0$	Matriz simétrica e positiva semi-definida;
$\lambda_{max}\{A\}$	Autovalor máximo da matriz $A$ ;
$\lambda_{min}\{A\}$	Autovalor mínimo da matriz $A$ ;
$\kappa\{A\}$	Número de condição da matriz $A$ ;
$\delta_R(\Sigma_\theta, \Phi_\theta)$	Distância de Riemann entre as matrizes $\Sigma_\theta$ e $\Phi_\theta$ ;



# Sumário

<b>1</b>	<b>Introdução</b>	<b>27</b>
1.1	Motivação	27
1.2	Contribuições da tese	30
1.3	Organização do documento	34
<b>2</b>	<b>Interface Cérebro-Máquina</b>	<b>37</b>
2.1	Introdução	37
2.2	Breve descrição anatômica e fisiológica do cérebro humano.	37
2.3	Eletroencefalografia - EEG.	40
2.3.1	Origem dos sinais elétricos cerebrais	41
2.3.2	Ritmos da atividade cerebral.	43
2.3.3	Localização de sensores de EEG no couro cabeludo	45
2.4	Outras técnicas para a medição da atividade cerebral	46
2.5	Tipos de Sistemas de ICM	48
2.6	Conclusões do Capítulo	52
<b>3</b>	<b>Algoritmos básicos aplicados à ICM</b>	<b>53</b>
3.1	Introdução	53
3.2	Filtragem Temporal	57
3.2.1	Filtros de Resposta Impulsiva Finita - FIR	58
3.3	Filtragem Espacial	58
3.3.1	Padrões Espaciais Comuns - CSP	60
3.3.2	Propriedades do filtro CSP	62
3.4	Extração de Características	64
3.4.1	Logaritmo da variância	65
3.4.2	Matriz de covariância	66
3.5	Classificação	66
3.5.1	Análise de Discriminantes Lineares - LDA	67
3.5.2	Classificação baseada na distância de Riemann	70
3.5.2.1	Ferramentas da Geometria Diferencial	71

3.5.2.2	Classificador via Ferramentas de Riemann	74
3.6	Métodos de avaliação de desempenho de sistemas ICM.	75
3.6.1	Acurácia de classificação . . . . .	75
3.6.2	Matriz de confusão . . . . .	76
3.7	Conclusões do Capítulo . . . . .	77
<b>4</b>	<b>Representação Compacta de Sinais de EEG . . . . .</b>	<b>79</b>
4.1	Introdução . . . . .	79
4.2	Breve revisão sobre produto interno e energia de sinais .	80
4.2.1	Base Ortogonal de Sinais . . . . .	81
4.2.2	Base de Sinais Senoidais . . . . .	83
4.3	Projeção dos sinais de EEG na base Senoidal . . . . .	84
4.4	Interpretação da Matriz de Coeficientes e Sub-bandas .	90
4.5	Conclusões do Capítulo . . . . .	92
<b>5</b>	<b>Parametrização da Transformação Linear . . . . .</b>	<b>95</b>
5.1	Introdução . . . . .	95
5.2	Transformação Linear . . . . .	96
5.3	Matriz de Covariância dos Sinais Transformados . . . . .	97
5.4	Classificação dos sinais de EEG via matriz de covariância	101
5.5	Desigualdades Matriciais Lineares . . . . .	105
5.5.1	Problema de Otimização Convexa . . . . .	106
5.5.2	Complemento Schur . . . . .	107
5.6	Determinação da Matriz de Parâmetros via Problema de Otimização . . . . .	108
5.7	Conclusões do Capítulo . . . . .	109
<b>6</b>	<b>Classificação dos Sinais de EEG via Ferramentas de Rie- mann com Otimização . . . . .</b>	<b>111</b>
6.1	Introdução . . . . .	111
6.2	Análise preliminar . . . . .	113
6.2.1	Número de condição de matriz simétrica . . . . .	114
6.2.2	Conexão entre número de condição e Distância de Riemann . . . . .	117

6.3	Formulação do problema de otimização para EEG . . . .	118
6.3.1	O problema . . . . .	119
6.4	Resultados Numéricos . . . . .	122
6.5	Conclusões do Capítulo . . . . .	125
<b>7</b>	<b>Determinação da Matriz de Parâmetros: abordagem CSP-LDA . . . . .</b>	<b>129</b>
7.1	Introdução . . . . .	129
7.2	Análise preliminar . . . . .	130
7.3	Separar o vetor de característica médio . . . . .	132
7.4	Reduzir a dispersão . . . . .	133
7.5	Resultados numéricos . . . . .	135
7.6	Considerações do Capítulo . . . . .	137
<b>8</b>	<b>Sub-bandas com Projeção na Base Senoidal . . . . .</b>	<b>139</b>
8.1	Introdução . . . . .	139
8.2	CSP via projeção na Base Senoidal . . . . .	141
8.3	Extração de Características via Sub-bandas . . . . .	142
8.4	Pontuação das Sub-bandas . . . . .	142
8.5	Classificador Bayesiano . . . . .	143
8.6	Máquina de Vetor Suporte - SVM . . . . .	145
8.7	Método Adotado para o Ajuste do Algoritmo . . . . .	145
8.8	Experimentos usando os dados IVa da III Competição de Sistemas ICM . . . . .	146
8.8.1	Descrição do conjunto de dados . . . . .	146
8.8.2	Procedimento de avaliação adotado . . . . .	147
8.8.3	Pré-processamento dos sinais . . . . .	147
8.8.4	Resultados e discussão . . . . .	147
8.8.4.1	Algoritmo CSP/LDA padrão como referência . . . . .	147
8.8.4.2	SBCSP com restrição no banco de filtros entre 8 and 30 Hz . . . . .	148
8.8.4.3	SBCSP com um banco de filtro com restrições de 24 bandas entre 0 e 50 Hz . . . . .	149

8.8.4.4	SBCSP com o banco de filtros sem restrições . . . . .	150
8.9	Experimentos usando dados do conjunto 2a da competição IV de ICM . . . . .	151
8.9.1	Descrição dos dados . . . . .	151
8.9.2	Procedimento de avaliação do sistema . . . . .	151
8.9.3	Pré-processamento . . . . .	152
8.9.4	Resultados e discussões . . . . .	152
8.10	Conclusões do Capítulo . . . . .	153
<b>9</b>	<b>Considerações Finais . . . . .</b>	<b>155</b>
	<b>Referências . . . . .</b>	<b>161</b>
	<b>Apêndices . . . . .</b>	<b>171</b>
	<b>APÊNDICE A Otimização para o Classificador . . . . .</b>	<b>173</b>
	<b>APÊNDICE B Resumos dos Artigos Científicos Produzidos ao longo do trabalho . . . . .</b>	<b>175</b>
	<b>Anexos . . . . .</b>	<b>179</b>
	<b>ANEXO A Classificador baseado em Hiperplanos . . . . .</b>	<b>181</b>
	A.1 Máquina de Vetor Suporte - SVM . . . . .	182
	A.2 Sobreposição entre exemplos de classes diferentes . . . . .	186
	<b>ANEXO B Algoritmo para a estimativa da média geométrica em Matrizes Positivas Definidas . . . . .</b>	<b>189</b>

# 1 Introdução

## 1.1 Motivação

Nas últimas duas décadas, o constante avanço em pesquisas nas áreas de neurociências, de instrumentação biomédica e de processamento digital de sinais tem proporcionado uma melhor compreensão do funcionamento do cérebro humano, bem como o aperfeiçoamento de instrumentos de medição de sinais biomédicos proporcionando maior confiabilidade e também, o desenvolvimento de algoritmos para o processamento dos sinais resultantes da medição da atividade cerebral.

Esses fatores tornaram possível o desenvolvimento de sistemas artificiais capazes de traduzir a intenção de pessoas em comandos diretamente da atividade cerebral. Os comandos são gerados a partir da medição e análise de biosinais resultantes da atividade cerebral e em seguida, são enviados para dispositivos que realizam as ações desejadas (SHIH; KRUSIENSKI; WOLPAW, 2012).

Esses sistemas foram denominados de interface cérebro-máquina (ICM) e, foram inicialmente propostos para facilitar o processo de reabilitação de indivíduos com a capacidade motora comprometida (WOLPAW et al., 2002). No entanto, diversas outras aplicações têm sido propostas, dentre elas destacam-se o controle de robôs móveis (MILLAN et al., 2004), aplicações em ambientes computacionais (EBRAHIMI, 2007) ou mesmo, o comando de dispositivos aéreos não-tripulados (LA-FLEUR et al., 2013).

Essencialmente, os comandos gerados por um sistema ICM tem origem em diferentes características da atividade cerebral. Logo, os usuários da ICM devem alterar, voluntariamente ou não, a atividade cerebral com a finalidade de gerar as ações de controle desejadas e previamente listadas em um protocolo experimental adotado (GANDHI,

2015).

Há diferentes maneiras de induzir voluntariamente atividades cerebrais distintas. Uma abordagem comumente utilizada para sistemas de ICM é a imagética motora (PFURTSCHELLER; SILVA, 1999). Nessa abordagem, os participantes são instruídos a imaginar o movimento dos membros superiores, dos inferiores ou mesmo da língua (PFURTSCHELLER; NEUPER, 2001). A imaginação do movimento de diferentes membros altera a atividade cerebral em diferentes regiões do cérebro segundo Palaniappan (2005) e assim, pode-se detectar a mudança de atividade através de sensores posicionados no couro cabeludo dos usuários da ICM (TEPLAN, 2002).

Todos os sistemas de ICM são projetados para medir a atividade cerebral e traduzi-la em ações (AMIRI; FAZEL-REZAI; ASADPOUR, 2013). O usual para realizar a medição da atividade cerebral é utilizar sensores posicionados diretamente no couro cabeludo. Esse procedimento permite obter sinais da atividade elétrica cerebral com resolução temporal adequada e sem a necessidade de procedimentos cirúrgicos (WOLPAW; WOLPAW, 2012), sendo conhecido por eletroencefalograma (EEG).

O desafio imposto a sistemas de ICM está em extrair informação, a partir dos sinais de EEG, sobre a intenção do usuário do sistema e gerar um comando, de forma confiável e robusta (KRUSIENSKI et al., 2011). Assim uma variedade de algoritmos têm sido propostos para o processamento de sinais de EEG aplicados a sistemas de ICM (BASHASHATI et al., 2007).

Em geral, os algoritmos propostos consideram propriedades dos sinais de EEG nos domínios do tempo, da frequência e do espaço (RODRÍGUEZ-BERMÚDEZ et al., 2013). O principal objetivo dos algoritmos é o de ressaltar características dos sinais associadas à atividade cerebral de interesse e, com isso, facilitar a tradução da intenção do usuário em ações por parte do sistema de ICM (DORNHEGE et al., 2006). Portanto, duas etapas essenciais para o processamento

dos sinais são a extração de características e a classificação dos sinais decodificando-os em ações de controle.

Em sistemas ICM baseados em imagética motora, utiliza-se o fenômeno da redução de energia, observada em alguns sensores e em ondas cerebrais específicas do sinal de EEG, como gatilho para se gerar as ações desejadas. Esse fenômeno é explorado por sistemas ICM a fim de se tirar proveito de mudanças nas propriedades frequenciais e espaciais dos sinais de EEG. (AGHAEI; MAHANTA; PLATANIOTIS, 2016).

Em geral, os sinais de EEG são filtrados com o uso de filtros convencionais, do tipo passa-banda por exemplo, com a finalidade de selecionar uma faixa de frequência relacionada ao fenômeno de imagética motora (BLANKERTZ et al., 2008). Ou ainda, adota-se uma abordagem utilizando-se um banco de filtros, como em Novi et al. (2007a), com o objetivo de considerar características em faixas de frequência distintas separadamente, mas com o prejuízo de aumentar a demanda computacional.

Há também informação sobre a atividade cerebral de interesse armazenada na distribuição espacial dos sensores do EEG no couro cabeludo (LOTTE; GUAN, 2011). Assim, para melhorar a qualidade dos sinais, diferentes abordagens de processamento baseadas em matriz de covariância como fonte de informação espacial dos sinais são utilizadas. Duas abordagens, que se encaixam nesse contexto, são o uso de algoritmos de filtragem espacial do tipo Padrões Espaciais Comuns - CSP (do inglês para *Common Spatial Pattern*) (RAMOSER; GERKING; PFURTSCHELLER, 2000) em conjunto com um separador linear, Análise de Discriminantes Lineares - LDA (do inglês para *Linear Discriminant Analysis*) (LOTTE et al., 2007a) assim como, algoritmos de classificação de sinais de EEG baseados em geometria de Riemann (BARACHANT et al., 2012).

O objetivo principal dessas duas abordagens é o de extrair informações úteis sobre a intenção do usuário a ser detectada a partir de

matrizes de covariância dos sinais de EEG. Para tanto, essas abordagens possuem baixa complexidade e apresentam boas taxas de classificação dos sinais (LOTTE et al., 2007a; CONGEDO; BARACHANT; ANDREEV, 2013). No entanto, a ausência de graus de liberdade parece ser a principal dificuldade na direção de melhorar ainda mais as taxas de sucesso dessas técnicas.

De fato, tanto a abordagem baseada nos algoritmos CSP-LDA quanto a baseada em geometria de Riemann, não possuem graus extras de liberdade a serem explorados (SILVA et al., 2016). Portanto, os seus parâmetros são ajustados sem a possibilidade de se adaptar e/ou otimizar o método para um problema específico. Como resultado, estes métodos tendem a ser severamente comprometidos por variações nos sinais de EEG processados.

Neste contexto, esses algoritmos para o processamento e extração de informações a partir de matrizes de covariância dos sinais de EEG utilizados em sistemas de ICM podem ser aperfeiçoados com a introdução de graus extras de liberdade e com a respectiva otimização para cada método. Esse é o foco na presente tese de doutorado.

## 1.2 Contribuições da tese

Nesta tese de doutorado, o foco é dado na otimização de um extrator de características e classificador para aplicações em sistemas de ICM baseado em imagética motora. Os resultados mostrados aqui, contudo, podem ser estendidos a qualquer aplicação que envolva a matriz de covariância como fonte de informação dos sinais a serem classificados.

A metodologia apresentada consiste em adicionar novos graus de liberdade que são expressos linearmente nas matrizes de covariância dos sinais. Assim a etapa de extração de características baseada em matrizes de covariância pode ser expressa como um problema de otimização via Desigualdades Matriciais Lineares - LMI (do inglês para *Linear Matrix Inequalities*) (BOYD et al., 1994). As restrições do pro-

blema, que consistem em LMIs a serem resolvidas, e a função objetivo a ser otimizada são determinadas a partir dos conceitos empregados nas abordagens via CSP-LDA e via geometria de Riemann.

Na abordagem com o uso dos algoritmos CSP-LDA, por exemplo, as restrições LMIs são formuladas tal que a dispersão entre as matrizes de covariância dos sinais e a matriz de covariância média seja reduzida. Ao mesmo tempo, a função objetivo do problema explora o afastamento máximo entre as médias de matrizes de covariância de duas classes diferentes. Uma vez resolvido o problema de otimização, os algoritmos CSP-LDA são utilizados para classificar os sinais.

Para a abordagem com o uso de ferramentas da geometria de Riemann, as restrições LMIs de um problema quasi-convexo foram formuladas para minimizar o número de condição do produto de duas matrizes de covariância, uma representando o sinal a ser classificado e a outra a inversa do valor médio da classe. Quanto mais próximo o número de condição estiver da unidade, mais representativo da classe é o sinal. O número de condição pode ser reduzido por meio de procedimento iterativo e com isso, espera-se reduzir a distância de Riemann entre a matriz que representa o sinal e a matriz média da classe. Uma vez que os parâmetros livres são determinados como solução de problema quasi-convexo, as ferramentas da geometria de Riemann são utilizadas para realizar a classificação do sinais de EEG.

Ao rearranjar as expressões das duas abordagens mencionadas, conseguimos expressar as soluções como um problema LMI e, portanto, otimizar sua ação sobre o procedimento de extração de características. Os graus extras de liberdade, uma das principais contribuições dessa tese, correspondem aos parâmetros de um filtro que são transformados e expressos como uma variável matricial de decisão nas LMIs.

A fim de reduzir a dimensão do problema a ser resolvido, é utilizada uma representação compacta dos sinais EEG. Essa nova representação é obtida em termos de uma matriz de coeficientes e uma base senoidal de sinais definida em uma faixa de frequência de inte-

resse. Destaca-se ainda, que a nova representação, além de compacta, contém informação do conteúdo espectral dos sinais de EEG de forma explícita. Isso se dá, pois a representação é obtida a partir de uma base senoidal em frequências específicas.

Uma terceira abordagem proposta nesta tese, consiste em definir a partir da representação compacta dos sinais de EEG matrizes de covariância associadas a um conjunto de sub-faixas de frequências, emulando assim um banco de filtros. Assim, para cada sub-faixa de frequência pode-se aplicar um filtro espacial CSP seguido de um classificador LDA. Em seguida, formar um vetor de características aumentado e utilizar um algoritmo de classificação do tipo *Support Vector Machine* (SVM) (SCHOLKOPF; SMOLA, 2003). Mostra-se que a etapa de filtragem convencional via filtros FIR ou IIR pode ser substituída, sem perda significativa de informação, pela projeção dos sinais na base senoidal, tendo como principal vantagem a redução no tempo de ajuste dos algoritmos medido por um número menor de operações realizadas.

Os resultados numéricos apresentados nesta tese utilizaram conjuntos de sinais de EEG registrados durante a realização de tarefas cognitivas de imagética motora. Esses dados estão disponibilizados *online* gratuitamente e foram gerados por duas competições internacionais promovidas para testar e validar algoritmos aplicados a sistemas de ICM. A acurácia de classificação foi utilizada nesta tese para avaliar as abordagens estudadas.

A introdução e o ajuste de novos graus de liberdade mostrou-se eficiente em obter resultados numéricos de classificação animadores quando comparados aos métodos convencionais, tanto via Riemann quanto via CSP-LDA. Ainda que apenas alguns graus de liberdade foram explorados. O que demonstra um vasto caminho a ser perseguido para o contínuo aperfeiçoamento dos métodos de classificação dos sinais.

As principais contribuições da presente tese são sintetizadas a seguir:

- Introdução de uma representação dos sinais de EEG filtrados em uma faixa de frequência de interesse de forma compacta. A representação compacta dos sinais filtrados é obtida a partir dos coeficientes resultantes da projeção dos sinais de EEG brutos (não filtrados) numa base de sinais senoidais definida em uma faixa de frequência de interesse. A energia do erro médio quadrático relativo de aproximação entre os sinais de EEG filtrados e os sinais resultantes da reconstrução a partir da base de sinais senoidais e dos coeficientes resultantes da projeção é pequena na faixa de frequência considerada.
- Introdução de uma transformação linear nos sinais de EEG tal que os parâmetros que definem a transformação são expressos como elementos de uma matriz de parâmetros simétrica e positiva semidefinida que parametriza a matriz de covariância dos sinais transformados;
- Formulação de um problema para se determinar a matriz de parâmetros e conseqüentemente a transformação linear, através da solução de um problema de otimização convexa com restrições expressas em termos de desigualdades lineares matriciais - LMIs;
- Definição de um problema de otimização convexa para determinar uma matriz de parâmetros, onde a minimização da função objetivo corresponde em afastar os valores médios de duas classes. Enquanto que a verificação de um conjunto de restrições LMIs implicam em redução da dispersão dos elementos de uma mesma classe em torno do valor médio. A função objetivo e o conjunto de restrições são escolhidas baseadas nos algoritmos de filtragem espacial (CSP) e de classificação dos sinais (LDA);
- Definição de um conjunto de restrições LMIs do problema de otimização convexo para determinar a matriz de parâmetros e reduzir a dispersão em torno da matriz de covariância dos sinais transformados média e utilizar as ferramentas da geometria de Riemann para classificar os sinais de EEG;

- Uso da informação frequencial dos sinais de EEG disponível explicitamente na representação compacta dos sinais para aproximar um banco de filtros e com isso reduzir o custo computacional e manter o desempenho de classificação do sistema ICM.

A possibilidade de ajustar os parâmetros de um filtro com a finalidade de extrair características mais significativas dos sinais de EEG e com isso melhorar as taxas de classificação das abordagens baseadas nos algoritmos CSP e LDA e ferramentas da geometria de Riemann, é um aspecto importante da presente tese de doutorado.

### 1.3 Organização do documento

Este documento está organizado em 9 Capítulos.

No Capítulo 2, são apresentados os conceitos básicos relacionados à área de neurociências utilizados na área de sistemas interface cérebro-máquina - ICM.

No Capítulo 3, uma visão detalhada dos algoritmos comumente utilizados para o processamento e classificação de sinais de EEG utilizados em sistemas ICM baseados em imagética motora é apresentada.

No Capítulo 4 é apresentada a revisão teórica sobre projeção de sinais em bases de sinais ortogonais e como utilizar esse procedimento para obter uma representação compacta dos sinais de EEG que contém informação frequencial explícita.

No Capítulo 5 mostra-se que os parâmetros de uma transformação linear podem ser expressos linearmente na matriz de covariância dos sinais. Esses parâmetros adicionam novos graus de liberdade que podem ser determinados como solução de um problema de otimização convexo.

No Capítulo 6 ferramentas de Riemann são utilizadas para a classificação dos sinais a partir da matriz de covariância. Os graus de

liberdade são ajustados para reduzir a dispersão em torno da matriz de covariância média.

No Capítulo 7 os graus de liberdade são obtidos como solução de uma problema de otimização, onde a função objetivo conduz a solução a ter valores médios afastados, e o conjunto de restrições LMI impõem limites para a dispersão entre os elementos. Os sinais são classificados utilizando a abordagem padrão CSP-LDA.

No Capítulo 8 faz-se uso da informação frequencial contida na representação compacta dos sinais de EEG para tratar as sub-bandas de frequência de forma distinta, e assim formar um vetor de característica aumentado. O algoritmo SVM - *Support Vector Machine* é utilizado para classificar os sinais.

Finalmente, no Capítulo 9 as considerações finais do trabalho são apresentadas.

Ao final do documento, encontram-se os apêndices e os anexos. No apêndice A apresenta-se uma breve discussão sobre uma abordagem alternativa em que os parâmetros livres podem ser ajustados para otimizar a saída do classificador. No apêndice B, encontram-se os resumos dos artigos científicos, publicados e/ou sob revisão, além da descrição de uma colaboração realizada ao longo do desenvolvimento desta tese. Os anexos A e B são apresentados com o objetivo tornar o documento autocontido. Neles são apresentados, respectivamente, o algoritmo SVM, assim como um algoritmo iterativo para se determinar a média geométrica entre matrizes positivas.



## 2 Interface Cérebro-Máquina

### 2.1 Introdução

O objetivo desse Capítulo é apresentar conceitos básicos, relacionados a neurociências, fundamentais para a compreensão do princípio de operação de um sistema ICM baseado em sinais de EEG.

Serão destacados também, aspectos em relação à origem dos sinais de EEG, dos ritmos da atividade cerebral estudados via EEG e dos padrões neurofisiológicos detectados no EEG que são utilizados em sistemas ICM para carregar a informação da intenção do usuário.

### 2.2 Breve descrição anatômica e fisiológica do cérebro humano.

O cérebro humano adulto abriga aproximadamente cem bilhões de neurônios que são responsáveis por gerar a atividade elétrica que é medida pelo eletroencefalograma (EEG) e utilizado por sistemas ICM.

O estudo de sistemas de interface cérebro-máquina envolve a necessidade de compreender a organização anatômica e a fisiologia básica do sistema nervoso humano. Esse entendimento permite interpretar os sinais registrados da atividade cerebral e processá-los para decodificar a intenção da pessoa e assim gerar o sinal de comando, saída do sistema de ICM.

A unidade funcional do sistema nervoso humano é o neurônio. Os neurônios são células que apresentam três partes: o corpo celular, o axônio e a árvore dendrítica. Os dendritos recebem os sinais de entrada, enquanto que o axônio conduz as informações para a saída. Os neurônios têm a capacidade de propagar sinais elétricos gerados pelo desequilíbrio entre íons do meio intracelular e do meio extracelular.

O sistema nervoso humano pode ser dividido didaticamente em duas partes principais: o sistema nervoso central (SNC) e o sistema nervoso periférico (SNP). O primeiro consiste em duas outras partes, o encéfalo e a medula espinal. Já o SNP consiste em neurônios responsáveis por captar informações sobre os meios interno e externo ao corpo humano, sendo denominados de neurônios sensoriais. Há também neurônios que conduzem a informação do SNC em direção ao SNP, sendo estes denominados de neurônios motores. De forma resumida, pode-se entender o funcionamento do SNC como a ação de integrar os sinais que chegam via neurônios sensoriais e, determinar ou não, uma resposta a ser enviada via neurônios motores (SILVERTHORN, 2010; SANEI; CHAMBERS, 2007).

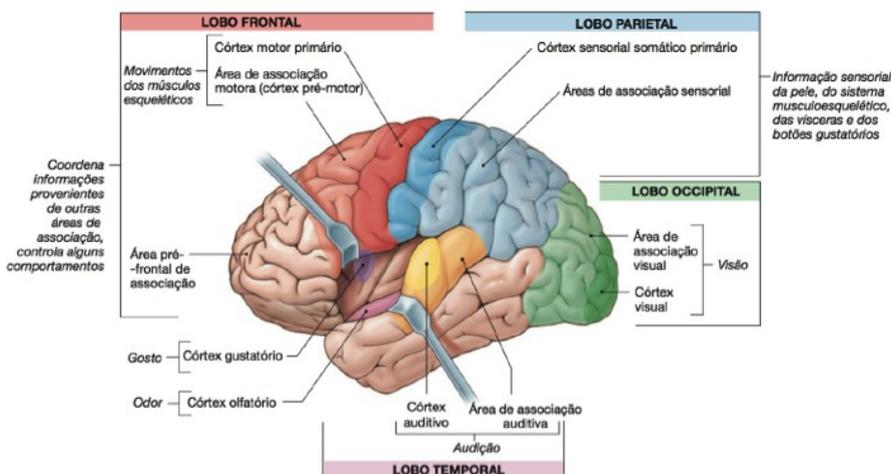
O cérebro humano constitui a maior parte do encéfalo e é responsável pelo processamento das funções superiores, como consciência, memória, linguagem, etc. O córtex cerebral é a camada mais externa do cérebro e é dividido em dois hemisférios, o direito e o esquerdo. Cada hemisfério está dividido anatomicamente em quatro lobos.

Na Figura 1 é apresentado um esquema do cérebro, destacando-se as regiões anatômicas e os lobos frontal, parietal, occipital e temporal.

O lobo frontal é responsável pelo processamento de informações da atividade motora e do planejamento racional. O lobo parietal contém o córtex sensorial e áreas de associação sensorial. É nessa região do córtex cerebral, por exemplo, que há um conjunto de neurônios envolvidos com o processamento de informações com origem nos dedos da mão. O curioso é que neurônios vizinhos a esta região processam a informação originária da palma da mão (SILVERTHORN, 2010). Já a informação visual é processada no lobo occipital, enquanto que a informação proveniente da audição é processada no lobo temporal.

Do ponto do vista funcional, o cérebro pode ser dividido em três especializações: áreas sensoriais, áreas motoras e áreas de associação.

Figura 1: O córtex cerebral e suas regiões, retirado de (SILVERTHORN, 2010)

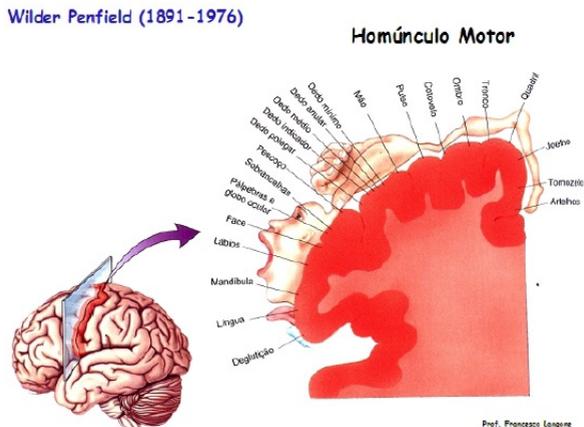


As áreas sensoriais são responsáveis por receber os estímulos provenientes do SNP e os traduzem em percepção. As áreas motoras geram os padrões de comando a serem enviados aos músculos esqueléticos, responsáveis principalmente pela sustentação e movimento do corpo humano. As áreas de associação integram as informações das áreas sensoriais e motoras, e podem comandar comportamentos voluntários. Em geral, a informação é processada por mais de uma área cerebral (SILVERTHORN, 2010).

Na Figura 2 é apresentada uma representação distorcida do corpo humano no cérebro. Nela é possível observar como diferentes partes do corpo humano são representadas no córtex, porém de uma forma diferente, tanto no tamanho quanto na disposição, de como são

encontradas no corpo.

Figura 2: O Homunculus de Penfield.



### 2.3 Eletroencefalografia - EEG.

O conceito de neurofisiologia surgiu com Carlo Matteucci (1811 – 1868) e Emil Du Bois-Reymond (1818 – 1896) quando ambos registraram pela primeira vez sinais elétricos emitidos de nervos musculares usando um galvanômetro (SANEI; CHAMBERS, 2007; SILVA, 2005).

Já os conceitos de Eletroencefalograma(EEG), eletro-(referente ao registro de sinais elétricos vindos do cérebro) encéfalo-(referente a sinais emitidos da cabeça) e grama(ou grafia que significa escrita ou desenho) surgiram com Richard Caton (1842 – 1926), cientista Inglês que fez a observação do primeiro sinal elétrico proveniente do escalpo humano em 1875 (SANEI; CHAMBERS, 2007; SILVA, 2005).

Desde o seu surgimento, a técnica de EEG evoluiu consideravelmente através das décadas, destacando importantes marcos históri-

cos que fizeram com que fosse possível atingir o nível de complexidade que os exames de EEG possuem atualmente. Alguns desses eventos podem ser citados como, por exemplo, o primeiro relatório produzido a partir de um EEG por Hans Berger (1873 – 1941), muito conhecido por grande parte dos Eletroencefalografistas, em 1929 incluindo o ritmo alfa como a principal componente de sinais EEG, o advento de um amplificador diferencial para realizar as aquisições de EEG em 1932 pela Fundação Rockefeller e o reconhecimento da importância de uma aquisição de sinais EEG utilizando mais de um canal (SILVA, 2005; SANEI; CHAMBERS, 2007).

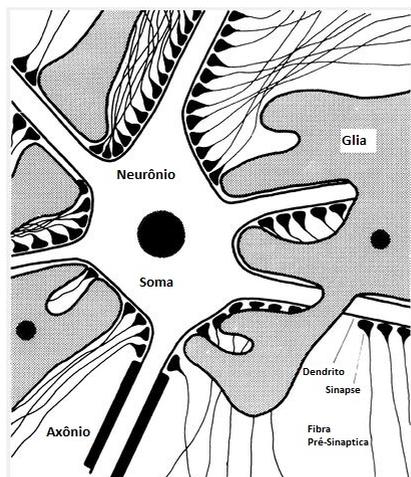
Atualmente, EEGs são realizados tanto de forma invasiva quanto não-invasiva, com diversas aplicações e diferentes equipamentos que possuem ferramentas de processamento de sinais no estado da arte, eletrodos específicos para as tarefas necessárias e capacidade de armazenamento para realizar exames de longa duração (SILVA, 2005; SANEI; CHAMBERS, 2007).

### 2.3.1 Origem dos sinais elétricos cerebrais

Para que haja um bom entendimento de como são gerados os sinais elétricos medidos pelo EEG, é interessante ter uma noção básica da estrutura de um neurônio, bem como a sua interatividade com o seu meio externo, incluindo outros neurônios.

A estrutura de um neurônio é mostrada na Figura 3, onde pode-se observar seus principais componentes estruturais, o soma (corpo celular), responsável pelo metabolismo celular, os dendritos e o axônio. Diversos processos ocorrem através da interação dessas estruturas, como, por exemplo, o axônio, cuja presença é unitária, pode-se ramificar e realizar contato com outras células nervosas ou até mesmo órgãos, enquanto que diversos dendritos podem estar presentes em uma única célula nervosa, também realizando contato com outros neurônios. Além desses tipos de comunicação entre neurônios, existem as sinapses, que também podem ser observadas na Figura 3 (SILVA, 2005).

Figura 3: Estrutura básica de um neurônio. Ênfase para o corpo celular (soma), dendritos e axônio. Extraído de (SILVA, 2005)



Os sinais observados em EEGs são provenientes de variações de potenciais no espaço e no tempo, gerados principalmente a partir de células neuronais do tipo piramidais presentes no córtex cerebral, que por conta de sua disposição paralela e proximidade de regiões superficiais de seus dendritos, domina os potenciais observados em sinais de EEG (SANEI; CHAMBERS, 2007).

Quando os neurônios são ativados por estímulos externos, como reações químicas, eletricidade, entre outros; correntes sinápticas são geradas e por sua vez dão origem a campos elétricos no couro cabeludo, medidos por equipamentos de EEG (SANEI; CHAMBERS, 2007).

Essas correntes sinápticas são produzidas pela entrada e saída de íons (sódio,  $Na^+$ , potássio,  $K^+$ , cálcio,  $Ca^{++}$ , cloro,  $Cl^-$ , entre outros) dos neurônios, processo que ocorre quando potenciais sinápticos são gerados entre essas células, que por sua vez mudam a diferença de potencial de equilíbrio da membrana celular (aproximadamente 70mV com a polaridade negativa dentro da célula) (SANEI; CHAMBERS, 2007; SILVA, 2005). Para maiores detalhes sobre como funciona esse

processo de entrada e saída de íons nos neurônios, uma boa explicação passo a passo é dada em (SANEI; CHAMBERS, 2007).

Ao realizar a aquisição de sinais da atividade cerebral há uma grande dificuldade devido à estrutura de tecidos presentes entre a localização dos sensores e o córtex cerebral (couro cabeludo, ossos, superfície externa do córtex). Esses tecidos atenuam o sinal consideravelmente, principalmente os ossos, obrigando, em geral, a adição de um alto ganho aos potenciais medidos pelos sensores para que a visualização do sinal de EEG seja possível (SANEI; CHAMBERS, 2007).

Além da atenuação, ruídos internos e externos ao couro cabeludo são gerados, dificultando ainda mais a visualização. Devido a essa adversidade, apenas a atividade sincronizada de grandes populações de neurônios, tanto na forma temporal quanto espacial são capazes de gerar potenciais suficientemente fortes para serem medidos no EEG (SANEI; CHAMBERS, 2007).

### 2.3.2 Ritmos da atividade cerebral.

Sinais de EEG podem ser utilizados para a análise dos ritmos cerebrais, relacionados tanto a certos estados mentais quanto a doenças do sistema nervoso (SANEI; CHAMBERS, 2007). Estes ritmos são, na verdade, sinais com características de amplitude e frequência relativamente definidas (apesar destas características variarem de acordo com alguns fatores, como por exemplo a idade), o que possibilita a identificação das mesmas por especialistas ou por sistemas automatizados. Os principais ritmos cerebrais são descritos a seguir. A Tabela 1 apresenta um resumo com as principais características dos ritmos cerebrais e faixas de frequências (SANEI; CHAMBERS, 2007).

- **Ritmo delta:** apresenta faixa de frequência característica entre 0,5 a 4 Hz, e está associada principalmente com o estado de sono profundo, porém pode também estar presente no estado em que o indivíduo está acordado (BLINOWSKA; DURKA, 2006). Por

Tabela 1: Principais ritmos da atividade cerebral

Ritmos	Frequência (Hz)	Características
Delta	0.5 - 4.0	Sono profundo, vigília
Theta	4.0 - 7.0	Lapsos de consciência em direção a sonolência
Alfa	8.0 - 13	Relaxado, sem atenção
Beta	13 - 30	Esta faixa de frequência também está associada com movimentos das mãos
Gamma	$\geq 30$	Funções motoras ou cognitivas. Encontrado quando o sujeito está prestando atenção ou está tendo alguma estimulação sensorial
Mu	9 – 11	Esta faixa de frequência está associada a movimentos das mãos. Encontrado sobre o córtex motor e somatossensorial

apresentar frequência relativamente baixa, é facilmente confundida com outros sinais de origem não cerebral, como a atividade muscular, causados por movimento dos grandes músculos do pescoço e mandíbula que estão próximos à pele (logo, estes sinais estão sujeitos a uma menor atenuação do que a sofrida pelos sinais gerados no cérebro) (SANEI; CHAMBERS, 2007).

- **Ritmo theta:** possui frequência na faixa de 4 a 7.5 Hz, e aparece na transição do estado consciente para o estado de sonolência. Este ritmo está também associado a acessos ao inconsciente, inspiração criativa e meditação profunda. Normalmente é acompanhado de outras frequências, e parece estar relacionado ao nível de motivação(SANEI; CHAMBERS, 2007).
- **Ritmo alpha:** este ritmo aparece na região occipital do cérebro, e está situado na faixa de frequências entre 8 a 13 Hz. Este ritmo parece indicar um estado de relaxamento, sem qualquer atenção ou concentração. Muitos indivíduos apresentam este tipo de onda com os olhos fechados, razão pela qual chegou a afirmar-se que

o ritmo alpha não passa de um estado de espera produzido pelas regiões do cérebro relacionadas a visão (SANEI; CHAMBERS, 2007).

- **Ritmo beta:** Ritmo de frequência na faixa de 14 a 26 Hz relacionado aos estados de atenção e pensamento ativos, foco no mundo exterior ou resolução de problemas concretos. Um nível alto de ritmo deste tipo é encontrado em indivíduos no estado de pânico. Quanto à localização, o ritmo beta pode ser medido principalmente nas regiões frontal e central do escalpo. Os ritmos beta podem ser suprimido por atividades motoras ou por estímulos táteis(SANEI; CHAMBERS, 2007).
- **Ritmo gamma:** Ritmo com frequência acima de 30 Hz (principalmente até 45 Hz). Embora bastante raro e de baixa amplitude ( $< 2\mu V$ ), este ritmo pode ser um bom indicativo de sincronização associado a evento (SANEI; CHAMBERS, 2007; GROSSE-WENTRUP; SCHOLKOPF; HILL, 2011).
- **Ritmo mu:** ritmos mu são um tipo especial de ritmo alpha que são observados dentro da faixa de frequências entre 9 a 11 Hz. Estão quase sempre presentes quando o sujeito está relaxado e desaparecem quando o sujeito move uma mão ou um dedo no lado contralateral, ou seja, os ritmos mu desaparecem sobre o hemisfério esquerdo do cérebro quando a mão direita é movida e vice-versa (FELZER, 2001). Assim, esta banda de frequência pode ser associada com o movimento da mão (e também com a imaginação do movimento da mão) e pode ser utilizada para distinguir entre os movimentos de cada uma das mãos (SANEI; CHAMBERS, 2007).

### 2.3.3 Localização de sensores de EEG no couro cabeludo

O posicionamento dos sensores de EEG no couro cabeludo é padronizado pelo sistema internacional 10-20 (JASPER, 1958). Conforme pode ser visualizado na Figura 4, há uma distribuição uniforme

da localização dos sensores pelo couro cabeludo utilizando como referência marcos anatômicos da cabeça, como as posições *Nasion*, *Inion* e orelhas (ponto pré-auricular) (SANEI; CHAMBERS, 2007; JASPER, 1958).

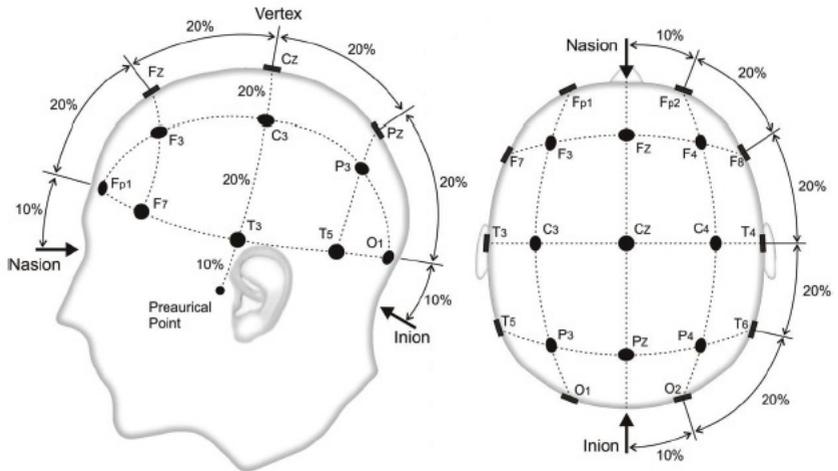
A localização dos sensores é identificada por letras e por números. As letras indicam regiões do cérebro sobre as quais os sensores se encontram, por exemplo, regiões frontal (F), central (C), temporal (T), parietal (P) e occipital (O). Os sensores posicionados sobre a linha central recebem o índice "z". Já as posições sobre o hemisfério esquerdo do cérebro recebem uma numeração ímpar enquanto que as posições sobre o hemisfério direito do cérebro recebem uma numeração par. Assim, por exemplo, a localização do sensor em  $P_z$  indica que ele se encontra sobre a região parietal e está posicionado sobre a linha central, enquanto que o sensor  $C_4$  está posicionado sobre a região central e no hemisfério direito do cérebro, como pode ser visto na Figura 4 (WOLPAW; WOLPAW, 2012).

## 2.4 Outras técnicas para a medição da atividade cerebral

No contexto de sistemas de ICM, o registro da atividade cerebral através de sinais de EEG mantém posição de destaque, principalmente por características como: baixo custo relativo dos equipamentos para medição (sistema de aquisição de dados); por dispensar procedimentos invasivos; e por medir sinais da atividade cerebral com resolução temporal elevada (MAKEIG et al., 2012). Tipicamente, o EEG mede mudanças na atividade cerebral na ordem de milissegundos (*ms*) com intensidade em microvolts ( $\mu V$ ). Mesmo assim, outras técnicas para a medição da atividade cerebral têm sido estudadas com o objetivo de aplicar em sistemas ICM.

Dentre elas, citam-se o eletrocorticograma (ECoG), a magnetoencefalografia (MEG), imageamento por ressonância magnética funcional (fMRI) e espectroscopia do infravermelho próximo (NIRS), como exemplos de técnicas disponíveis para registro da atividade cerebral.

Figura 4: Sistema Internacional 10-20 de Posicionamento dos Sensores para EEG (SANEI; CHAMBERS, 2007; WOLPAW; WOLPAW, 2012).



Segue uma breve descrição dessas técnicas alternativas para a medição da atividade cerebral:

- Eletrocorticografia - ECoG: O registro da atividade elétrica é realizado com o posicionamento de sensores na superfície do córtex. A atividade registrada é proveniente de populações de neurônios próximos aos eletrodos. Esta técnica de registro combina alta resolução temporal e espacial. No entanto, é uma técnica invasiva, com a necessidade de procedimentos cirúrgicos com riscos inerentes a saúde do sujeito (MILLER et al., 2007).
- Magnetoencefalografia - MEG: Nessa técnica a atividade elétrica

cerebral é medida através dos campos magnéticos produzidos por correntes iônicas. O método é não-invasivo, apresenta boa resolução temporal e baixa resolução espacial. No entanto, os equipamentos para medição possuem custo relativo elevado, pois os campos magnéticos possuem intensidade baixa, o que dificulta a medição (MELLINGER et al., 2007).

- Imageamento por Ressonância Magnética funcional - fMRI: Uma maneira de medir a atividade cerebral de forma indireta é monitorar o nível de oxigenação do sangue que nutre os circuitos neuronais. Quanto maior a atividade de uma região cerebral, maior o gasto de energia e maior é a quantidade de sangue necessária para repor esse gasto energético. Em imageamento por ressonância magnética funcional mede-se pequenas mudanças no fluxo sanguíneo. Esta técnica se caracteriza por baixa resolução temporal e equipamentos com custo elevados (SITARAM et al., 2007).
- Espectroscopia do Infravermelho próximo - NIRS: A espectroscopia do infravermelho próximo utiliza como princípio de funcionamento o comprimento de onda da luz infravermelha que penetra na cabeça e permite um mapeamento funcional do córtex cerebral. Diferentes níveis de oxigenação do sangue resultam em diferentes propriedades ópticas que podem ser medidas por NIRS (NAGAOKA et al., 2010).

## 2.5 Tipos de Sistemas de ICM

Um dos princípios de funcionamento de sistema ICM é a capacidade de reconhecer padrões nos sinais de EEG associados a mudanças da atividade cerebral do usuário. Em geral, essas mudanças na atividade cerebral podem ocorrer em resposta a estímulos externos ou ao desenvolvimento de tarefas cognitivas por parte do usuário da ICM. Essas mudanças são agrupadas em duas categorias principais: os potenciais relacionados à eventos (ERP, da língua inglesa, Event Rela-

ted Potencial) e a Sincronização/Dessincronização relacionada à evento (ERS/ERD, da língua inglesa, Event Related (De)Synchronization ).

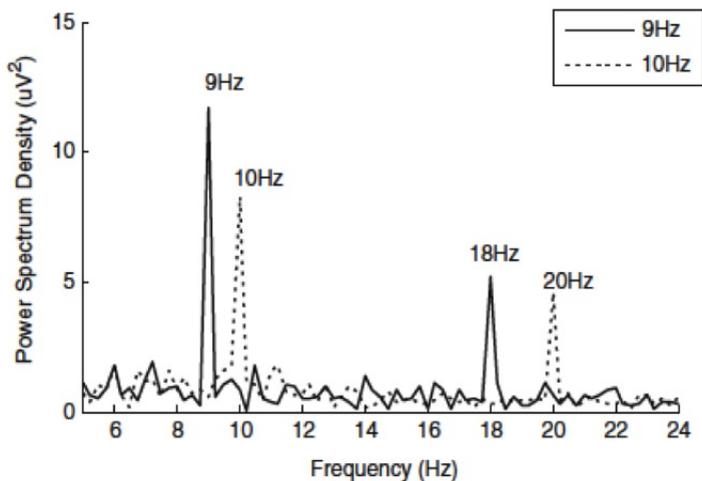
Os ERP são sinais elétricos produzidos pelo cérebro em resposta a um estímulo recebido pelo usuário. Esse estímulo pode ser visual, auditivo, tátil, etc. Em geral, esses estímulos são apresentados ao usuário em momentos pré-determinados e o sistema ICM os detecta sincronamente. Dois ERPs bastante utilizados em ICM são: os potenciais evocados visuais em regime permanente (SSVEP ) e os potenciais P300 (FATOURECHI et al., 2007).

Em ICM baseadas em SSVEP, os estímulos visuais são apresentados a taxas superiores a 4 ciclos por segundo e provocam uma resposta oscilatória contínua nas vias do sistema visual (NORCIA et al., 2015). Então o usuário deve focar a atenção através do olhar para apresentar um potencial SSVEP. Cada estímulo visual pode apresentar diferentes propriedades, como por exemplo, frequências de oscilação distintas, e assim comandos diferenciados podem ser gerados pela ICM. Na Figura 5 é ilustrado um exemplo de mudanças no espectro do sinal de EEG relacionados a propriedades frequências do estímulo.

Os potenciais P300 são gerados pelo cérebro quando estímulos pouco frequentes, no entanto desejáveis e portanto significativos, estão intercalados com outros estímulos frequentes ou rotineiros. A presença desses estímulos pouco frequentes evoca uma atividade cerebral que é percebida no EEG como um potencial de pico positivo, em aproximadamente 300ms, após o início do estímulo, localizado na região centro-parietal do córtex (em geral, melhor visualizado no sensor localizado na posição  $P_z$  do sistema 10 – 20).

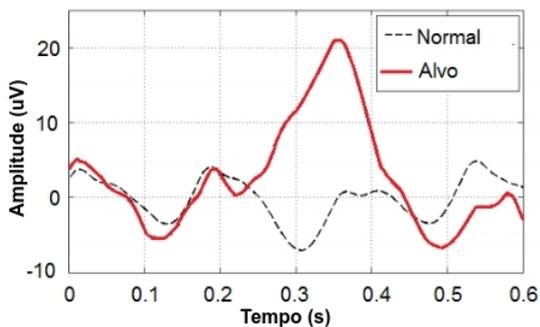
Em sistemas ICM baseados em P300, é apresentado ao usuário uma lista de estímulos, que podem ser visuais, auditivos ou táteis, cada um dos quais representa uma saída em particular. O usuário deve prestar atenção ao estímulo que representa a ação desejada. Na Figura 6 é mostrada um exemplo de potencial P300 (curva contínua) obtido através da média aritmética de várias repetições sincronizadas

Figura 5: Densidade espectral de potência do sensor na posição  $O_2$ . As frequências do estímulo utilizadas foram:  $9\text{Hz}$  e  $10\text{Hz}$ . Destacam-se os picos referentes as frequências fundamentais e a segunda harmônica. Retirado de (LI; ANG; GUAN, 2010).



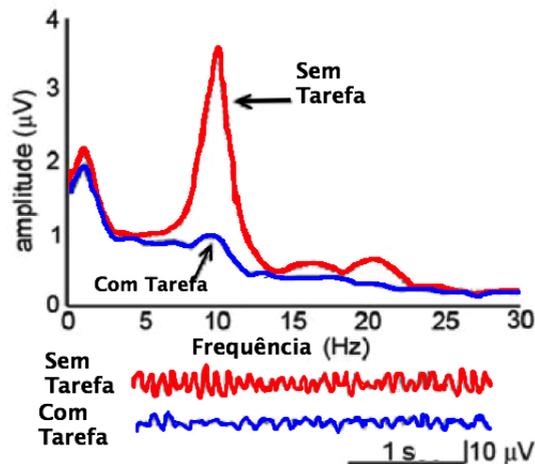
no tempo (FATOURECHI et al., 2007; LOTTE et al., 2007b).

Figura 6: (Curva contínua) Potencial P300 obtido a partir da média de realizações de experimentos de um único sensor (localizado na posição  $P_z$ ). (Curva pontilhada) Ausência de potencial P300. Adaptado de (RIVET et al., 2009)



Chamam-se de ERD e ERS os sinais gerados em resposta à tarefas cognitivas como, por exemplo, o movimento ou a preparação para o movimento dos membros. O ERD se traduz em uma redução da potência dos sinais de EEG em uma faixa de frequências específica nos sensores localizados em uma certa região do couro cabeludo, como resultado da dessincronização de um conjunto de neurônios, como pode ser visualizado na curva na cor azul da Figura 7. O ERS se traduz em um aumento da potência de uma faixa de frequências em uma certa zona, e reflete também uma sincronização do ritmo oscilatório de um grupo de neurônios, como esboçado na curva em vermelho na Figura 7.

Figura 7: (Curva azul) Redução de energia (ERD) no espectro do EEG na faixa de frequência em torno de  $10\text{Hz}$ . (Curva vermelha) Espectro do EEG sem dessincronização. Adaptado de (LI; ANG; GUAN, 2010)



Em sistemas ICM baseados em ERS/ERD, a imaginação do movimento dos membros é normalmente utilizada como tarefa mental de gatilho. Por exemplo, a imaginação do movimento da mão direita é conhecida por gerar uma redução de potência nos sensores do hemisfério

esquerdo do cérebro, mais precisamente, em torno da localização do sensor  $C_3$ .

## 2.6 Conclusões do Capítulo

Nesse Capítulo, os principais conceitos de neurociências aplicados a sistemas ICM foram apresentados, como uma descrição breve sobre a anatomia e fisiologia do cérebro humano, origens e ritmos dos sinais elétricos cerebrais e também, das principais abordagens utilizadas para se implementar sistemas ICM.

No próximo Capítulo, serão apresentados alguns algoritmos básicos utilizados ao se implementar um sistema ICM baseado em imagética motora.

# 3 Algoritmos básicos aplicados à ICM

## 3.1 Introdução

O objetivo deste Capítulo é descrever quais são as etapas básicas, do ponto de vista do processamento dos sinais de EEG, que um sistema de interface cérebro-máquina executa a fim de traduzir a intenção do usuário da ICM em comandos.

Figura 8: Diagrama com as principais etapas de processamento de sinal de EEG em sistemas ICM.



Como pode ser visto na Figura 8, uma cadeia típica do processamento de sinais de um sistema de ICM consiste em uma etapa para a aquisição de dados (sinais), seguida das etapas de pré-processamento dos sinais, de extração e seleção de características dos sinais associados

à intenção do usuário da ICM, na classificação das características em categorias relacionadas aos comandos a serem enviados ao dispositivo a ser controlado.

A primeira etapa consiste na medição da atividade cerebral. Como já destacado no Capítulo 2, há uma variedade de técnicas para a aquisição de sinais relacionados à atividade cerebral. Nesse trabalho temos interesse em sistemas para aquisição de sinais que registram os sinais de Eletroencefalograma - EEG que são amplamente utilizados em aplicações com sistemas de ICM.

Em geral, os sinais de EEG são medidos de acordo com protocolos experimentais específicos. Para sistemas de ICM baseados em imagética motora, um exemplo de protocolo experimental que pode ser utilizado consiste em medir os sinais de EEG enquanto os indivíduos estão sentados em uma cadeira confortável posicionada em frente a uma tela de computador e com os sensores do sistema de aquisição fixados com auxílio de um gel, que facilita a condução elétrica e fixação dos sensores, sobre o couro cabeludo (LEEB et al., 2007).

Através da tela do computador, os indivíduos são instruídos a desenvolver tarefas cognitivas pré-estabelecidas pelo protocolo experimental, como por exemplo, a imaginação do movimento dos membros do corpo. Os sinais são registrados com marcações em instantes de tempo pré-definidos pelo protocolo experimental para possibilitar uma análise posterior (WOLPAW; WOLPAW, 2012).

O que é esperado com esse procedimento é a obtenção de sinais de EEG em sincronia com o desenvolvimento das tarefas cognitivas pré-definidas. Assim, é possível utilizar os sinais de EEG aquisitados para realizar um ajuste de modelos computacionais com o objetivo de detectar características associadas as atividades cerebrais de interesse.

Em geral, os indivíduos que estão sob coleta dos sinais de EEG são instruídos a repetir várias vezes as tarefas cognitivas definidas pelo protocolo experimental considerado. Cada repetição é definida por uma seqüência de instruções dirigidas aos indivíduos. Um exemplo dessa

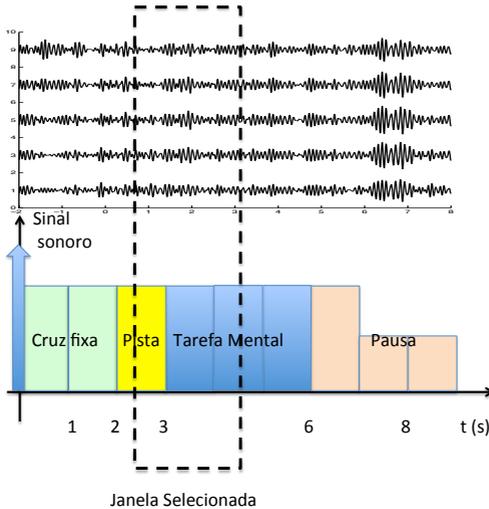
seqüência de instruções é mostrada na Figura 9.

A seqüência de instruções tem duração média de 6 segundos seguida de uma breve pausa. Tudo se inicia com um alerta sonoro (sinal sonoro) para indicar ao indivíduo que uma seqüência de instruções será apresentada. Concomitante com o alerta sonoro, na tela do computador é apresentada ao indivíduo uma indicação (uma cruz fixa) com duração de tempo de 2s. O significado da instrução é para que o usuário relaxe e concentre-se. Em seguida, é adicionado à indicação uma pista (dica) sobre qual tarefa cognitiva o indivíduo deve executar no intervalo de tempo subsequente. A pista permanece disponível na tela do computador ao indivíduo por 1s e some (desaparece), permanecendo na tela apenas a indicação inicial (a cruz). Esse evento indica ao usuário que ele deve iniciar a desenvolver a tarefa cognitiva indicada. Aguarda-se por 3s, intervalo de tempo em que o indivíduo deve executar a tarefa cognitiva indicada pela pista. Ao final desse período de tempo a indicação inicial (a cruz) desaparece indicando que a seqüência terminou e uma pausa de descanso é dada ao indivíduo.

A etapa de pré-processamento é aplicada aos sinais de EEG para detectar, reduzir ou remover ruídos, sejam eles de natureza técnica ou com origem na atividade elétrica de outras partes do corpo humano (artefatos). O objetivo é melhorar a relação sinal-ruído (SNR, do inglês para *Signal-Noise Ratio*) e isolar os sinais de interesse. Um dos ruídos presentes nos sinais de EEG é resultante da interferência da rede de alimentação (ruído da rede). Em geral, esse ruído está localizado em uma faixa de freqüências em torno de 50/60 Hz. Uma etapa de filtragem com filtros do tipo *notch* é normalmente utilizada para remover esse ruído.

Outro ruído presente nos sinais de EEG tem origem na montagem dos sensores no couro cabeludo do indivíduo. Ao longo do experimento, em geral é observado um aumento na impedância no contato entre os sensores e o couro cabeludo. Flutuações na impedância podem causar distorções nos sinais de EEG e também polarização. Ao

Figura 9: Cronologia indicando um exemplo de uma seqüência de instruções que pode ser apresentada aos indivíduos sob coleta de sinais de EEG com a finalidade de se obter sinais para ajuste dos algoritmos.



se detectar essas distorções, um procedimento que pode ser adotado é, simplesmente, descartar (remove) os sinais de EEG distorcido e, também reaplicar o gel nos contatos entre os sensores e o couro cabeludo. Para maiores detalhes sobre o procedimento de colocação dos sensores no couro cabeludo, consulte [Teplan \(2002\)](#).

Os ruídos descritos até o momento têm origem de natureza técnica, seja por causa da montagem dos sensores no couro cabeludo (aumento de impedância) ou seja por interferências causadas por conta da fonte de alimentação dos circuitos elétricos (ruídos da rede de alimentação). No entanto, há outros tipos de ruídos que possuem como origem a atividade elétrica de outras partes do corpo humano que não o cérebro. Esses tipos de ruídos são denominados de artefatos.

Dentre os principais artefatos presentes em sinais de EEG destacam-se os que têm origem, por exemplo, na atividade elétrica devido ao movimento do globo ocular (EOG), de piscadas (EOG), da atividade muscular de músculos que sustentam o crânio ou controlam

o movimento da mandíbula (EMG) e até mesmo batimento cardíaco (HRV). Maiores detalhes são apresentados na referência (FATOURECHI et al., 2007).

Uma grande quantidade de algoritmos tem sido desenvolvidos nos últimos anos com a finalidade de realizar o pré-processamento dos sinais de EEG aplicados a sistemas de ICM, como pode ser verificado em (BASHASHATI et al., 2007). Dentre eles destacam-se o uso de: filtros temporais (também conhecidos como filtros espectrais ou filtros frequenciais) com a finalidade de selecionar faixas de frequências de interesse, normalmente associadas a características das tarefas cognitivas desenvolvidas e também atenuar artefatos; filtros espaciais para melhorar a relação sinal-ruído dos sinais de EEG, principalmente, combinando os sinais obtidos em diferentes sensores para atenuar a presença de artefatos e concentrar as características de interesse em um número menor de parâmetros (BASHASHATI et al., 2007).

## 3.2 Filtragem Temporal

A filtragem dos sinais de EEG é uma estratégia comumente adotada nos estágios iniciais do processamento dos sinais em sistemas ICM. Com a filtragem, busca-se melhorar a qualidade dos sinais, principalmente através da seleção de faixas de frequência de interesse.

A seleção da faixa de frequência está associada às propriedades que se deseja ressaltar nos sinais. Por exemplo, o desenvolvimento de tarefas cognitivas, como a imaginação de movimento das mãos, altera propriedades nos sinais de EEG em faixas de frequências específicas, como nos ritmos  $\mu$  e  $\beta$  (PFURTSCHELLER et al., 2005), veja Tabela 1.

A presença de ruídos e artefatos também pode ser atenuada através da filtragem dos sinais. Um ruído típico presente nos sinais biomédicos, dentre eles o EEG, é interferência da rede de alimentação elétrica dos circuitos normalmente localizada na faixa de frequências de 50/60 Hz. Filtros do tipo rejeita faixa são utilizados para atenuar

esse ruído.

Métodos de filtragem convencionais, como os filtros de resposta impulsiva finita - FIR e filtros de resposta impulsiva infinita - IIR, são normalmente utilizados para limitar os sinais de EEG em faixa de frequências. A análise e o projeto de filtros FIR e IIR estão fora do escopo desse trabalho. Há uma vasta literatura dedicada a esse fim (SANEI; CHAMBERS, 2007). Nesta seção, segue uma breve revisão com as principais características dos filtros FIR.

### 3.2.1 Filtros de Resposta Impulsiva Finita - FIR

Os filtros FIR são sistemas lineares e invariantes no tempo que utilizam as  $n_c$  últimas amostras do sinal não filtrado,  $z(n)$ , para determinar o sinal filtrado  $x(n)$  (IFEACHOR; JERVIS, 2002).

$$x(n) = \sum_{k=0}^{n_c} a_k z(n-k) \quad (3.1)$$

onde  $a_k$  são os coeficientes do filtro. A operação em Eq. (3.1) define a convolução do sinal com o núcleo dado pelos coeficientes  $\{a_0, \dots, a_{n_c}\}$ . Os coeficientes  $a_k$  são determinados em função da resposta impulsiva desejada para o filtro FIR. Dentre as possibilidades, destacam-se os filtros FIR passa-baixas, passa-bandas, passa-altas e o rejeita faixa (IFEACHOR; JERVIS, 2002).

Os filtros FIR tem sido utilizados para a filtragem de sinais biomédicos principalmente por, quando devidamente projetados, não introduzir distorções de fase nos sinais.

## 3.3 Filtragem Espacial

Os sinais de EEG apresentam como característica uma relação sinal-ruído baixa indicando que padrões associados a tarefas mentais de interesse podem ser atenuados ou distorcidos pela presença de artefatos. Uma maneira de melhorar a relação sinal-ruído de sinais de EEG é aplicar transformações espaciais para destacar características

associadas aos padrões de interesse. Essas transformações, em geral, combinam as informações dos sensores registrados em várias posições do escalpo para realçar propriedades dos sinais.

De forma geral, seja  $X$  uma matriz de dimensões  $p \times q$  que representa um segmento de sinal de EEG amostrado a uma taxa  $f_s$ , com  $p$  sensores e  $q$  amostras por sensor. Define-se a transformação linear como,

$$Y = WX \quad (3.2)$$

onde  $W$  é a matriz de transformação com dimensões  $p \times p$  e  $Y$  são os sinais resultantes da transformação, denominados de componentes.

Em geral, as transformações  $W$  são obtidas a partir de informações de estatísticas de primeira e de segunda ordem, como os valores médios e a variância dos sinais (KEINOSUKE, 1990).

Seja a partição do segmento de sinal  $X \triangleq [x_1 \cdots x_q] \in \mathbb{R}^{p \times q}$ , com  $x_k \in \mathbb{R}^p$ ,  $\forall k = 1, \dots, q$ . Define-se o vetor média entre as  $q$  amostras do segmento de sinal  $X$  como

$$\text{med}\{X\} \triangleq \frac{1}{q} \sum_{k=1}^q x_k, \quad (3.3)$$

onde  $\text{med}\{X\} \in \mathbb{R}^p$  representa o valor médio temporal dos  $p$  sinais em  $X$ .

A matriz de covariância de um segmento de sinal com média nula ( $\text{med}\{X\} = 0$ ) é definida como

$$\text{cov}\{X\} \triangleq \frac{1}{q} XX^T, \quad (3.4)$$

em que  $\text{cov}\{X\} \in \mathbb{R}^{p \times p}$  é uma matriz simétrica que representa a dispersão do sinal no entorno do valor médio (nulo). Ela representa a energia do sinal.

Um procedimento para se determinar a matriz de transformação  $W$  que demonstrou-se efetivo em ressaltar características dos sinais de EEG associados a duas tarefas mentais diferentes é conhecida na literatura como padrões espaciais comuns - CSPCSP (do inglês, Com-

mon Spatial Pattern) (RAMOSER; GERKING; PFURTSCHELLER, 2000). Esse procedimento é descrito na próxima seção.

### 3.3.1 Padrões Espaciais Comuns - CSP

Considere que segmentos de sinais de EEG estão agrupados em duas classes  $\mathcal{A}$  e  $\mathcal{B}$ , indicadas pelo rótulo  $\boldsymbol{\rho} \in \{\mathbf{a}, \mathbf{b}\}$ , contendo  $n_a$  e  $n_b$  elementos, respectivamente. Cada segmento de sinal,  $X_{\boldsymbol{\rho}_i} \in \mathbb{R}^{p \times q}$ , contém  $p$  sensores e  $q$  amostras por sensor, amostrados a uma taxa  $f_s$ .

A matriz de covariância normalizada de cada segmento é obtida fazendo-se,

$$C_{\boldsymbol{\rho}_i} \triangleq \frac{\text{cov}\{X_{\boldsymbol{\rho}_i}\}}{\text{tr}\{\text{cov}\{X_{\boldsymbol{\rho}_i}\}\}}, \quad (3.5)$$

onde  $\text{tr}\{A\}$  é o operador traço da matriz  $A$  e o fator  $\text{tr}\{\text{cov}\{X_{\boldsymbol{\rho}_i}\}\}$  é responsável pela normalização de  $C_{\boldsymbol{\rho}_i}$ . Observe que a soma dos autovalores de  $C_{\boldsymbol{\rho}_i}$  é igual a unidade como resultado da normalização.

Para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente, as médias das matrizes de covariância são definidas pelas equações Eq. (3.6) e Eq. (3.7), respectivamente.

$$M_{\mathbf{a}} = \frac{1}{n_a} \sum_{i=1}^{n_a} C_{\mathbf{a}_i}, \quad (3.6)$$

e

$$M_{\mathbf{b}} = \frac{1}{n_b} \sum_{i=1}^{n_b} C_{\mathbf{b}_i}. \quad (3.7)$$

Matrizes de covariância são simétricas e positivas semi-definidas por construção (KEINOSUKE, 1990). Assim a decomposição espectral da soma das matrizes médias  $M_{\mathbf{a}}$  e  $M_{\mathbf{b}}$  é dada por,

$$M = M_{\mathbf{a}} + M_{\mathbf{b}} = U\Lambda U^T, \quad UU^T \triangleq I, \quad (3.8)$$

onde as matrizes  $U$  e  $\Lambda$  são as matrizes com os autovetores e autovalores de  $M$ , respectivamente. A matriz  $\Lambda$  é uma matriz diagonal, cujo elementos são aos autovalores de  $M$ . Considere a matriz de transformação

$Q$ , definida como segue

$$Q = \Lambda^{-\frac{1}{2}} U^T. \quad (3.9)$$

Observe que ao multiplicar ambos os lados a Eq. (3.8) por  $Q$  obtém-se

$$Q(M_{\mathbf{a}} + M_{\mathbf{b}})Q^T = I, \quad (3.10)$$

onde utilizou-se  $QMQ^T \triangleq I$ . Defina  $S_{\mathbf{a}} = QM_{\mathbf{a}}Q^T$  e  $S_{\mathbf{b}} = QM_{\mathbf{b}}Q^T$ . Então a Eq. (3.10) pode ser re-escrita como segue,

$$S_{\mathbf{a}} + S_{\mathbf{b}} = I \quad (3.11)$$

implica que  $S_{\mathbf{a}}S_{\mathbf{b}} = S_{\mathbf{b}}S_{\mathbf{a}}$ .

A igualdade na Eq. (3.11) implica que as matrizes  $S_{\mathbf{a}}$  e  $S_{\mathbf{b}}$  compartilham os mesmos autovetores e a soma dos autovalores é igual à unidade. Para visualizar essa propriedade entre as matrizes  $S_{\mathbf{a}}$  e  $S_{\mathbf{b}}$  considere as decomposições espectrais dessas matrizes:  $S_{\mathbf{a}} = V\Lambda_{\mathbf{a}}V^T$  e  $S_{\mathbf{b}} = V\Lambda_{\mathbf{b}}V^T$ , onde  $V$  é a matriz de autovetores e  $\Lambda_{\rho}$  são matrizes diagonais cujo elementos são os autovalores, com  $\rho \in \{\mathbf{a}, \mathbf{b}\}$ . Então,

$$S_{\mathbf{a}} + S_{\mathbf{b}} = V(\Lambda_{\mathbf{a}} + \Lambda_{\mathbf{b}})V^T = I, \quad VV^T \triangleq I \quad (3.12)$$

Considere que os autovetores de  $S_{\rho}$  estão nas colunas da matriz  $V$ , tendo-se  $V \triangleq [v_1 \cdots v_p]$ , com  $v_j \in \mathbb{R}^p$  o autovetor associado ao autovalor  $\lambda_{\rho_j}$  para  $j = 1, \dots, p$ . Destaca-se que a partir da Eq. (3.12) tem-se  $\lambda_{\mathbf{a}_j} + \lambda_{\mathbf{b}_j} = 1$ .

A partir das equações Eq. (3.9) e Eq. (3.12), define-se a matriz de transformação  $W$ , como segue,

$$W = V^T Q \quad (3.13)$$

Ao aplicar  $W$  na equação Eq. (3.8) obtém-se a diagonalização das matrizes  $M_{\mathbf{a}}$  e  $M_{\mathbf{b}}$ , simultaneamente. Destaca-se, novamente, que a soma dos autovalores após a transformação é unitária.

$$W(M_{\mathbf{a}} + M_{\mathbf{b}})W^T = \Lambda_{\mathbf{a}} + \Lambda_{\mathbf{b}} = I \quad (3.14)$$

**Algoritmo 3.1.** *Algoritmo para determinar o filtro espacial CSP.*

- **Entradas:** Conjunto de matrizes de covariância normalizada,  $\{C_{\mathbf{a}_1}, C_{\mathbf{a}_2}, \dots, C_{\mathbf{a}_{n_a}}, C_{\mathbf{b}_1}, C_{\mathbf{b}_2}, \dots, C_{\mathbf{b}_{n_b}}\}$  segundo Eq. (3.5);
  - **Saídas:** Transformação espacial CSP:  $W$
1. Calcula a matriz de covariância média para cada uma das classes, Eq. (3.6) e Eq. (3.7);
  2. Determina a matriz de transformação para normalizar os sinais, Eq. (3.9);
  3. Calcula matriz com os autovetores, Eq. (3.12);
  4. Determinar o CSP via Eq. (3.13)

### 3.3.2 Propriedades do filtro CSP

Considere aplicar a transformação definida em Eq. (3.13) a cada um dos segmentos de sinal  $X_{\rho_i}$ , obtém-se,

$$Y_{\rho_i} = W X_{\rho_i} \quad (3.15)$$

onde os sinais obtidos após a transformação são denominados de componentes. Note que, a transformação Eq. (3.13) pode ser interpretada em duas etapas: a primeira realizada pela matriz  $Q$  e a segunda etapa, realizada pela matriz  $V$ .

Na primeira etapa é realizada uma normalização dos sinais em relação à matriz de covariância média, definida em Eq. (3.8). Na segunda etapa, os sinais normalizados são projetados nas direções definidas pela matriz de autovetores  $V$ . Observe que nem todas as direções produzem variância máxima para uma classe e mínima para a outra (LOTTE; GUAN, 2010). Lembre que os autovalores das matrizes  $S_{\mathbf{a}}$  e  $S_{\mathbf{b}}$  estão entre zero e um, Eq. (3.12), logo as direções associadas aos maiores e aos menores autovalores produzirão as componentes com maior e menor variâncias, respectivamente. Por exemplo, os autovalores que satisfazem a relação,

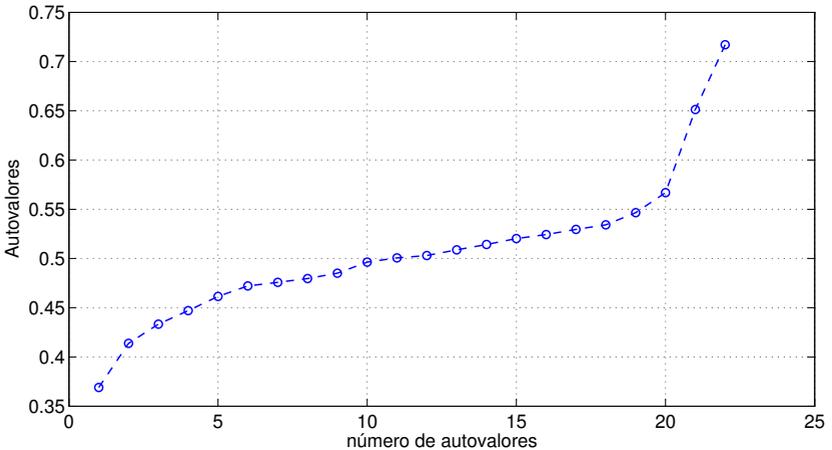
$$|\lambda_{\rho_i} - 0,5| \geq \lambda_{\ell}, \forall i = 1, \dots, p. \quad (3.16)$$

onde  $0 \leq \lambda_\ell < 0,5$  é o valor que define o limiar para a escolha dos autovetores associados aos autovalores que satisfazem Eq. (3.16). Na Figura 10 são ilustrados os autovalores de uma matriz  $S_{\mathbf{a}}$ .

A transformação definida em Eq. (3.13) é conhecida na literatura de sistemas de ICM como filtro espacial CSP. Este filtro é utilizado para realçar propriedades nos sinais de EEG realizando um contraste entre duas classes. Ao aplicar uma transformação linear no espaço dos sensores do sinal de EEG, o filtro CSP determina direções em que a variância de uma classe é máxima, e simultaneamente é mínima para a outra classe.

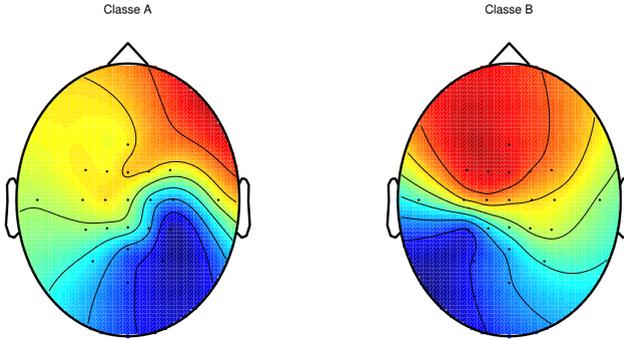
Destaca-se que a eficiência do filtro CSP, definido pela Eq. (3.13), está relacionada ao quanto as médias normalizadas das matrizes de covariância, em Eq. (3.6) e Eq. (3.7) são representativas das classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente.

Figura 10: Exemplo de  $p = 22$  autovalores da matriz  $S_p$  da Eq. (3.12) em ordem crescente. Observe que os valores estão limitados ao intervalo entre 0 e 1. Em geral, os autovetores associados aos pares com maior e menor valores, são escolhidos para compor a transformação definida em Eq. (3.13)



As colunas da matriz  $W^{-1}$  são denominadas de padrões espa-

Figura 11: Exemplo de dois padrões espaciais comuns obtidos para o caso em que as classes  $\mathcal{A}$  e  $\mathcal{B}$  são: imaginação do movimento da mão esquerda e direita, respectivamente.



ciais comuns. Os padrões espaciais comuns representam a contribuição de cada componente para a reconstrução do sinal de EEG e podem ser utilizadas para visualizar a distribuição topográfica das componentes do CSP. A Figura 11 ilustra os padrões espaciais comuns associados aos autovalores máximo e mínimo, para o caso em que se tem duas classes.

Algumas variações do algoritmo CSP foram propostas nos últimos anos, dentre elas destacam-se versões regularizadas do CSP [Lotte e Guan \(2011\)](#), que consideram informações frequenciais através de sub-bandas de frequências ([NOVI et al., 2007b](#)), que combinam de forma ótima filtros espaciais e temporais ([DORNHEGE et al., 2006](#)) e finalmente, informações espectrais dos sinais ([LEMM et al., 2005](#))

### 3.4 Extração de Características

A extração de características de sinais de EEG para fins de classificação, em geral, tem como objetivo representar os sinais de forma compacta e ao mesmo tempo, com propriedades que destaquem diferenças entre as diferentes classes de interesse ([DUDA et al., 2001](#)). É desejável também, que o conjunto de características selecionadas sejam invariantes à presença de ruído nos sinais.

Nesta seção será descrito o método de extração de características baseado na variância dos sinais de EEG após a transformação espacial via CSP, e também a utilização direta da matriz de covariância como característica dos sinais de EEG.

### 3.4.1 Logaritmo da variância

Em sinais de EEG pré-processados por filtros espaciais, em especial o filtro CSP descrito na seção anterior, têm-se componentes com variância máxima para uma classe e simultaneamente, mínima para a outra classe. Nesse sentido, pode-se utilizar a variância das componentes para formar um vetor de características.

Note que a variância de cada componente pode ser obtida a partir dos elementos da diagonal da matriz de covariância dos sinais de EEG após a transformação realizada pela matriz  $W$ , Eq. (3.13). Então, tem-se que o vetor de característica é dado por,

$$z_{\rho_i} \triangleq \text{diag}\{WC_{\rho_i}W^T\} \in \mathbb{R}^p, \forall i = 1, \dots, n_a + n_b \quad (3.17)$$

onde  $\text{diag}\{A\}$  é o operador que extrai os elementos da diagonal da matriz  $A$ .

O vetor de característica deve capturar as propriedades relevantes do sinal associadas aos padrões de interesse (HERMAN et al., 2008). Nesse sentido, como visto anteriormente, nem todas as componentes dos sinais apresentam a propriedade de variância (energia) máxima para uma classe e mínima para a outra, veja Figura 10. Assim, pode-se selecionar um subconjunto de filtros espaciais na Eq. (3.13) e com isso, reduzir a dimensão do vetor de características.

Um critério para a seleção de um subconjunto de filtros espaciais com maior relevância para a discriminação entre as classes é o dado pela Eq. (3.16), baseado em (BARACHANT et al., 2010). Define-se um valor para o limiar  $\lambda_\ell$  e seleciona-se apenas os filtros espaciais que satisfazem a Eq. (3.16). Suponha que após determinar o valor de  $\lambda_\ell$ , a dimensão do vetor de características Eq. (3.17) é reduzido para  $r \leq p$ .

Uma segunda abordagem, segundo (LOTTE; GUAN, 2011), bastante difundida é a simples escolha de três pares de filtros espaciais associados aos maiores e menores autovalores em Eq. (3.16).

O vetor de característica em Eq. (3.17) pode ainda ser normalizado e mapeado pela função log com o objetivo de tornar a distribuição dos vetores próximos da distribuição normal, segundo Blankertz et al. (2008). Assim, tem-se ainda como alternativa a Eq. (3.17),

$$\tilde{z} = \log(z/c) \quad (3.18)$$

onde  $c$  é constante e definida como  $c = \sum_{j=1}^r z_j$ .

### 3.4.2 Matriz de covariância

A matriz de covariância dos sinais de EEG é utilizada para se obter transformações espaciais como descrito nas seções anteriores (MOWLA et al., 2016). No entanto, uma abordagem alternativa que tem se mostrado eficiente (BARACHANT et al., 2012), é utilizar as matrizes de covariância diretamente como portadoras de informação para fins de classificação. Isso é possível ao se fazer uso do arcabouço teórico de geometria diferencial, onde matrizes positivas definidas formam os elementos que compõem um espaço não-Euclidiano (BHATIA, 2013).

As ferramentas utilizadas para classificar os sinais a partir das matrizes de covariância serão descritas na seção 3.5.2.

## 3.5 Classificação

Uma etapa essencial para o processamento de sinais em sistemas ICM é a classificação. Nela tem-se como objetivo assinalar rótulos aos vetores de características a fim de classificá-los. Nessa etapa, em geral, os vetores de características são mapeados para um novo espaço onde a decisão quanto ao rótulo do vetor e, portanto, quanto à classe, é tomada. Nesse novo espaço, denominado de espaço de decisão, regiões específicas determinam a classe dos sinais.

Considere o vetor  $z \in \mathbb{R}^r$  representando o vetor de características na entrada do classificador. Seja  $f$  o mapeamento do vetor  $z$  para a saída do classificador denominado de espaço de decisão. Então, tem-se

$$y = f(z), \quad (3.19)$$

onde  $y \in \mathbb{R}$ .

O problema de classificação consiste em determinar uma função  $f$  adequada para realizar a classificação dos vetores de entrada no espaço de características para o espaço de decisão (LOTTE et al., 2007b).

Considere  $z_{\rho_i} = [z_1 \ z_2 \ \cdots \ z_r]^T \in \mathbb{R}^r$  o  $i$ -ésimo vetor de característica de entrada com dimensão  $r$  e rótulo  $\rho$ , com  $i = 1, 2, \dots, n_v$ , onde  $n_v$  indica o número de vetores de entrada. O índice  $\rho$  é a função rótulo que define a classe do vetor  $z_{\rho_i}$ . A função rótulo assume valores no conjunto alfabético  $\rho \in \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots\}$ .

Seja  $Z_{\rho}$  a matriz contendo em suas  $n_v$  colunas, os vetores  $z_{\rho}$  pertencentes à classe definida pelo rótulo  $\rho$ , ou seja,

$$Z_{\rho} = \begin{bmatrix} z_{\rho_1} & z_{\rho_2} & \cdots & z_{\rho_i} & \cdots & z_{\rho_{n_v}} \end{bmatrix}.$$

Por exemplo, no caso em que se tem duas classes  $\mathcal{A}$  e  $\mathcal{B}$  indicadas pelo rótulo  $\rho \in \{\mathbf{a}, \mathbf{b}\}$  e cada uma contendo  $n_v$  vetores, definem-se as matrizes  $Z_{\mathbf{a}}$  e  $Z_{\mathbf{b}}$ , para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente.

Em sistemas de ICM, vários métodos têm sido utilizados para fins de classificação (LOTTE et al., 2007b). Nesta seção é realizada uma breve descrição dos seguintes métodos: análise de discriminantes lineares (LDA) (LOTTE et al., 2007b) e método baseado em geometria de Riemann (BARACHANT et al., 2012).

### 3.5.1 Análise de Discriminantes Lineares - LDA

O algoritmo LDA realiza um mapeamento linear dos vetores de característica para o espaço de decisão. Esse mapeamento determina

um hiperplano de separação entre os vetores de característica de classes diferentes (LOTTE et al., 2007b).

Supõe-se que os vetores de entrada possuem distribuição normal. Para duas classes, deseja-se determinar o vetor de pesos  $w = [w_1 \ w_2 \ \cdots \ w_r]^T \in \mathbb{R}^r$  que projeta os vetores de entrada, do espaço de características, para regiões diferentes no espaço de decisão de acordo com as classes. O hiperplano de separação entre as classes é dado por

$$f(z_\rho) = w^T z_\rho = \sum_{j=1}^r w_j z_j \quad (3.20)$$

O vetor  $w$  é determinado pelo LDA com base na estrutura de espalhamento dos vetores de entrada,  $z_{\rho_i}$ , para  $i = 1, 2, \dots, n_e$  que formam o conjunto de treinamento com  $n_e$  elementos<sup>1</sup> (LOTTE et al., 2007b).

Considere  $Z_{\mathbf{a}} \in \mathbb{R}^{r \times n_a}$  a matriz que contém em suas colunas os vetores  $z_{\mathbf{a}}$  que pertencem à classe  $\mathcal{A}$  e  $Z_{\mathbf{b}} \in \mathbb{R}^{r \times n_b}$  idem para a classe  $\mathcal{B}$ , com  $n_e = n_a + n_b$ .

Sejam  $m_{\mathbf{a}} = \frac{1}{n_a} \sum_{i=1}^{n_a} z_{\mathbf{a}_i}$  e  $m_{\mathbf{b}} = \frac{1}{n_b} \sum_{i=1}^{n_b} z_{\mathbf{b}_i}$  a média entre os vetores para as classes  $\mathbf{a}$  e  $\mathbf{b}$ , respectivamente. Definem-se as matrizes  $S_{\mathbf{a}} = Z_{\mathbf{a}} Z_{\mathbf{a}}^T - m_{\mathbf{a}} m_{\mathbf{a}}^T$  e  $S_{\mathbf{b}} = Z_{\mathbf{b}} Z_{\mathbf{b}}^T - m_{\mathbf{b}} m_{\mathbf{b}}^T$  que são as matrizes de espalhamento para cada uma das classes, respectivamente (KEINOSUKE, 1990). Então, o vetor de pesos  $w$  é obtido por:

$$w = (S_{\mathbf{a}} + S_{\mathbf{b}})^{-1} (m_{\mathbf{a}} - m_{\mathbf{b}}) \quad (3.21)$$

O vetor determinado em Eq. (3.21) é ótimo no sentido de que é o minimizador da função custo, dada pelo quociente de Rayleigh (LOTTE et al., 2007b):

$$J(w) = \frac{w^T S_e w}{w^T S_w w} \quad (3.22)$$

<sup>1</sup> Conjunto de treinamento é um subconjunto dos dados utilizados para obtenção de parâmetros do modelo de classificação.

onde  $S_e$  é a matriz de espalhamento entre classes, definida por  $S_e = (m_{\mathbf{a}} - m_{\mathbf{b}})(m_{\mathbf{a}} - m_{\mathbf{b}})^T$  e  $S_w$  é a matriz de espalhamento intraclasse, definida por  $S_w = S_{\mathbf{a}} + S_{\mathbf{b}}$ .

Em geral, o vetor de projeção  $w$  opera como um maximizador da distância entre os valores médios das classes e, ao mesmo tempo, minimizador dos espalhamentos intraclasse no espaço de decisão (DUDA et al., 2001).

Ao se determinar  $w$ , como na equação Eq. (3.21), um hiperplano é formado através da Eq. (3.20). Assim, a decisão sobre a que classe um dado vetor de entrada  $z_{\rho_i}$  pertence, pode ser tomada com base em um limiar dado pelo parâmetro  $\ell$ :

$$f(z_{\rho_i}) = \begin{cases} w^T z_{\rho_i} \geq \ell, & \text{classe } \mathcal{A} \\ w^T z_{\rho_i} < \ell, & \text{classe } \mathcal{B} \end{cases}$$

O limiar de decisão  $\ell$  associado a Eq. (3.21) é dado por:

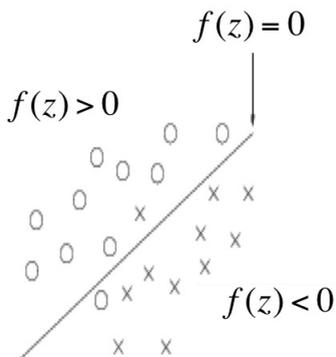
$$\ell = \frac{1}{2} w^T (m_{\mathbf{a}} + m_{\mathbf{b}}) \quad (3.23)$$

Uma limitação do algoritmo LDA é a baixa capacidade de generalização (LI; ANG; GUAN, 2010). Uma vez que o vetor de pesos  $w$  é determinado utilizando vetores  $z_{\rho_i}$  que pertencem ao conjunto de treinamento, os resultados obtidos quando se aplica  $w$  ao conjunto de teste<sup>2</sup> é deteriorado.

A solução encontrada é ótima quando os vetores de entrada apresentam distribuição normal (gaussiana) (DUDA et al., 2001). Para o caso binário, conclui-se a partir da equação Eq. (3.21) que os vetores que representam o valor médio para cada uma das classes devem ser diferentes, pois do contrário o vetor  $w$  será nulo (KEINOSUKE, 1990). Assim, para distribuições com matrizes de covariância diferentes, mas valores médios iguais, o algoritmo LDA não retornará resultados adequados de classificação.

<sup>2</sup> Conjunto de teste é subconjunto dos dados utilizado para avaliar o modelo obtido, nesse caso, o vetor de pesos  $w$  dado pela equação Eq. (3.21)

Figura 12: Esboço do hiperplano de separação linear entre duas classes via LDA. Adaptado de (LOTTE et al., 2007b)



Em geral, o LDA é robusto, fácil de usar com baixa complexidade, tem requisitos computacionais muito baixos e geralmente fornece bons resultados. Ele tem sido utilizado dentro da comunidade ICM por muitos grupos de pesquisa, alguns dos quais são discutidos em referências (LOTTE et al., 2007b)

### 3.5.2 Classificação baseada na distância de Riemann

O algoritmo CSP utiliza as matrizes de covariância dos sinais de EEG para determinar uma transformação linear que realça propriedades dos sinais e, como consequência, a etapa de classificação é facilitada. Uma abordagem alternativa ao uso de filtros espaciais CSP e classificadores lineares do tipo LDA, é utilizar ferramentas da geometria Riemanniana (BARACHANT et al., 2012) para classificar os sinais.

Isso se faz possível, pois matrizes de covariância apresentam propriedades estruturais interessantes. Matrizes de covariância são por construção matrizes simétricas e positivas semi-definidas e portanto, formam um hipercone não-Euclidiano (BHATIA, 2013).

Do arcabouço de geometria Riemanniana, tem-se que variedades são espaços topológicos em que localmente são similares a espaços Euclidianos e globalmente possuem uma estrutura diferencial bem definida (JUNIOR; GIRALDI, 2012).

O espaço de matrizes de covariância é uma variedade Riemanniana (BHATIA, 2013) e com isso, as ferramentas da geometria Riemanniana mostram-se adequadas para se trabalhar com matrizes de covariância, pois consideram a geometria intrínseca que os dados carregam (BARACHANT et al., 2012).

### 3.5.2.1 Ferramentas da Geometria Diferencial

Seja  $\mathcal{M}^n$  variedade de Riemann de dimensão  $n$ . Considere  $\mathcal{S}^n = \{S \in \mathbb{R}^{n \times n}, S = S^T\}$  o espaço de matrizes simétricas de ordem  $n$  e  $\mathcal{P}^n = \{P \in \mathcal{S}^n, P \succ 0\}$  o subconjunto de matrizes simétricas positivas definidas (SPD). O espaço de matrizes positivas definidas forma uma variedade diferenciável em que métricas de Riemann são válidas, portanto é denominado de variedade de Riemann (BARACHANT et al., 2012; MOAKHER, 2005).

Seja  $\Omega \in \mathcal{P}^n$  um elemento da variedade de Riemann. Existe um plano tangente à  $\Omega$  representado por  $T_\Omega \mathcal{M}$  que forma um espaço Euclidiano de matrizes simétricas. Seja  $S_1, S_2 \in T_\Omega \mathcal{M}$  duas matrizes simétricas que pertencem ao plano tangente à  $\Omega$ . Considere  $s_i \triangleq \text{vec}\{S_i\}$  para  $i = 1, 2$  os vetores obtidos a partir das colunas da matriz  $S_i$  concatenadas.

O produto interno entre dois vetores que pertencem ao plano tangente é definido por,

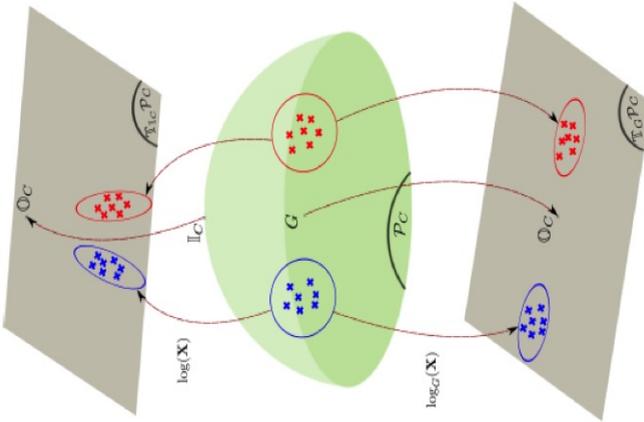
$$\langle s_1, s_2 \rangle_\Omega \triangleq \text{tr}\{\Omega^{-1} S_1 \Omega^{-1} S_2\} \quad (3.24)$$

onde  $\text{tr}\{A\}$  representa o operador traço da matriz  $A$ .

O mapeamento de elementos da variedade de Riemann para um dado plano tangente, e vice-versa, é definido em (BARACHANT et al., 2012). Na Figura 13 é possível observar um esboço de uma variedade

e dois mapeamentos sendo realizados para planos tangentes (YGER, 2013).

Figura 13: Esboço de uma Variedade de Riemann e espaços tangentes associados. Observa-se que para dois planos tangentes tem-se mapeamentos diferentes. Adaptado de (YGER, 2013)



Há um mapeamento único entre os elementos do espaço vetorial  $T_{\Omega}\mathcal{M}$  para a variedade de Riemann. Seja  $S \in T_{\Omega}\mathcal{M}$  e  $\Phi, \Omega \in \mathcal{M}$  o mapeamento  $T_{\Omega}\mathcal{M} \mapsto \mathcal{M}$  é definido por

$$\Phi \triangleq \Omega^{\frac{1}{2}} \exp(\Omega^{-\frac{1}{2}} S \Omega^{-\frac{1}{2}}) \Omega^{\frac{1}{2}}, \quad (3.25)$$

onde  $S \in \mathcal{S}^n$ . Considere a decomposição espectral de uma matriz simétrica  $A = U\Lambda U^T$ , então  $\exp(A) \triangleq U \exp(\Lambda) U^T$ , assim como  $\ln(A) \triangleq U \ln(\Lambda) U^T$ .

O mapeamento inverso, ou seja, de elementos da variedade de Riemann para o espaço tangente à  $\Omega$  é definido por,

$$S \triangleq \Omega^{\frac{1}{2}} \ln(\Omega^{-\frac{1}{2}} \Phi \Omega^{-\frac{1}{2}}) \Omega^{\frac{1}{2}} \quad (3.26)$$

A menor curva que une dois elementos de uma variedade de Riemann  $\mathcal{M}$  é denominada de geodésica (BARACHANT et al., 2012). Dados dois elementos quaisquer,  $\Phi, \Omega \in \mathcal{M}^n$ , a geodésica  $\Theta_\beta(\Phi, \Omega) \in \mathcal{M}^n$ , entre esses elementos é uma curva parametrizada que se inicia no elemento  $\Phi$  e vai até o elemento  $\Omega$ , ou vice-versa (BARACHANT et al., 2012).

$$\Theta_\beta(\Phi, \Omega) \triangleq \Omega^{\frac{1}{2}} (\Omega^{-\frac{1}{2}} \Phi \Omega^{-\frac{1}{2}})^\beta \Omega^{\frac{1}{2}}, \quad \beta \in [0, 1] \quad (3.27)$$

O parâmetro  $\beta$  pode assumir valores no intervalo  $\beta \in [0, 1]$  tal que  $\Theta_0 = \Omega$  e  $\Theta_1 = \Phi$ .

A distância entre dois elementos de uma variedade de Riemann é dado pelo comprimento da curva geodésica entre esses elementos. O comprimento entre dois elementos é denominado por distância de Riemann e é definida por (MOAKHER, 2005):

$$\delta_R(\Phi, \Omega) \triangleq \|\ln(\Phi^{-1}\Omega)\|_F \quad (3.28)$$

onde  $\|A\|_F$  indica a norma Frobenius da matriz  $A$ .

A distância de Riemann pode ser representada a partir dos autovalores da matriz  $\Phi^{-1}\Omega$ :

$$\delta_R(\Phi, \Omega) \triangleq \sqrt{\sum_{i=1}^n \ln^2(\xi_i)}, \quad (3.29)$$

onde  $\xi_i$  são os autovalores de  $\Phi^{-1}\Omega$ .

Considere o conjunto  $\mathcal{W} = \{\Omega_i \in \mathcal{P}^n, i = 1, \dots, n_r\}$ . A média geométrica entre as matrizes  $\Omega_i \in \mathcal{W}$  deve satisfazer, segundo (MOAKHER, 2005):

$$M_g^* = \operatorname{argmin}_{M_g} \sum_{i=1}^{n_r} \delta_R^2(M_g, \Omega_i) \quad (3.30)$$

Observa-se que a média geométrica das matrizes deve minimizar a soma das distâncias de Riemann ao quadrado entre cada matriz  $\Omega_i$  e a matriz  $M_g$ .

A média geométrica resultante de Eq. (3.30) deve satisfazer a equação matricial não-linear, dada por:

$$\sum_i \ln(\Omega_i^{-1} M_g) \triangleq 0 \quad (3.31)$$

Não há um resultado analítico para o problema em Eq. (3.30), em geral, utilizam-se aproximações resultantes de métodos iterativos (MO-AKHER, 2005).

### 3.5.2.2 Classificador via Ferramentas de Riemann

Observe que as ferramentas da geometria Riemanniana, descritas brevemente na seção anterior, podem ser utilizadas para classificar sinais de EEG. Uma abordagem com esse fim é utilizar as matrizes de covariância dos sinais como elementos da variedade de Riemann (BARACHANT et al., 2012). Lembre-se que as matrizes de covariância são simétricas e positivas semi-definidas por construção, como em Eq. (3.4).

Considere as matrizes de covariância definidas em Eq. (3.5) sem a normalização, então

$$\Omega_{\rho_i} \triangleq \text{cov}\{X_{\rho_i}\} \in \mathbb{R}^{p \times p}, \quad \forall i = 1, \dots, n_a + n_b \quad (3.32)$$

onde  $\rho \in \{\mathbf{a}, \mathbf{b}\}$  são os rótulos que indicam, respectivamente, classe  $\mathcal{A}$  e  $\mathcal{B}$ .

Seja a média geométrica para as matrizes de covariância da classe  $\mathcal{A}$  definidas como,

$$M_{g_{\mathbf{a}}}^* = \underset{M_{g_{\mathbf{a}}}}{\operatorname{argmin}} \sum_{i=1}^{n_a} \delta_R^2(M_{g_{\mathbf{a}}}, \Omega_{\mathbf{a}_i}) \quad (3.33)$$

e para as matrizes de covariância da classe  $\mathcal{B}$ ,

$$M_{g_{\mathbf{b}}}^* = \underset{M_{g_{\mathbf{b}}}}{\operatorname{argmin}} \sum_{i=1}^{n_b} \delta_R^2(M_{g_{\mathbf{b}}}, \Omega_{\mathbf{b}_i}) \quad (3.34)$$

Assim, a matriz resultante da média geométrica entre as matrizes de covariância para cada uma das classes é utilizada como referência para indicar a classe (BARACHANT et al., 2012). Então cada classe é de fato caracterizada pela matriz média entre as matrizes de covariância, como em Eq. (3.33) ou Eq. (3.34). Dada uma matriz de covariância em que se deseja assinalar a classe, basta verificar a distância de Riemann em relação à matriz média de cada classe.

Considere  $\Omega_{\rho}$  uma matriz de covariância que se deseja classificar. Suponha que se tem apenas duas classes, então o rótulo  $\rho \in \{\mathbf{a}, \mathbf{b}\}$  e as matrizes de covariância médias são dadas por  $M_{g_{\mathbf{a}}}$  e  $M_{g_{\mathbf{b}}}$ , para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente. Então a função  $f$  que classifica os sinais é definida por,

$$f(\Omega_{\rho}) = \delta_R(\Omega_{\rho}, M_{g_{\mathbf{a}}}) - \delta_R(\Omega_{\rho}, M_{g_{\mathbf{b}}}) \quad (3.35)$$

e a decisão de classificação é realizada como segue,

$$f(\Omega_{\rho}) = \begin{cases} \delta_R(\Omega_{\rho}, M_{g_{\mathbf{b}}}) \geq \delta_R(\Omega_{\rho}, M_{g_{\mathbf{a}}}), & \text{classe } \mathcal{A} \\ \delta_R(\Omega_{\rho}, M_{g_{\mathbf{b}}}) < \delta_R(\Omega_{\rho}, M_{g_{\mathbf{a}}}), & \text{classe } \mathcal{B} \end{cases}$$

### 3.6 Métodos de avaliação de desempenho de sistemas ICM.

O projeto de um sistema ICM envolve a seleção de vários parâmetros e métodos. Alguns deles foram descritos nas seções anteriores. No entanto, se faz necessário a definição de métricas para a avaliação global do sistema. Métricas que tornem possível a comparação entre configurações diferentes de sistemas ICM, como por exemplo, protocolos de aquisição de sinais e diferentes métodos de pré-processamento e classificação.

Vários métodos para a avaliação de desempenho de sistemas ICM podem ser utilizados (RAO; SCHERER, 2010). Neste trabalho, são descritos dois métodos de avaliação utilizados em sistemas ICM: o método com base na acurácia de classificação e o método baseado na matriz de confusão, ambos descritos na próxima seção.

#### 3.6.1 Acurácia de classificação

A acurácia de classificação é utilizada em muitos trabalhos em sistemas ICM (RAO; SCHERER, 2010). Ela relaciona o número de exemplos corretamente classificados,  $p_c$  com o número total de exemplos realizados  $p_{tot}$  durante um experimento. A equação Eq. (3.36) calcula a acurácia de classificação para um dado experimento em porcentagem

(%).

$$Acc(\%) = \frac{p_c}{p_{tot}} \times 100 \quad (3.36)$$

A acurácia de classificação é uma medida efetiva do desempenho de sistemas de ICM. No entanto, ela não considera aspectos de sensibilidade, especificidade e seletividade.

### 3.6.2 Matriz de confusão

Como forma de quantificar o desempenho do sistema ICM, são utilizadas métricas estatísticas que consideram a taxa com a qual o sistema de ICM consegue detectar um determinado evento de interesse (tarefa mental) (RAO; SCHERER, 2010). As tentativas, ou épocas, fornecidas ao sistema ICM são agrupadas em quatro categorias distintas, de acordo com a marcação real dos dados e a marcação realizada pelo sistema ICM. Alguns rótulos utilizados são:

- **Verdadeiro Positivo (VP)**: Evento ocorreu; Sistema ICM identificou evento
- **Verdadeiro Negativo (VN)**: Evento não ocorreu; Sistema ICM não identificou evento
- **Falso Negativo (FN)**: Evento ocorreu; Sistema ICM não identificou evento
- **Falso Positivo (FP)**: Evento não ocorreu; Sistema ICM identificou evento

O conjunto VP e VN agrupa os acertos do sistema ICM, enquanto que FN e FP agrupa os erros. Em seguida, são calculadas métricas de desempenho de acordo com os grupos formados:

$$Acc = \frac{VP + VN}{VP + VN + FN + FP} \quad (3.37)$$

$$Seletividade = \frac{VP}{VP + FP} \quad (3.38)$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3.39)$$

A equação Eq. (3.37) expressa a taxa de acerto global do sistema ICM, ou seja, o número de acertos dividido pelo número total de tentativas. A seletividade, definida na equação Eq. (3.38), mostra a taxa de acerto do sistema para tentativas em que ocorreu o evento. A expressão Eq. (3.39) estima a taxa de acurácia do algoritmo ao classificar uma tentativa como uma em que ocorreu um evento. Em outras palavras, a sensibilidade expressa a capacidade do classificador em identificar a ocorrência do evento.

As duas métricas apresentadas nesta seção representam apenas alguns fatores que podem ser considerados na avaliação de sistemas ICM. Outros fatores importantes que podem ser utilizados são a necessidade de procedimentos invasivos ou não, o tempo necessário para o ajuste dos algoritmos, a facilidade e o conforto da utilização dos equipamentos, entre outros. A importância desses fatores podem variar de acordo com a aplicação final do sistema ICM (RAO; SCHERER, 2010).

### 3.7 Conclusões do Capítulo

Nesse Capítulo, um conjunto de algoritmos básicos que podem ser aplicados para implementar um sistema ICM foi apresentado. Os algoritmos foram apresentados na mesma seqüência com que eles se encontram na cadeia de processamento do sinal de EEG. Embora a quantidade de algoritmos que pudessem ser apresentados nesse Capítulo seja diverso, apenas os algoritmos que foram utilizados para comparação dos métodos propostos foram descritos.

No próximo Capítulo, dá-se início à apresentação do método proposto nesse trabalho. Os sinais de EEG serão projetados numa base de sinais senoidais e, com isso, uma representação compacta dos sinais será definida e utilizada para os desenvolvimentos futuros.



# 4 Representação Compacta de Sinais de EEG

## 4.1 Introdução

Nesse Capítulo é apresentado o procedimento para a obtenção de uma representação compacta dos sinais de EEG. Essa nova representação desponta como uma alternativa à representação original por amostras do sinal no tempo.

A representação compacta é obtida a partir da projeção dos sinais de EEG em uma base de sinais ortogonais formada por senoides. Cada elemento da base está associado a uma única frequência. Assim na representação obtida a informação sobre o conteúdo frequencial dos sinais é expresso de forma explícita.

A partir da representação compacta e da base de sinais é possível reconstruir a representação original de forma aproximada. O erro cometido por essa aproximação é minimizado no sentido de mínimos quadrados.

Outro aspecto interessante a ser ressaltado está no fato de que quando apenas um subconjunto de sinais da base e os respectivos termos da representação compacta são considerados, a aproximação obtida corresponde aos sinais originais filtrados na mesma faixa de frequências que contém as frequências correspondentes ao subconjunto.

No processamento de sinais de EEG em sistemas ICM, a primeira etapa realizada consiste na filtragem frequencial dos sinais. A finalidade é selecionar faixas de frequências relacionadas à atividade cerebral de interesse, assim como atenuar a influência de ruídos e artefatos.

Em geral, filtros lineares convencionais são empregados para

realizar a filtragem dos sinais de EEG, retornando sinais filtrados com a mesma dimensão dos sinais originais. Ou ainda, quando faixas de frequências não consecutivas são de interesse, diferentes filtros, um correspondente a cada faixa, podem ser utilizados. Em ambas as situações, a representação compacta dos sinais desponta como alternativa viável.

## 4.2 Breve revisão sobre produto interno e energia de sinais

Considere dois sinais reais  $z(n)$  e  $b(n)$  no intervalo  $\Delta = (n_1, n_q)$ . O produto interno entre  $z(n)$  e  $b(n)$  em  $\Delta$  é definido como:

$$\langle z(n), b(n) \rangle_{\Delta} \triangleq \sum_{n=n_1}^{n_q} z(n)b(n) \quad (4.1)$$

Sinais que apresentam o produto interno nulo são denominados sinais ortogonais no intervalo  $\Delta$ . O produto interno é utilizado para definir a energia de um sinal: para um sinal  $z(n)$  a energia é definida como:

$$E_z = \langle z(n), z(n) \rangle_{\Delta} = \sum_{n=n_1}^{n_q} z^2(n) \quad (4.2)$$

Agora considere a aproximação do sinal  $z(n)$  realizada pelo sinal  $b(n)$  ponderado por uma constante  $\gamma$  no intervalo  $\Delta$ , tal que:

$$z(n) \approx \gamma b(n); \quad n_1 < n < n_q \quad (4.3)$$

onde  $\gamma \in \mathbb{R}$  (constante). Defina o sinal de erro de aproximação  $e_a(n)$  entre  $z(n)$  e  $\gamma b(n)$  no intervalo  $\Delta$ , dado por,

$$e_a(n) = z(n) - \gamma b(n); \quad n_1 < n < n_q \quad (4.4)$$

Uma medida do tamanho do sinal de erro de aproximação é dada pela energia do sinal  $e_a(n)$  no intervalo  $\Delta$ ,

$$E_e = \sum_{n=n_1}^{n_q} e_a^2(n) \quad (4.5)$$

Observa-se que a energia do erro  $E_e$  é função apenas da constante  $\gamma$ , logo o valor da constante  $\gamma$  que torna a energia do sinal de

erro de aproximação tão pequena quanto possível pode ser determinada resolvendo-se a equação,

$$\frac{dE_e}{d\gamma} = 0 \quad (4.6)$$

A constante  $\gamma$  que satisfaz Eq. (4.6) é obtida, primeiramente, substituindo-se a equação Eq. (4.4) na equação Eq. (4.5). Em seguida, substitui-se o resultado em Eq. (4.6) e assim obtém-se a constante  $\gamma$ , dada por:

$$\gamma^* = \frac{\langle z(n), b(n) \rangle_{\Delta}}{\langle b(n), b(n) \rangle_{\Delta}} \quad (4.7)$$

A melhor aproximação do sinal  $z(n)$  a partir do sinal  $b(n)$  é dada por  $\hat{z}(n) = \gamma^* b(n)$ , onde  $\gamma^*$  é dado por Eq. (4.7) no intervalo  $\Delta$  no sentido de Eq. (4.6).

Em analogia a vetores, o sinal  $\hat{z}(n)$  pode ser interpretado como a projeção do sinal  $z(n)$  na direção  $b(n)$ , assim o coeficiente  $\gamma^*$  corresponde ao coeficiente da projeção entre os sinais.

### 4.2.1 Base Ortogonal de Sinais

Na seção anterior, foi visto como determinar uma aproximação de um sinal a partir de outro sinal ponderado, de tal forma que a energia do sinal de erro de aproximação entre os dois sinais foi minimizada. Algo semelhante, será realizado nesta seção. No entanto, um dos sinais será considerado como uma combinação linear de sinais ortogonais.

Considere o conjunto  $\mathcal{O} = \{b_1(n), \dots, b_{n_b}(n)\}$ , com  $n_b$  sinais reais no intervalo  $\Delta = (n_1, n_q)$ , tal que,

$$\langle b_k(n), b_l(n) \rangle_{\Delta} \triangleq 0, \quad \forall k \neq l \quad (4.8)$$

para  $k, l = 1, 2, \dots, n_b$ . Note que para  $k = l$ , tem-se a energia do sinal  $b_k(n)$ . Logo, o conjunto  $\mathcal{O}$  que satisfaz a Eq. (4.8) é formado por sinais ortogonais.

Considere o problema de determinar a "melhor" aproximação entre um sinal real  $z(n)$  e uma combinação linear de  $m$  sinais ortogonais

do conjunto  $\mathcal{O}$ .

$$z(n) \approx \sum_{k=1}^m \gamma_k b_k(n) \quad (4.9)$$

onde  $m \leq n_b$  e  $\gamma_k \in \mathbb{R}$  são constantes para  $k = 1, 2, \dots, m$ . Defina o erro de aproximação entre esses sinais por:

$$e_a(n) = z(n) - \sum_{k=1}^m \gamma_k b_k(n) \quad (4.10)$$

Novamente, a energia do sinal de erro de aproximação em Eq. (4.10) será utilizada como métrica da qualidade da aproximação em Eq. (4.9), assim o conjunto de coeficientes  $\gamma_k$  que minimiza a energia do sinal de erro,  $E_e$ , em Eq. (4.10) é obtida quando

$$\frac{\partial E_e}{\partial \gamma_k} = 0, \quad k = 1, 2, \dots, m \quad (4.11)$$

é satisfeita. Então os coeficientes  $\gamma_k$  podem ser obtidos a partir de

$$\gamma_k^* = \frac{\langle z(n), b_k(n) \rangle_{\Delta}}{\langle b_k(n), b_k(n) \rangle_{\Delta}} \quad k = 1, 2, \dots, m \quad (4.12)$$

Observa-se que o termo no denominador em Eq. (4.12) representa a energia do sinal  $b_k(n)$  do conjunto  $\mathcal{O}$  no intervalo  $\Delta$  e é independente do sinal  $z(n)$ . Logo, quando os sinais do conjunto  $\mathcal{O}$  satisfazem a Eq. (4.8) e,

$$\langle b_k(n), b_l(n) \rangle_{\Delta} = 1, \quad \forall k = l \quad (4.13)$$

o conjunto  $\mathcal{O}$  é denominado ortonormal.

Uma vez que os elementos do conjunto de sinais são mutuamente ortogonais, pois satisfazem Eq. (4.8), a determinação dos coeficientes  $\gamma_k$  via Eq. (4.12) é independente para cada elemento do conjunto  $\mathcal{O}$ . Isto implica que se considerarmos mais sinais, ou seja,  $m + 1 \leq n_b$  do conjunto dos sinais  $\mathcal{O}$ , a energia do sinal de erro será reduzida e o novo coeficiente a ser determinado,  $\gamma_{m+1}$ , não perturbará os valores dos coeficientes determinados anteriormente,  $\{\gamma_1, \gamma_2, \dots, \gamma_m\}$ .

Importante notar que, uma vez que o conjunto de sinais  $\mathcal{O}$  com  $n_b$  elementos é dado, ele forma uma base de aproximação de dimensão

$n_b$  para o espaço de sinais. Logo, os coeficientes  $\gamma_k$  são uma representação do sinal real  $z(n)$  na base de sinais formada pelo conjunto  $\mathcal{O}$ .

Se para um dado número de elementos da base de sinais,  $n_b$ , a energia do sinal de erro de aproximação é zero, o conjunto de sinais da base é denominado conjunto completo. Neste caso,  $\mathcal{O}$  é de fato uma base.

### 4.2.2 Base de Sinais Senoidais

Sinais senoidais são conhecidos na literatura por formarem um conjunto de sinais ortogonais, onde cada elemento que pertence a esse conjunto possui associado uma informação de frequência (DURKA, 2007).

Considere que os elementos do conjunto de sinais  $\mathcal{O}$  são definidos no intervalo  $\Delta = (n_1, n_q)$ , como segue:  $b_k(n) = \sin(\omega_k n)$  e  $b_{k+m} = \cos(\omega_k n)$ ,  $k = 1, 2, \dots, m$ , com  $\omega_k = 2\pi \frac{fk}{f_s}$  e  $f_s$  é a frequência de amostragem considerada, onde o número total de elementos da base de sinais senoidais é dado por  $n_b = 2m$  e caracterizam  $m$  frequências distintas.

A aproximação de um sinal real  $z(n)$  por uma combinação linear de elementos do conjunto  $\mathcal{O}$  é dada como segue,

$$z(n) \approx \hat{z}(n) = \sum_{k=1}^m \gamma_k \sin(\omega_k n) + \gamma_{k+m} \cos(\omega_k n) \quad (4.14)$$

e perceba que o par de coeficientes  $\{\gamma_k, \gamma_{k+m}\}$ , que reduzem a energia do sinal de erro de aproximação, pode ser determinado como descrito nas seções anteriores, repetido aqui por conveniência:

$$\gamma_k = \frac{\langle z(n), \sin(\omega_k n) \rangle_{\Delta}}{\langle \sin(\omega_k n), \sin(\omega_k n) \rangle_{\Delta}} \quad e \quad \gamma_{k+m} = \frac{\langle z(n), \cos(\omega_k n) \rangle_{\Delta}}{\langle \cos(\omega_k n), \cos(\omega_k n) \rangle_{\Delta}} \quad (4.15)$$

para  $k = 1, 2, \dots, m$ .

Como mencionado anteriormente, o conjunto de coeficientes obtido via Eq. (4.15) representa o sinal  $z(n)$  na base de sinais senoidais.

### 4.3 Projção dos sinais de EEG na base Senoidal

Conforme mencionado no Capítulo 3, durante a etapa de coleta dos sinais de EEG um conjunto de sensores é posicionado no couro cabeludo do usuário do sistema de ICM. Essa configuração dos sensores caracteriza a natureza multivariada do sinal de EEG o que nos permite uma representação vetorial do sinal. Considere o vetor  $\mathbf{z}(n) = [z_1(n), z_2(n), \dots, z_p(n)]^T$  ser o sinal de EEG bruto coletado em  $p$  sensores no instante de tempo discreto  $n$ .

Seja  $Z \in \mathbb{R}^{p \times q}$  uma matriz de posto completo que representa um segmento de sinal de EEG com  $p$  sensores e  $q$  amostras por sensor, cada um registrado a uma taxa de amostragem  $f_s$ . O segmento  $Z$  considerado aqui é suposto não ter recebido nenhum processamento sendo portanto, denominado de sinal bruto.

Seja  $X \in \mathbb{R}^{p \times q}$  uma matriz de posto completo representando as amostras do sinal  $Z$  de EEG, após um filtro passa-faixa ideal com faixa de frequência de interesse  $\Delta f = (f_l, f_u)$ , onde  $f_l$  e  $f_u$  são, respectivamente, as frequências de corte inferior e superior do filtro. Considere a partição  $X = [\mathbf{x}(1) \cdots \mathbf{x}(q)]$  em colunas, onde cada coluna de  $X$  é dada pelo vetor  $\mathbf{x}(n) = [x_1(n) \cdots x_p(n)]^T$ . A matriz  $X$  é denominada como o segmento do sinal de EEG filtrado, como mostrado na Figura 15a.

Seja ainda,  $B \in \mathbb{R}^{2m \times q}$  uma matriz de posto completo contendo  $2m$  sinais que formam a base de sinais senoidais, sendo  $m$  sinais senoidais e  $m$  sinais cossenoidais, respectivamente. Um exemplo da base de sinais é mostrado na Figura 17b. Cada sinal da base contém  $q$  amostras, coletadas com a frequência de amostragem  $f_s$ , de um segmento de sinal da base  $\mathbf{b}(n)$  como:

$$\mathbf{b}(n) = \begin{bmatrix} b_1(n) \\ \vdots \\ b_{2m}(n) \end{bmatrix}; \begin{cases} b_k(n) = \sin\left(\frac{2\pi f_k}{f_s} n\right) \text{ para } k = 1, \dots, m \\ b_{k+m}(n) = \cos\left(\frac{2\pi f_k}{f_s} n\right) \text{ para } k = 1, \dots, m \end{cases} \quad (4.16)$$

onde  $n \in \{1, \dots, q\}$  define o tempo discreto.

Seja o conjunto  $\mathcal{F} = \{f_1, \dots, f_k, \dots, f_m\}$  contendo as frequências discretas na faixa de frequências de interesse, tal que  $f_l \leq f_k \leq f_u$ . Supõe-se que  $f_u < \frac{f_s}{2}$  e é definida uma resolução em frequência  $\delta f \triangleq f_{k+1} - f_k = \frac{f_s}{q}$ . Note que  $\delta f$  é o menor intervalo de frequência tal que os elementos que compõem a base de sinais  $B$  sejam ortogonais (DURKA, 2007).

Considere o problema de determinar a "melhor" aproximação do segmento de sinal de EEG bruto  $Z$  usando o segmento da base de sinais senoidais  $B$ , como indicado abaixo

$$\hat{X} = \Gamma B, \quad (4.17)$$

onde a matriz  $\Gamma \in \mathbb{R}^{p \times 2m}$  representa a matriz de coeficientes da aproximação e  $\hat{X} \in \mathbb{R}^{p \times q}$ . As linhas de  $\Gamma$  estão associadas aos  $p$  sensores do sinal  $Z$  e as colunas de  $\Gamma$  estão relacionadas aos  $m$  sinais da base distintos em frequência.

A matriz de coeficientes  $\Gamma$  é determinada resolvendo-se o problema de minimizar a função custo definida na Eq. (4.18).

$$\Gamma^* = \arg \min_{\Gamma} \|Z - \Gamma B\|_F^2 \quad (4.18)$$

onde o asterisco  $*$  é utilizado para indicar a solução ótima e  $\|A\|_F$  indica a norma Frobenius da matriz  $A$ .

O problema expresso na equação Eq. (4.18) tem solução na forma exata,

$$\Gamma^* = ZB^T(BB^T)^{-1} \quad (4.19)$$

onde o  $A^T$  indica a operação de transposição da matriz  $A$ . Para simplificar a notação, o asterisco  $*$  será omitido deste ponto em diante do Capítulo.

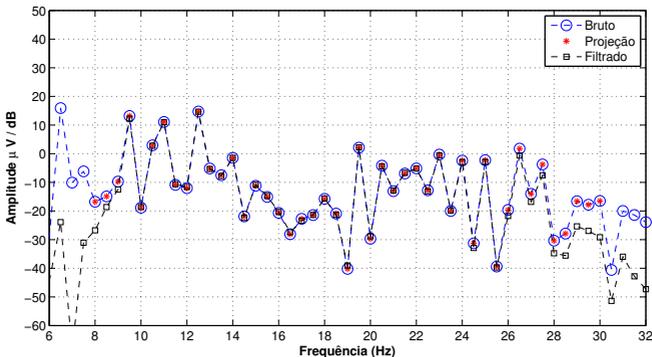
Observe que ao substituir a Eq. (4.19) na expressão  $\hat{X} = \Gamma B$  obtém-se uma estimativa ótima de  $Z$ , no sentido de minimizar a Eq. (4.18), a partir do conjunto de sinais que formam a base  $B$ , que é dada por:

$$\hat{X} = ZB^T(BB^T)^{-1}B \quad (4.20)$$

O sinal  $\hat{X}$  é o sinal resultante da projeção do sinal  $Z$  na base de sinais senoidais  $B$ . Observe ainda que o conteúdo frequencial do sinal  $\hat{X}$ , na Eq. (4.20), é definido pelos  $m$  elementos da base de sinais senoidais  $B$ . Por sua vez, os  $m$  elementos de  $B$  são amostras discretas da banda passante do filtro ideal. Assim, o espectro do sinal  $\hat{X}$  corresponde ao espectro do sinal  $Z$  para as frequências definidas pela base de sinais  $B$ , como pode ser visto na Figura 14.

Um aspecto interessante a ser notado, é que o espectro do sinal  $X$ , resultante da filtragem ideal na banda passante  $\Delta f$ , corresponde por definição, ao espectro do sinal  $Z$ , para as frequências contidas em  $\Delta f$ . Nesse sentido, há uma correspondência entre os espectros dos sinais  $X$ ,  $Z$  e  $\hat{X}$  para as frequências discretas que formam o conjunto  $\mathcal{F} \subset \Delta f$ .

Figura 14: Exemplo de espectro de amplitude dos sinais  $Z$ (bruto),  $\hat{X}$ (projeção) e  $\tilde{X}$  (filtrado, filtro não-ideal). A base de sinais senoidais foi definida com 45 frequências diferentes, igualmente espaçadas ( $\delta f = \frac{1}{2}$ ) entre as frequências de corte inferior e superior, 8 e 30Hz, respectivamente. Nota-se que os sinais  $Z$ (marcador círculo e cor azul) e  $\hat{X}$  (marcador asterisco e cor vermelha) apresentam o mesmo conteúdo frequencial para as frequências da base de sinais  $B$ . O sinal  $\tilde{X}$  (marcador quadrado e cor preta) foi obtido na saída de um filtro passa-faixa realizável, portanto não-ideal. Por essa razão, o espectro de  $\tilde{X}$  é diferente do espectro de  $Z$  e  $\hat{X}$ , principalmente, para as frequências próximas as frequências de corte. Especificadamente, para as frequências abaixo de 8,5 Hz e acima de 28 Hz.

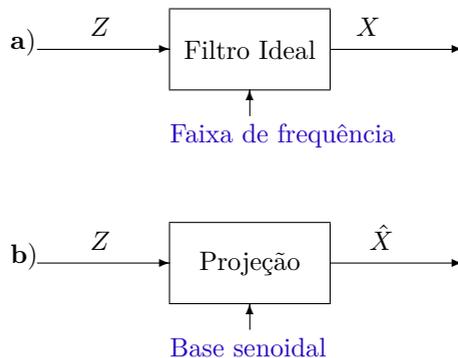


Em síntese, quando a base senoidal está definida na faixa de frequência  $\Delta f$ , com resolução em frequência dada por  $\delta f = \frac{f_s}{q}$ , e o sinal de EEG bruto é filtrado por um filtro passa-faixa ideal definido na mesma faixa de frequência  $\Delta f$ , tanto os sinais obtidos da projeção de  $Z$  em  $B$ , como na Eq. (4.20), quanto os obtidos por filtragem ideal  $X$ , compartilham o mesmo conteúdo espectral, para as  $m$  frequências que formam a base de senoides.

Nota-se ainda que, um filtro ideal é não-realizável, portanto, para fins de ilustrar as propriedades mencionadas, é utilizado um filtro não-ideal do tipo *Butterworth* de quarta ordem. O sinal  $Z$  filtrado pelo filtro passa-faixa não-ideal será representado por  $\tilde{X}$ . Para ilustrar este ponto, na Figura 14 tem-se um exemplo do espectro dos sinais  $Z$ ,  $\hat{X}$  e  $\tilde{X}$ .

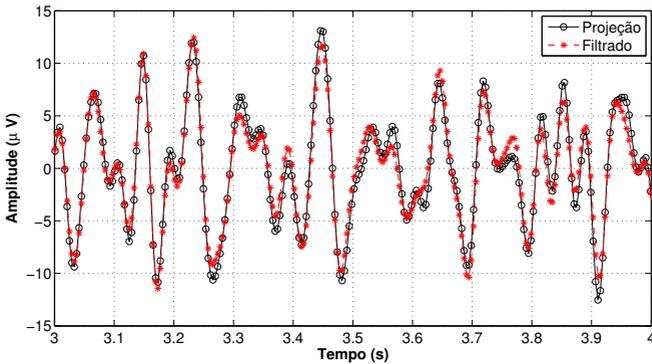
O problema definido na Eq. (4.18) é formulado de tal forma a minimizar o erro entre os sinais  $Z$  e  $\hat{X}$ . Note que os sinais filtrados,  $X$  e  $\tilde{X}$  não são considerados em Eq. (4.18). No entanto, o erro de aproximação no tempo entre os sinais  $\hat{X}$  e  $\tilde{X}$  pode ser aferida pela energia do erro médio quadrático relativo, isto é, calculando-se:  $\frac{\|\tilde{X} - \hat{X}\|^2}{\|\tilde{X}\|^2}$ . As Figuras 15 e 16 ilustram esse cenário.

Figura 15: a) O sinal  $X$  é a versão idealmente filtrada do sinal  $Z$  na faixa de frequência de interesse  $\Delta f = (f_i, f_u)$ . b) O sinal  $\hat{X}$  é uma projeção do sinal  $Z$  na base de sinais  $B$  definida na faixa de frequência de interesse.



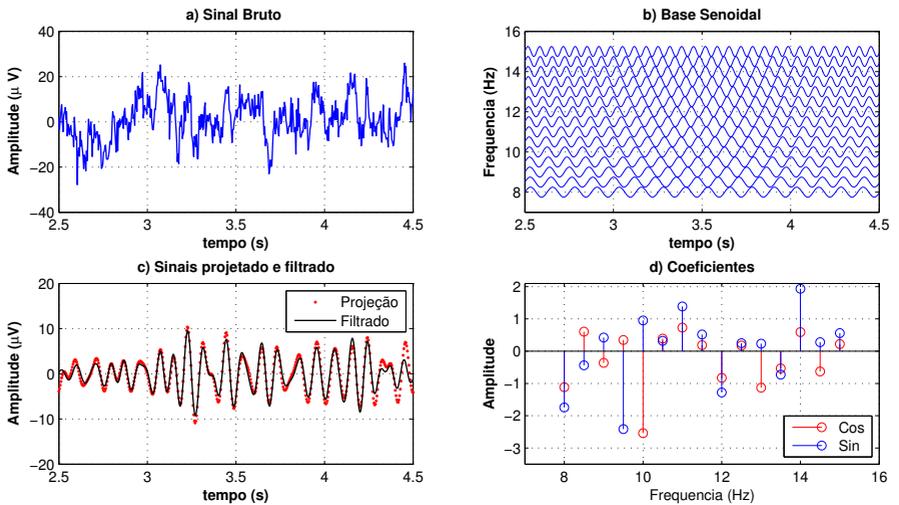
Na Figura 16 é mostrada uma comparação no tempo entre dois sinais unidimensionais selecionados das matrizes  $\tilde{X}$  e  $X$ . Estes sinais foram obtidos, respectivamente, do sinal  $Z$  por filtragem (não-ideal) e por projeção, como na Figura 15, utilizando a mesma faixa de frequências para o filtro passa-faixa e para a base de senoides. A partir da Figura 16 pode-se notar que o erro de aproximação entre  $\tilde{X}$  e  $\hat{X}$  é pequeno. Para o conjunto de sinais testados, obteve-se erro quadrático médio normalizado inferior a 10%.

Figura 16: Um segmento de um canal dos sinais filtrado e projetado é mostrado em vermelho (com marcador estrela) e em preto (com marcador círculo), respectivamente. O erro quadrático médio entre esses sinais é:  $MSE = 1.6114$  que corresponde à 7,34% da energia do sinal filtrado.



A Figura 17 está sub-dividida em quatro quadros. No quadro a) tem-se um esboço de um sinal de EEG (unidimensional) com duração de 2 segundos, amostrado a  $f_s = 250Hz$ , denominado sinal bruto. No quadro b) tem-se o esboço de uma base de sinais senoidais, com a mesma duração no tempo do sinal bruto e contendo as frequências discretas entre 8 e 15  $Hz$ , igualmente espaçadas de  $\delta f = 0,5Hz$ . No quadro c) tem-se dois sinais, são eles: o sinal filtrado (filtro passa-faixa com banda passante entre 8 e 15  $Hz$ ) e o sinal resultante da projeção do sinal do quadro a) na base de sinais senoidais do quadro b). Finalmente, no quadro d) são apresentados os coeficientes obtidos ao se projetar o

Figura 17: a) Sinal de EEG bruto obtido do sensor localizado em  $C_z$  com  $q = 500$  amostras ( $f_s = 250Hz$ ). b) Esboço dos termos senoidais da base de sinais senoidais ( $m = 15$ ), no intervalo entre as frequências  $f_1 = 8$  e  $f_m = 15$  Hz com  $\delta f = 0,5$ . c) Em vermelho e marcador círculo, a projeção de do sinal de EEG na base de sinais. Em preto e marcador contínuo, sinal de EEG filtrado por um filtro passa-faixa com frequências de corte  $f_l = 8Hz$  e  $f_u = 15Hz$ , respectivamente. d) Coeficientes obtidos da projeção do sinal em a) na base de sinais em b). As cores (vermelho e azul) diferenciam os coeficientes que correspondem aos termos cossenoidais e senoidais, respectivamente, da base de sinais.



sinal do quadro a) na base de sinais senoidais do quadro b).

Na Figura 17, particularmente, nos itens c) e d) é possível verificar a redução obtida na representação do sinal. Em c) tem-se um sinal com  $q = 500$  amostras e em d) um conjunto de  $2m = 30$  coeficientes. Nesse sentido, no Capítulo 5 a matriz de coeficientes  $\Gamma$  é utilizada para representar os sinais de EEG de forma compacta.

O erro de aproximação entre os sinais filtrado e a projeção do sinal na base pode ser observado na Figura 17 c). Nota-se que a ope-

ração de projeção do sinal bruto numa base de sinais senoidais é capaz de extrair informação frequencial dos sinais contida nos coeficientes resultantes da projeção.

#### 4.4 Interpretação da Matriz de Coeficientes e Sub-bandas

Os elementos da matriz  $\Gamma$  correspondem aos coeficientes da base senoidal Eq. (4.16) conduzindo à melhor aproximação do sinal bruto  $Z$  em relação à base  $B$ , no sentido de minimizar a Eq. (4.18). Observe que  $B$  é usado para especificar a banda de frequência de interesse.

Cada par de colunas da matriz  $\Gamma$  pode ser visto como a projeção dos  $p$  canais de  $Z$  em termos de um seno e um cosseno de uma frequência específica.

Para enfatizar esse ponto, denotemos  $\gamma(j) \in \mathbb{R}^p$  como  $j$ -ésima coluna da matriz  $\Gamma$  e a partição de  $\Gamma$  como mostrado abaixo:

$$\Gamma = \begin{bmatrix} \gamma(1) & \cdots & \gamma(m) & \gamma(m+1) & \cdots & \gamma(2m) \end{bmatrix} \quad (4.21)$$

Os pares de colunas  $\{\gamma(j), \gamma(j+m)\}$  para todos os índices  $j = 1, \dots, m$  representam os coeficientes de uma senoide e cossenoide de frequência  $f_j/f_s$  na reconstrução dos  $p$  sinais de  $Z$  a partir da base senoidal definida na Eq. (4.16). Portanto, as colunas da matriz  $\Gamma$  contêm informações sobre o conteúdo de frequência dos canais. Como será visto no Capítulo 8, esta propriedade de  $\Gamma$  será útil para definir níveis de importância para cada componente de frequência dos sinais de EEG.

Além disso, a matriz  $\Gamma$  tende a ter dimensão muito menor do que o segmento de sinal  $Z$  porque normalmente  $2m$  é pequeno quando comparado a  $q$  (DURKA, 2007). Para ilustrar este ponto, é mostrada na Figura 17 c), a representação do sinal filtrado no domínio do tempo  $\tilde{X}$  (amostras de sinal) e na Figura 17 d) os elementos da matriz de coeficientes,  $\gamma$ , no domínio da frequência. Observe que a matriz de coeficientes pode representar o sinal EEG filtrado de uma forma mais compacta, sem perda relevante de informação, como mostrado na Figura 17.

Conforme explicado anteriormente, as informações de frequência do sinal  $Z$  na faixa de frequência de interesse podem ser obtidas a partir das colunas da matriz  $\Gamma$ .

Na sequência, expressamos a matriz de covariância dos sinais resultantes da projeção de  $Z$  na base  $B$  na faixa de frequência de interesse como uma função da matriz de coeficientes  $\Gamma$ .

Seja  $cov\{\hat{X}\}$  a matriz de covariância de  $\hat{X}$  definida como

$$cov\{\hat{X}\} = \frac{1}{q} \hat{X} \hat{X}^T \quad (4.22)$$

Note que ao substituir a igualdade  $\hat{X} = \Gamma B$  na Eq. (4.22) obtém-se,

$$cov\{\hat{X}\} = \frac{1}{q} \Gamma B B^T \Gamma^T = \frac{1}{2} \Gamma \Gamma^T \quad (4.23)$$

onde  $\frac{1}{q} B B^T = \frac{1}{2} I_{2m}$ .

A Eq. (4.23) pode ser re-escrita para destacar as colunas da matriz  $\Gamma$  como,

$$cov\{\hat{X}\} = \frac{1}{2} \sum_{j=1}^m \gamma(j) \gamma(j)^T + \gamma(j+m) \gamma(j+m)^T, \quad (4.24)$$

Observe que a base senoidal  $B$  é um conjunto de senoides com frequências no conjunto  $\mathcal{F}$ , como pode ser visto a partir da seção 4.3. Assim, a matriz de covariância na Eq. (4.24) está associada com senoides de frequências entre  $f_l$  e  $f_u$ .

Dada a banda (ou faixa) de frequência definida pelo conjunto  $\mathcal{F}$ , consideremos um conjunto de  $n_s$  sub-bandas, cada uma com largura de banda  $w_b$  e um deslocamento  $\varpi_d$  entre duas sub-bandas consecutivas, isto é,  $w_b - \varpi_d$  é a sobreposição entre duas sub-bandas consecutivas. Para ajustar as sub-bandas  $n_s$  em  $\mathcal{F}$ , o número de sub-bandas  $n_s$  deve satisfazer a relação  $m = w_b + \varpi_d(n_s - 1)$ .

Seja  $cov_\ell\{\hat{X}\}$  a matriz de covariância do sinal  $\hat{X}$  quando apenas os coeficientes da  $\ell$ -ésima sub-banda de frequência é considerado. Neste caso, a matriz de covariância associada a esta sub-banda é dada

por:

$$\text{cov}_\ell\{\hat{X}\} = \frac{1}{2} \sum_{j=1}^{w_b} \gamma_\ell(j)\gamma_\ell(j)^T + \gamma_\ell(j+m)\gamma_\ell(j+m)^T \quad (4.25)$$

onde  $\ell = 1, 2, \dots, n_s$  indica a sub-banda, e  $\gamma_\ell(j) \triangleq \gamma(j + (\ell - 1)\varpi_d)$ , para um dado  $\ell$ , denota os coeficientes da  $\ell$ -ésima sub-banda de frequências, para  $j = 1, 2, \dots, w_b$ .

Note que, para  $\varpi_d = \frac{w_b}{2}$ , temos 50% de sobreposição entre duas sub-bandas consecutivas e obtemos  $n_s = 2\frac{m}{w_b} - 1$  sub-bandas.

As matrizes de covariância definidas na Eq. (4.24) e Eq. (4.25) possuem propriedades de interesse sobre a relação entre os sensores de EEG. Por exemplo, nos elementos da diagonal tem-se a informação de energia média dos sensores que formam o segmento de sinal de EEG. Nos elementos de fora da diagonal da matriz de covariância, há informação de correlação cruzada de energia entre os sensores.

Ao se definir a matriz de covariância para uma sub-banda de frequência, a partir dos coeficientes resultantes da projeção do sinal  $Z$  na base de sinais senoidais  $B$ , como em Eq. (4.25), evidencia-se que o uso dos coeficientes, torna a informação de frequência dos sinais explícita na representação compacta. Além disso, permite selecionar sub-bandas de frequências e suas relações de dependência entre os sensores para sub-faixas de frequência dos sinais através da matriz de covariância por sub-banda. Essa propriedade será explorada nos próximos Capítulos.

## 4.5 Conclusões do Capítulo

Neste Capítulo foi mostrado como se obter uma representação compacta dos sinais de EEG. A representação é obtida ao se projetar os sinais de EEG brutos em uma base de sinais senoidais definida em uma faixa de frequência de interesse. A compactação pode ser aferida, principalmente, através da redução drástica nas dimensões da nova matriz que representa os sinais, por exemplo, tem-se inicialmente, uma matriz

contendo os sinais de EEG brutos  $Z \in \mathbb{R}^{p \times q}$ , enquanto a nova representação possui dimensões reduzidas, isto é,  $\Gamma \in \mathbb{R}^{p \times 2m}$ , onde  $2m \ll q$ .

Outro aspecto a ser ressaltado é que de posse da representação compacta dos sinais de EEG e da base de sinais senoidais definida numa faixa de frequências de interesse, os sinais de EEG filtrados na mesma faixa podem ser aproximados com erro pequeno. Essa propriedade ainda é válida quando uma sub-banda caracterizada por um subconjunto de sinais consecutivos da base de sinais senoidais é utilizada, com os respectivos coeficientes, para a reconstrução do sinal. Nesse caso, o erro de aproximação é pequeno em relação ao sinal de EEG filtrado na mesma faixa de frequências contidas no subconjunto de sinais da base.

Finalmente, destaca-se ainda que a representação compacta dos sinais de EEG mantém a informação espacial, intrínseca da distribuição dos sensores ao longo do couro cabeludo, e também, disponibiliza a informação frequencial dos sinais de forma explícita.

No próximo Capítulo, essas propriedades serão utilizadas para reduzir o número de parâmetros de uma parametrização de uma transformação linear aplicada aos sinais de EEG. Os parâmetros dessa transformação serão expressos linearmente na matriz de covariância dos sinais após a transformação.



# 5 Parametrização da Transformação Linear

## 5.1 Introdução

No Capítulo 4 foi mostrado como se obter uma representação compacta do sinal filtrado de EEG. A representação compacta dos sinais é expressa através de uma matriz de coeficientes que é resultante da projeção do sinal de EEG bruto em uma base de senoides que por sua vez está definida em uma faixa de frequência de interesse.

Durante o pré-processamento dos sinais de EEG, uma abordagem comum dos algoritmos é utilizar transformações lineares que combinam a informação de diferentes sensores do sinal de EEG. Essas transformações, em geral, resultam em novos sinais com propriedades de interesse ressaltadas. Dentre os algoritmos utilizados com essa finalidade, destaca-se o filtro espacial CSP aplicado ao sinal de EEG filtrado, descrito no Capítulo 3.

No domínio do tempo discreto (amostras do sinal), o pré-processamento realizado consiste, em geral, em filtrar os sinais de EEG para limitá-los em faixa de frequências de interesse através de filtros convencionais, como discutido na seção 3.2.

Um aspecto interessante a ser notado é que com a filtragem convencional a dimensão dos sinais resultantes é mantida, enquanto que com a representação compacta dos sinais há uma redução significativa nas dimensões da nova representação sem perda significativa de informação.

Nesse Capítulo, inicialmente é apresentada uma transformação linear definida para o espaço das amostras dos sinais de EEG filtrados. Mostra-se que os parâmetros que definem a transformação podem ser

expressos como elementos de uma matriz (matriz de parâmetros) que parametriza linearmente a matriz de covariância dos sinais transformados.

Mostra-se também que a dimensão da matriz de parâmetros pode ser reduzida ao se utilizar a representação compacta dos sinais em contraste aos sinais filtrados. Assim, a transformação linear pode ser obtida de forma indireta através da matriz de parâmetros como solução de problemas de otimização convexa com restrições por desigualdades lineares matriciais - LMI.

Os problemas de otimização podem ser formulados para realçar propriedades relevantes dos sinais de EEG com o objetivo de facilitar a classificação dos sinais entre as tarefas cognitivas utilizadas por sistemas ICM.

## 5.2 Transformação Linear

Nessa seção, é definida uma transformação linear aplicada no espaço das amostras dos sinais de EEG. Sendo assim, seja  $X \in \mathbb{R}^{p \times q}$  um segmento de sinal de EEG resultante da filtragem por um filtro passa-banda ideal do segmento de sinal de EEG bruto  $Z \in \mathbb{R}^{p \times q}$ , ambos contendo  $p$  sensores e  $q$  amostras por sensor, que foram amostradas a taxa de  $f_s$ , como definido no Capítulo 4. Seja  $G \in \mathbb{R}^{q \times q}$  uma matriz quadrada.

Considere a transformação linear aplicada no espaço das amostras do sinal, definida por:

$$Y = ZG \quad (5.1)$$

onde  $Y \in \mathbb{R}^{p \times q}$  é a matriz resultante da transformação de  $Z$  realizada por  $G$ .

A transformação linear  $G$ , assim como os filtros espaciais, pode ser determinada para ressaltar propriedades dos sinais para fins de classificação. No entanto, diferentemente dos filtros espaciais, a matriz  $G$  opera no espaço das amostras do segmento de sinal  $Z$  (colunas de  $Z$ )

e portanto pode emular uma filtragem temporal. Por exemplo, no caso particular em que a matriz de transformação  $G$  é definida a partir da base de sinais senoidais  $B$ , isto é,  $G = B^T(BB^T)^{-1}B$  como na Eq. (4.20) do Capítulo 4, os sinais em  $Y$ , resultantes da transformação, são projeções do sinal  $Z$  na base de sinais senoidais  $B$  e, portanto, satisfazem a igualdade,  $Y = \hat{X}$ .

Para o caso em que a matriz  $G$  deve ser encontrada, a determinação direta de  $G$  pode demandar um custo computacional elevado, tendo em vista que as dimensões de  $G$  são grandes, em geral igual ao número de amostras  $q$  do sinal  $Z$ .

Uma alternativa à determinação direta da matriz de transformação  $G$  é considerar a matriz de covariância dos sinais transformados e, através dela, determinar a transformação  $G$ . Esta abordagem será discutida na próxima seção.

### 5.3 Matriz de Covariância dos Sinais Transformados

Nessa seção a matriz de covariância dos sinais transformados é definida. O objetivo é parametrizar a matriz de transformação  $G$  e expressar os parâmetros de forma linear na matriz de covariância dos sinais transformados.

Seja a matriz de covariância do segmento de sinal  $Y \in \mathbb{R}^{p \times q}$  definida por:

$$\text{cov}\{Y\} \triangleq \frac{1}{q}YY^T, \quad (5.2)$$

onde, assume-se que  $q \gg p$ .

A fim de expressar a matriz de covariância de  $Y$  como função da transformação  $G$ , substitui-se na Eq. (5.2) a Eq. (5.1) e obtém-se,

$$\text{cov}\{Y\} = \frac{1}{q}ZGG^TZ^T = ZHZ^T \quad (5.3)$$

onde foi utilizado a mudança de variável  $H \triangleq \frac{1}{q}GG^T$ . A matriz  $H \in \mathbb{R}^{q \times q}$  é simétrica e positiva semi-definida por construção, isto é,  $H^T = H \succeq 0$ .

Observa-se que a matriz  $H$  parametriza a transformação  $G$  e pode ser determinada para alterar propriedades da matriz de covariância dos sinais transformados. Por exemplo, para uma matriz  $H$  particular dada, a transformação  $G$ , na Eq. (5.3), pode ser obtida pela expressão:  $G \triangleq q^{\frac{1}{2}} H^{\frac{1}{2}}$ .

A dimensão da matriz  $H$  é a mesma da matriz  $G$ . O número de parâmetros da matriz  $H$  é função da quantidade de amostras consideradas no segmento de sinal  $Z$ , sendo dada pela expressão  $q(q+1)/2$ . Portanto, determinar a matriz  $H$  ainda pode demandar um custo computacional elevado, assim como a determinação da matriz  $G$ .

Observa-se, no entanto, que uma alternativa ao uso do segmento de sinal de EEG bruto,  $Z$ , nas equações Eq. (5.1) e Eq. (5.3), é utilizar o sinal de EEG filtrado por um filtro passa-faixa ideal  $X$ .

A escolha do sinal  $X$ , ao invés do sinal  $Z$ , é normalmente realizada, em sistemas ICM baseados em imagética motora, para selecionar uma faixa de frequência do conteúdo espectral do sinal associado ao fenômeno neurofisiológico de interesse. Além de ter como consequência secundária, a atenuação de componentes de frequência associadas a sinais não desejados, como por exemplo artefatos e ruídos.

Como já discutido no Capítulo 4, o sinal  $X$  possui o mesmo conteúdo espectral do sinal  $Z$  na banda passante do filtro ideal. Destaca-se que o sinal resultante da projeção do sinal  $Z$  na base de sinais senoidais, dado por  $\hat{X} = \Gamma B$ , apresenta o mesmo conteúdo frequencial de  $Z$  e de  $X$ , para as frequências que formam a base de sinais  $B$ .

Assim, uma alternativa ao uso, tanto dos sinais  $Z$ , quanto dos sinais  $X$ , nas equações Eq. (5.1) e Eq. (5.3), é definir a transformação linear  $G$  para o sinal  $\hat{X}$ , resultante da projeção de  $Z$  na base de sinais senoidais,  $B$ . Então, define-se,

$$\hat{Y} \triangleq \hat{X}G \tag{5.4}$$

onde  $\hat{Y}$  são os sinais resultantes da transformação de  $\hat{X}$  realizada pela matriz  $G$ . Análogo as Eq. (5.2) e Eq. (5.3) pode-se definir a matriz de

covariância de  $\hat{Y}$  como segue,

$$\text{cov}\{\hat{Y}\} \triangleq \frac{1}{q} \Gamma B G G^T B^T \Gamma^T = \Gamma H_0 \Gamma^T \quad (5.5)$$

onde utilizou-se a mudança de variável,

$$H_0 \triangleq \frac{1}{q} B G G^T B^T \in \mathbb{R}^{2m \times 2m}, \quad H_0 \succeq 0 \quad (5.6)$$

e  $\hat{X} \triangleq \Gamma B$  é a reconstrução dos sinais a partir da base de sinais  $B$  e da matriz de coeficientes  $\Gamma$ , discutidos na Capítulo 4. A matriz  $H_0$  na Eq. (5.6) é denominada de matriz de parâmetros.

Assume-se que, o erro de aproximação entre os sinais  $X$  e  $\hat{X}$  é pequeno, como pode ser visto na Figura 16. Além disso, os sinais  $X$  e  $\hat{X}$  compartilham o mesmo espectro de amplitude para as frequências que definem a base de sinais senoidais. Assim, as matrizes de covariância,  $\text{cov}\{\hat{X}\}$  e  $\text{cov}\{X\}$  são semelhantes, logo para uma mesma transformação linear  $G$ ,  $\text{cov}\{\hat{Y}\}$  e  $\text{cov}\{Y\}$ , com  $Y = XG$ , são semelhantes.

Observa-se ainda que o número de parâmetros livres a serem determinados na Eq. (5.5) é função da dimensão da matriz  $H_0$  que por sua vez é função do número  $m$  de senoides distintas que compõem a base de sinais considerada, isto é,  $B \in \mathbb{R}^{2m \times q}$ . Em geral, o número de senoides distintas é muito menor do que a quantidade de amostras do segmento do sinal filtrado  $X$  ou da aproximação  $\hat{X}$ , o que implica em  $2m \ll q$ . Portanto, determinar a matriz  $H_0$  demanda um custo computacional menor do que determinar a matriz  $H$ .

Nota-se ainda que a matriz de transformação  $G$  pode ser obtida a partir das matrizes  $H_0$  e  $B$ , como segue,

$$G = B^T (B B^T)^{-1} (q H_0)^{\frac{1}{2}} \quad (5.7)$$

mesmo que isso seja possível, nesse trabalho não utilizou-se com a matriz  $G$ , uma vez que para os interesses de classificação a determinação da matriz  $H_0$  é suficiente.

Observa-se na Eq. (5.7) que a matriz  $H_0$  parametriza a matriz de transformação  $G$ . Além disso, a matriz de parâmetros  $H_0$  é expressa

linearmente na matriz de covariância dos sinais transformados e apresenta como restrição intrínseca ser positiva semidefinida, como pode ser visto na Eq. (5.6). Isso permite o uso de algoritmos numéricos eficientes para a determinação de  $H_0$ , como será descrito na seção 5.5.

A redução do número de parâmetros livres de  $H$  com  $q(q+1)/2$  para  $H_0$  com  $m(2m+1)$  onde  $q \gg m$ , implica em uma demanda computacional menor para determiná-los. Logo, com a mudança de variável definida na Eq. (5.6), a matriz de covariância do sinal transformado é parametrizada pela matriz  $H_0$  ao invés da matriz  $H$ , como na Eq. (5.3).

Outro aspecto interessante, é que na representação compacta de  $\hat{X}$  via matriz de coeficientes  $\Gamma$  torna-se explícito o conteúdo frequencial do sinal  $\hat{X}$  na base de sinais senoidais  $B$ . Nota-se que cada coluna da matriz  $\Gamma \in \mathbb{R}^{p \times 2m}$  está associada a um elemento da base de senoides, como discutido na seção 4.4. Logo, a matriz de parâmetros  $H_0$ , na Eq. (5.5), pondera a colunas da matriz  $\Gamma$  e conseqüentemente, a informação em frequência dos sinais  $\hat{X}$  para obter a matriz de covariância do sinal transformado.

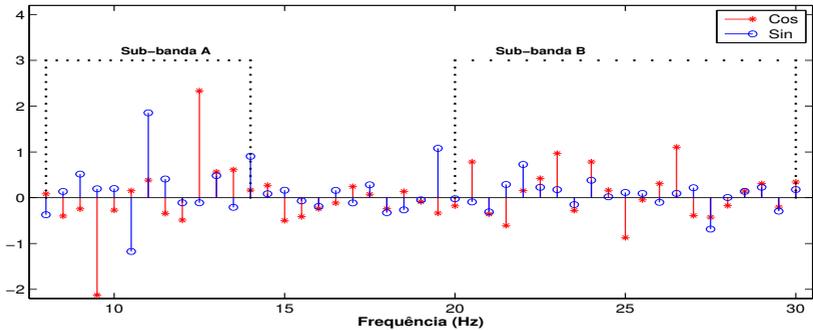
Nesse sentido, a matriz  $H_0$  pode ser escolhida para selecionar uma sub-banda de frequência do sinal  $\hat{X}$  com largura de banda  $\omega_b$ . Por exemplo, seja  $h_{0i,j}$  o elemento na  $i$ -ésima linha e na  $j$ -ésima coluna da matriz  $H_0$ , para  $i, j = 1, 2, \dots, 2m$ . Para selecionar uma sub-banda de frequência basta fazer os elementos da diagonal de  $H_0$  correspondentes aos coeficientes da sub-banda desejada serem iguais a 1, enquanto que os demais elementos, sejam nulos, isto é,

$$h_{0i,j} = \begin{cases} 1, & \forall i = j \text{ e } i \in \omega_b \\ 0, & \forall i \neq j \text{ e } i \ni \omega_b \end{cases} \quad (5.8)$$

onde  $\ni$  indica a não pertinência do elemento ao conjunto e conseqüentemente, apenas os elementos correspondentes a largura de banda da sub-banda de interesse são diferentes de zero. Ao aplicar a matriz  $H_0$  definida em Eq. (5.8) na Eq. (5.5), apenas os coeficientes que correspondem aos elementos de  $H_0$  diferentes de zero são considerados no cálculo da matriz de covariância de  $\hat{X}$ . Logo, a matriz de covariância

resultante corresponde a sub-banda de frequência associada aos coeficientes selecionados por  $\omega_b$ , como pode ser visto na Figura 18.

Figura 18: Exemplo de duas Sub-bandas de frequência: Sub-banda A e Sub-banda B, ambas definidas a partir dos coeficientes de um sensor resultantes da projeção dos sinais  $Z$  na base  $B$ . Tem-se os coeficientes associados aos sinais senoidais (círculo na cor azul) e aos sinais cosenoidais (asterisco na cor vermelha) da base de sinais. Para a sub-banda A tem-se  $\omega_b = 13$ , enquanto que para a sub-banda B tem-se  $\omega_b = 21$ . A escolha das sub-bandas A ou B, pôde ser realizada definindo-se a os elementos da matriz  $H_0$  na Eq. (5.8). A energia do sinal na sub-banda A corresponde a 59,51%, enquanto a energia na sub-banda B corresponde a 29,90% da energia total na faixa entre 8 a 30  $Hz$ .



Na próxima seção, será ilustrado uma forma simples de se obter a matriz de parâmetros  $H_0$  a partir das matrizes de coeficientes  $\Gamma_i$ . A estratégia adotada é análoga ao algoritmo CSP descrito na seção 3.3. A classificação dos sinais é realizada com base no algoritmo utilizando as ferramentas de Riemann, descritas na seção 3.5.2

## 5.4 Classificação dos sinais de EEG via matriz de covariância

Na seção anterior, mostrou-se que ao aplicar uma transformação linear nos sinais de EEG filtrados pode-se expressar os parâmetros da transformação linearmente na matriz de covariância dos sinais

transformados. Esses parâmetros podem ser interpretados como graus de liberdades adicionais que podem ser ajustados para, por exemplo, facilitar a classificação dos sinais de EEG entre categorias associadas a comandos a serem gerados.

Em particular, esses graus de liberdade são interessantes quando se utilizam as ferramentas de Riemann para classificar os sinais, discutidos na seção 3.5.2. Nessa abordagem, faz-se uso das matrizes de covariância dos sinais como características. Assim a classificação dos segmentos de sinais de EEG filtrados é realizada com base na distância de Riemann entre as matrizes de covariância dos sinais de EEG e a matriz de covariância média que caracteriza a classe considerada, como descrito na seção 3.5.2.

Uma propriedade da distância de Riemann entre matrizes é a sua invariância às transformações de similaridade e de congruência, por exemplo,  $\delta_R(\Phi, \Omega) = \delta_R(W\Phi W^T, W\Omega W^T)$ , segundo Congedo, Barachant e Andreev (2013). Essa propriedade torna a abordagem via Riemann sem parâmetros livres para serem ajustados, uma vez que os sinais de EEG estão filtrados em uma faixa de frequência de interesse e nenhuma transformação espacial pode ser determinada para auxiliar na classificação dos sinais.

Nesse sentido, o uso da matriz de covariância definida na Eq. (5.5) introduz graus de liberdade que podem ser ajustados para melhorar o desempenho de classificação utilizando distância de Riemann.

Nesta seção é apresentado um exemplo de como se determinar a matriz de parâmetros  $H_0$  a partir da matriz de coeficientes dos sinais. O problema formulado nesta seção, é análogo ao algoritmo CSP, apresentado na seção 3.3. Para classificar os sinais são utilizadas as ferramentas de Riemann, apresentadas na seção 3.5.2 aplicadas às matrizes de covariância dos sinais transformados.

Considere o problema de classificação dos sinais de EEG em duas classes,  $\mathcal{A}$  e  $\mathcal{B}$ . Cada uma das classes contendo  $n_a$  e  $n_b$  elementos,

respectivamente. Os elementos são definidos como,

$$\tilde{S}_i = (2m) \text{cov}\{\Gamma_i^T\} = \Gamma_i^T \Gamma_i \quad (5.9)$$

para todo  $i = 1, 2, \dots, n_a + n_b$  e  $\Gamma_i \in \mathbb{R}^{p \times 2m}$ . Observe que, por conveniência, estamos trabalhando com  $\text{cov}\{\Gamma_i^T\}$  ao invés de  $\text{cov}\{\Gamma_i\}$  que seria o normal. Esse artifício, também utilizado em (AGHAEI; MAHANTA; PLATANIOTIS, 2016) e simplifica consideravelmente a determinação da matriz de parâmetros  $H_0$  como veremos a seguir. Mais detalhes sobre esse artifício podem ser encontrados em (AGHAEI; MAHANTA; PLATANIOTIS, 2016).

Seja  $\tilde{S}_a$  e  $\tilde{S}_b$  a média entre as matrizes na Eq. (5.9) para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente, definidas como segue,

$$\tilde{S}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} \tilde{S}_i, \quad \forall \tilde{S}_i \in \mathcal{A} \quad (5.10)$$

e

$$\tilde{S}_b = \frac{1}{n_b} \sum_{i=1}^{n_b} \tilde{S}_i, \quad \forall \tilde{S}_i \in \mathcal{B} \quad (5.11)$$

A decomposição espectral da soma entre as matrizes  $\tilde{S}_a$  e  $\tilde{S}_b$  é

$$\hat{S} = \tilde{S}_a + \tilde{S}_b = \tilde{U} \tilde{\Lambda} \tilde{U}^T \quad (5.12)$$

onde  $\tilde{U}$  e  $\tilde{\Lambda}$  são as matrizes de autovetores e autovalores, respectivamente, de  $\hat{S}$ , com  $\tilde{U}^T \tilde{U} = I$ . Suponha  $\tilde{\Lambda}$  inversível e defina a matriz  $\tilde{Q} \triangleq \tilde{\Lambda}^{-\frac{1}{2}} \tilde{U}^T$  e

$$\tilde{Q} \tilde{S}_a \tilde{Q}^T = \tilde{V} \tilde{\Lambda}_a \tilde{V}^T \quad (5.13)$$

expressa a decomposição espectral de  $\tilde{Q} \tilde{S}_a \tilde{Q}^T$ , onde  $\tilde{V}$  e  $\tilde{\Lambda}_a$  são as matrizes de autovetores e autovalores de  $\tilde{Q} \tilde{S}_a \tilde{Q}^T$ , com  $\tilde{V}^T \tilde{V} = I$ . Caso  $\tilde{\Lambda}$  não seja inversível um método de regularização deve ser utilizado. Neste trabalho usaremos o método proposto em (BLANKERTZ et al., 2010).

Análogo ao discutido na seção 3.3, define-se a transformação,

$$\tilde{H} \triangleq \tilde{V}^T \tilde{Q} \quad (5.14)$$

e define-se  $\tilde{\Gamma} \triangleq \Gamma \tilde{H}^T$  e tem-se que  $\tilde{\Gamma}_i \tilde{\Gamma}_i^T = \Gamma_i \tilde{H}^T \tilde{H} \Gamma_i^T$ .

Logo, por analogia segue que

$$\text{cor}\{\hat{Y}_i\} = \Gamma_i H_0 \Gamma_i^T = \Gamma_i \tilde{H}^T \tilde{H} \Gamma_i^T \quad (5.15)$$

então define-se,

$$H_0 = \tilde{H}^T \tilde{H} \quad (5.16)$$

Observe que quando  $\tilde{\Lambda}$  é inversível, a regularização não se faz necessária e tem-se  $H_0 = \hat{S}^{-1}$ . Uma interpretação possível para esse resultado é que ele define uma transformação inspirada no método CSP aplicado no espaço dos sinais auxiliares  $\Gamma_i^T$ .

Para a classificação dos sinais de EEG, utilizam-se as ferramentas da geometria de Riemann, descritas na seção 3.5.2. Assim calcula-se a média geométrica entre as matrizes definidas em Eq. (5.15) para ambas as classes  $\mathcal{A}$  e  $\mathcal{B}$  a partir das equações Eq. (3.33) e Eq. (3.34), respectivamente.

Em seguida, utiliza-se a decisão apresentada em Eq. (3.35). Os resultados de acurácia de classificação estão mostrados na Tabela 2, tanto para o método com a transformação definida em Eq. (5.16), quanto para o caso onde a matriz  $H_0$  é a matriz identidade, que corresponde ao caso sem transformação.

Como o uso da distância de Riemann requer a matriz em Eq. (5.16) positiva definida, uma regularização do problema foi aplicada para cada indivíduo de acordo com Blankertz et al. (2010). Finalmente vale a pena salientar que Eq. (5.16) foi obtida através de argumentos baseados em transformações espaciais inspirados no CSP. Nenhum critério de otimalidade foi considerado para esse fim.

Os resultados apresentados na Tabela 2 corroboram o argumento de que os graus de liberdade introduzidos a partir da transformação linear  $G$  e obtidos por meio da matriz de parâmetros  $H_0$  podem ser utilizados para melhorar a taxa de classificação dos sinais de EEG aplicados a sistemas ICM. Especialmente, para métodos que utilizam matrizes de covariância/covariância como descritores dos sinais, como é o caso de classificação via distância de Riemann. Destaca-se ainda

Tabela 2: Resultados de classificação dos sinais de EEG em termos de acurácia utilizando as ferramentas de Riemann.  $\lambda$  é o parâmetro de regularização utilizado para cada sujeito. São mostrados os resultados com a transformação definida em Eq. (5.16) e sem a transformação,  $H_0 = I$ .

Sujeitos	1	2	3	4	5	6	7	8	9	Média
$\lambda$	0.8	0.0	0.6	0.9	0.3	0.25	0.25	0.1	0.9	-
com transformação	91,55	53,81	91,55	72,22	62,34	71,11	84,72	96,53	91,65	79,49
sem transformação	90,15	53,81	90,85	72,92	56,81	64,09	74,31	95,83	92,34	76,79

que, no caso da abordagem via Riemann, uma vez definida a faixa de frequências de interesse, os sinais são filtrados e as matrizes de covariância são obtidas, assim não há graus de liberdade para ajuste do método.

Assim, o uso das matrizes de covariância dos sinais transformados (definida na Eq. (5.5)) é uma forma alternativa à matrizes de covariância originais e adicionam novos graus de liberdade, que por simplicidade podem ser determinados como no método descrito nesta seção, alcançando resultados de classificação superiores quando comparados a sem a transformação.

Como mencionado anteriormente, a matriz  $H_0$  em Eq. (5.5) parametriza a transformação linear  $G$  e é expressa linearmente na matriz de covariância dos sinais transformados. Essa característica, nos permite utilizar ferramentas numéricas robustas para a determinação da matriz  $H_0$  através da formulação de problemas de otimização convexa. Esse é o tema da próxima seção.

## 5.5 Desigualdades Matriciais Lineares

Desigualdades matriciais lineares (LMI, *Linear Matrix Inequalities*) são ferramentas matemáticas que têm sido aplicadas em uma variedade de problemas, principalmente na área de Teoria de Controle e Sistemas (BALAKRISHNAN; VANDENBERGHE, 1998). Em geral,

esses problemas podem ser formulados como problemas de otimização convexa envolvendo restrições em termos de LMIs (BOYD et al., 1994).

Um aspecto interessante de formulações LMIs é que mesmo que o problema não apresente solução analítica (ou pelo menos, não se conheça até o momento), formulá-lo em termos de LMIs é garantia de solução em tempo polinomial com algoritmos de ponto interior (BOYD et al., 1994).

Há alguns pacotes computacionais, comerciais e públicos, disponíveis para a solução de problemas LMIs, que envolvem desde o desenvolvimento de interfaces amigáveis como o *Yalmip* (LOFBERG, 2004), até a disponibilização de resolvedores como o *SeDuMi* (STURM, 1998) e o *SDPT3* (TOH et al., 1998) para o ambiente *MATLAB* (MATLAB, 2010).

Nesta seção é realizada uma breve apresentação dos conceitos básicos sobre problemas de otimização envolvendo LMIs.

### 5.5.1 Problema de Otimização Convexa

Um problema de otimização convexa pode ser escrito na forma:

$$\min_x f_0(x), \text{ sujeito à: } f_i(x) \leq b_i, i = 1 \cdots, n_f. \quad (5.17)$$

onde as funções  $f_0(x), \cdots, f_{n_f}(x) : \mathbb{R}^d \mapsto \mathbb{R}$  são convexas, isto é, satisfazem a condição.

$$f_i(\xi x_1 + (1 - \xi)x_2) \leq \xi f_i(x_1) + (1 - \xi)f_i(x_2) \quad (5.18)$$

para todo  $x_1, x_2 \in \mathbb{R}^d$  e para todo  $0 \leq \xi \leq 1$ . Os problemas de otimização convexa possuem a propriedade de que um mínimo local corresponde ao mínimo global.

Os problemas de interesse nesse trabalho são formulados como problemas de otimização semidefinida (SDP): uma classe de problemas de otimização convexa na qual a função objetivo é linear no conjunto de variáveis de decisão (por exemplo,  $h = [h_1, \cdots, h_d]^T$ ) e as restrições

estão na forma de desigualdades matriciais lineares (LMIs),

$$\min_h c^T h, \text{ sujeito a: } F(h) \succeq 0 \quad (5.19)$$

onde a desigualdade matricial linear (LMI) é definida como,

$$F(h) \triangleq F_0 + \sum_{i=1}^d h_i F_i \succeq 0 \quad (5.20)$$

Os dados do problema são o vetor  $c \in \mathbb{R}^d$  e  $d + 1$  matrizes simétricas fixas  $F_0, \dots, F_d \in \mathbb{R}^{p \times p}$ , sendo  $h$  o vetor de variáveis de decisão.

A desigualdade  $F(h) \succeq 0$  significa que a matriz  $F(h)$  é positiva semidefinida, isto é,  $z^T F(h) z \geq 0$  para todo  $z \in \mathbb{R}^p$ . Note ainda que a desigualdade na Eq. (5.20) pode ser também estrita, isto é,  $\succ 0$ , ou ainda  $\preceq 0$  e  $\prec 0$ . Existem hoje pacotes computacionais eficientes para resolver problemas de otimização envolvendo LMI como restrições (LOFBERG, 2004; STURM, 1998; TOH et al., 1998).

Um conjunto  $\mathcal{C}$  é convexo se a linha entre quaisquer dois pontos do conjunto está completamente contida no conjunto:  $h_1, h_2 \in \mathcal{C} \rightarrow \xi h_1 + (1 - \xi) h_2 \in \mathcal{C}$  para todo  $0 \leq \xi \leq 1$ . Uma propriedade interessante das LMIs é que elas definem uma restrição convexa, ou seja, o conjunto de soluções  $\{h \in \mathbb{R}^d \mid F(h) \succeq 0\}$  que satisfaz as restrições é convexo.

O problema de otimização expresso na Eq. (5.19) engloba também uma classe de problemas mais simples, como por exemplo, determinar uma solução  $h$  que satisfaça a desigualdade  $F(h) \succeq 0$ . Esses problemas são denominados de problema de factibilidade.

### 5.5.2 Complemento Schur

Em geral, os problemas não aparecem diretamente na forma de desigualdades matriciais lineares. Um artifício utilizado com frequência para converter uma desigualdade (convexa) não-linear em uma LMI é o complemento Schur e será apresentado na seqüência.

Considere a matriz simétrica  $M$  particionada em quatro blocos

matriciais,

$$M = \begin{bmatrix} M_1 & M_2 \\ M_2^T & M_3 \end{bmatrix} \quad (5.21)$$

O complemento Schur de  $M_1$  e  $M_3$  em  $M$  são definidos, respectivamente, como segue,

$$(M/M_1) \triangleq M_3 - M_2^T M_1^{-1} M_2 \quad (5.22)$$

e

$$(M/M_3) \triangleq M_1 - M_2 M_3^{-1} M_2^T \quad (5.23)$$

Algumas propriedades que relacionam as matrizes  $(M/M_1)$  e  $(M/M_3)$  com a matriz  $M$ , são:

1. Se  $M_1 \succ 0$ , então  $M = M^T \succeq 0$  se e somente se  $(M/M_1) \succeq 0$ ;
2. Se  $M_3 \succ 0$ , então  $M = M^T \succeq 0$  se e somente se  $(M/M_3) \succeq 0$ ;

Para ilustrar o uso do complemento Schur em problemas de otimização, considere o exemplo de converter restrições com desigualdade matricial não-linear em restrições LMI. As desigualdades,

$$\text{tr}\{M_2^T(h)M_3(h)^{-1}M_2(h)\} < 1 \text{ e } M_3(h) \succ 0, \quad (5.24)$$

podem ser, de forma equivalente, expressas em termos das seguintes restrições lineares,

$$\text{tr}\{\tilde{M}\} < 1 ; \begin{bmatrix} \tilde{M} & M_2^T(h) \\ M_2(h) & M_3(h) \end{bmatrix} \succeq 0 \quad (5.25)$$

onde a matriz  $\tilde{M}$  é uma variável matricial auxiliar.

## 5.6 Determinação da Matriz de Parâmetros via Problema de Otimização

Note que a matriz de covariância dos sinais transformados na Eq. (5.5) é linear em  $H_0$ . Assim o problema de se determinar a matriz  $H_0$  pode ser formulado como um problema de otimização. Como

destacado anteriormente, a matriz  $H_0$  deve ser, no mínimo, positiva semi-definida a fim de garantir que a matriz de covariância do sinal após a transformação seja também, positiva semi-definida e portanto, uma matriz de covariância. A restrição,  $H_0 \succeq 0$ , é uma LMI pois pode ser representada na forma 5.20.

Genericamente, determinar a matriz  $H_0$  via solução de um problema de otimização pode ser formulado como segue:

$$\min_{H_0} J(H_0), \text{ sujeito à: } H_0 \in \mathcal{M} \quad (5.26)$$

onde a função  $J(H_0) : \mathbb{R}^{2m \times 2m} \mapsto \mathbb{R}$  é uma função custo linear em  $H_0$  que representa os objetivos do problema de otimização e  $\mathcal{M}$  representa o conjunto de restrições que se deseja satisfazer para o problema.

Versões particulares da função custo e do conjunto de restrições vão depender do problema que está se tentando resolver. Destaca-se que o problema dificilmente aparece na forma de Eq. (5.26) como uma função custo linear e restrições LMI. Contudo, há várias técnicas auxiliares que podem ser utilizadas para re-escrever o problema na forma desejada, sempre que isso for possível.

## 5.7 Conclusões do Capítulo

Neste Capítulo, mostrou-se como parametrizar uma transformação linear aplicada aos sinais de EEG. Os parâmetros (elementos de uma matriz simétrica e positiva semidefinida) que definem a transformação foram expressos linearmente na matriz de covariância dos sinais após a transformação.

Com o auxílio da representação compacta dos sinais de EEG, os parâmetros podem ser determinados como solução de um problema de otimização convexa com restrições LMI, no qual dispõe-se de pacotes computacionais eficientes para a sua solução.

Um exemplo de classificação dos sinais é mostrado. Nele as matrizes de covariância dos sinais são utilizadas como descritores dos

sinais para fins de classificação via ferramentas de Riemann. A matriz de parâmetros é determinada por um procedimento análogo ao algoritmo CSP. Com os resultados mostra-se que o ajuste dos parâmetros pode melhorar o desempenho de classificação dos sinais.

Nos Capítulos 6 e 7 são discutidas escolhas de funções objetivo e conjuntos de restrições que tornam o problema de otimização formulado nesse Capítulo particular. As abordagens via Riemann e CSP-LDA, respectivamente, são utilizadas como ponto de partida para a determinação das funções objetivo e do conjunto de restrições.

# 6 Classificação dos Sinais de EEG via Ferramentas de Riemann com Otimização

## 6.1 Introdução

No Capítulo 5 foi descrito como parametrizar uma transformação linear aplicada aos sinais de EEG. Os parâmetros que definem a transformação são estruturados como elementos de uma matriz de parâmetros, que é simétrica e positiva semi-definida.

Um aspecto interessante é que a matriz de parâmetros pode ser expressa linearmente na matriz de covariância dos sinais após a transformação linear, e portanto, determinar a matriz de parâmetros pode ser realizado como solução de um problema de otimização convexa.

Nesse Capítulo, os sinais de EEG são descritos através da matriz de covariância após a transformação linear apresentada no Capítulo anterior. Assim, um problema particular de otimização é formulado para se determinar a matriz de parâmetros. O objetivo é determinar a matriz de parâmetros que reduza o erro de classificação dos sinais de EEG.

Com esse fim, a função objetivo e o conjunto de restrições LMIs do problema de otimização são definidos com base em ideias simples de classificação, como a redução de dispersão dos elementos em torno do valor médio para cada uma das classes. Além disso, afastar, ao máximo possível, os valores médios de classes diferentes.

As métricas de Riemann, discutidas na seção 3.5.2, são utilizadas para a classificação dos sinais de EEG. Com elas, as matrizes de covariância dos sinais de EEG transformados são utilizadas como

características dos sinais para a classificação. A média geométrica entre as matrizes de covariância dos sinais é utilizada para caracterizar classes diferentes. A distância de Riemann entre uma dada matriz de covariância e a média geométrica é utilizada para assinalar a classe dos sinais.

Em geral, quando as ferramentas de Riemann são utilizadas para fins de classificação de sinais de EEG em sistemas ICM, o único pré-processamento realizado é a filtragem espectral dos sinais para selecionar a faixa de frequências de interesse. Assim, não há parâmetros disponíveis para ajuste do método.

Por outro lado, com a matriz de parâmetros linearmente expressa na matriz de covariância dos sinais transformados, introduz-se parâmetros que podem ser determinados para ressaltar propriedades dos sinais de EEG para reduzir o erro de classificação.

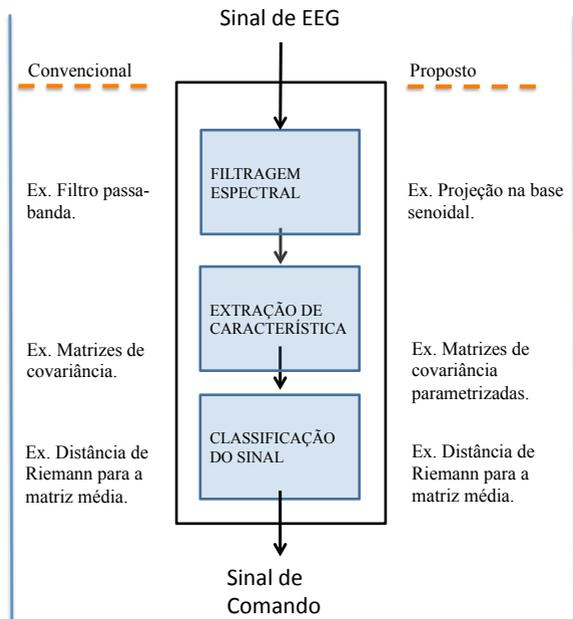
Por exemplo, pode-se determinar a matriz de parâmetros com o objetivo de reduzir a dispersão entre as matrizes de covariância dos sinais transformados em torno da média entre as matrizes de covariância para cada uma das classes de interesse.

Os resultados apresentados ao final desse Capítulo são obtidos seguindo a cadeia de processamento dos sinais esboçada na Figura 19. Como pode ser visto na Figura 19, as operações de filtragem espectral, extração de características e classificação dos sinais são realizadas de duas formas, uma identificada pela *Convencional* e a outra pela coluna *Proposto*. Na primeira, cada segmento de sinais de EEG é filtrado utilizando-se um filtro passa-faixa. As matrizes de covariância dos segmentos de sinal são utilizadas como características e a distância de Riemann é aplicada para classificar os sinais.

Na segunda coluna, ainda na Figura 19, a projeção dos sinais numa base de sinais senoidal é utilizada para substituir a filtragem pelo passa-faixa. Os coeficientes resultantes da projeção dos sinais na base senoidal são utilizados como uma representação compacta dos segmentos de sinais de EEG e, as matrizes de covariância são, então calculadas

como função dos coeficientes e parametrizadas pela matriz  $H_0$ . A distância de Riemann é utilizada para classificar os sinais.

Figura 19: Cadeia de processamento dos sinais de EEG utilizando as ferramentas de Riemann. Nas colunas na lateral do fluxo principal, tem-se uma breve descrição das diferenças entre a abordagem convencional e a abordagem proposta nessa tese.



## 6.2 Análise preliminar

De uma maneira intuitiva, é natural pensar que em um problema de classificação com parâmetros livres, estes devem ser ajustados para que os elementos que pertencem à mesma classe se concentrem em torno do valor médio da classe, caracterizando uma baixa dispersão em torno do médio. Além disso, os elementos que representam o valor médio de classes diferentes devem estar afastados, segundo alguma métrica de interesse.

Nesse sentido, a matriz de parâmetros  $H_0$ , que parametriza a transformação linear aplicada aos sinais e é expressa linearmente na matriz de covariância dos sinais transformados, pode ser determinada com o objetivo de reduzir a dispersão entre as matrizes de covariância de uma mesma classe e a matriz média da classe.

Na próxima seção, será discutido como reduzir o número de condição de uma matriz através da solução de um problema de otimização quasi-convexo. A minimização do número de condição será utilizado neste Capítulo como estratégia para aproximar matrizes e uma alternativa a distância de Riemann.

### 6.2.1 Número de condição de matriz simétrica

Seja  $A = A^T \in \mathbb{R}^{l \times l}$  uma matriz simétrica e positiva definida e  $\lambda_{max}\{A\}$  e  $\lambda_{min}\{A\}$  denotam os autovalores máximo e mínimo da matriz  $A$ , respectivamente. O número de condição de uma matriz  $A$  é definido como a relação entre o autovalor máximo e o autovalor mínimo de  $A$ , isto é, (HIGHAM, 1995),

$$\kappa\{A\} \triangleq \frac{\lambda_{max}\{A\}}{\lambda_{min}\{A\}} \tag{6.1}$$

Observe que o número de condição para uma dada matriz  $A = A^T \succeq 0$  será sempre maior ou igual a um, ou seja,  $\kappa\{A\} \geq 1$ , onde a igualdade ocorre somente para  $A = I_l$  (matriz identidade com dimensão apropriada).

Sejam  $\Sigma$  e  $\Phi$  duas matrizes simétricas e positivas definidas com dimensões  $l \times l$ . Agora, considere uma matriz  $A$ , tal que  $A = \Sigma^{-\frac{1}{2}} \Phi \Sigma^{-\frac{1}{2}}$ . Nesse caso, observe que, se  $\kappa\{A\} \approx 1$  implica que as matrizes  $\Phi$  e  $\Sigma$  satisfazem a relação de  $\Phi \approx \Sigma$ .

Seja a matriz  $P_\theta = P(\theta)$  uma matriz linear na variável  $\theta$ , tal que,  $P(\theta) \triangleq \sum_{i=1}^{n_\theta} \theta_i P_i$  para  $P_i = P_i^T$  dados e  $\theta = [\theta_1 \cdots \theta_{n_\theta}]^T$  ser um vetor de parâmetros a serem determinados.

Defina, agora, a matriz  $A_\theta \triangleq \Sigma_\theta^{-\frac{1}{2}} \Phi_\theta \Sigma_\theta^{-\frac{1}{2}}$ , onde  $\Phi_\theta$  e  $\Sigma_\theta$  são matrizes positivas semidefinidas e lineares na variável  $\theta$ . Suponha que se

deseja determinar a variável  $\theta$  tal que o número de condição da matriz  $A_\theta$  seja tão pequeno quanto possível, isto é, idealmente,  $\kappa\{A_\theta\} \approx 1$  e, como conseqüência, as matrizes  $\Phi_\theta \approx \Sigma_\theta$ .

O problema de otimização para se determinar o vetor  $\theta$  pode ser formulado como segue,

$$\min_{\theta} \quad \kappa\{A_\theta\} \quad (6.2)$$

$$\text{Sujeito à: } \Sigma_\theta \succ 0, \Phi_\theta \succ 0, A_\theta = \Sigma_\theta^{-\frac{1}{2}} \Phi_\theta \Sigma_\theta^{-\frac{1}{2}}$$

A função  $\kappa\{\cdot\}$  é uma função quasi-convexa, portanto uma solução aproximada do problema Eq. (6.2) pode ser obtida através da solução de uma seqüência de problemas de factibilidade convexo (LU; PONG, 2011).

Considere as LMIs  $\alpha I \succeq A_\theta$  e  $\alpha I \succeq A_\theta^{-1}$ , onde  $A_\theta^{-1} \triangleq \Sigma_\theta^{\frac{1}{2}} \Phi_\theta^{-1} \Sigma_\theta^{\frac{1}{2}}$  e  $\alpha \in \mathbb{R}_+^*$ . O escalar  $\alpha$  é um limitante superior para os autovalores das matrizes  $A_\theta$  e  $A_\theta^{-1}$ , respectivamente. Isto implica em  $\alpha \geq \lambda_{max}\{A_\theta\}$  e  $\alpha \geq \lambda_{max}\{A_\theta^{-1}\}$ . Note ainda que a seguinte relação entre os autovalores das matrizes  $A_\theta$  e  $A_\theta^{-1}$  é válida:  $\lambda_{min}\{A_\theta\} = \lambda_{max}\{A_\theta^{-1}\}$  e vice-versa.

Assim, o número de condição da matriz  $A_\theta$  pode ser expresso como função dos autovalores máximos de  $A_\theta$  e de  $A_\theta^{-1}$ , ou seja,

$$\kappa\{A_\theta\} = \frac{\lambda_{max}\{A_\theta\}}{\lambda_{max}\{A_\theta^{-1}\}} \quad (6.3)$$

Agora, analisemos as seguintes desigualdades matriciais lineares,

$$\alpha I \succeq \Sigma_\theta^{-\frac{1}{2}} \Phi_\theta \Sigma_\theta^{-\frac{1}{2}} \Leftrightarrow \alpha \Sigma_\theta - \Phi_\theta \succeq 0 \quad (6.4)$$

e

$$\alpha I \succeq \Sigma_\theta^{\frac{1}{2}} \Phi_\theta^{-1} \Sigma_\theta^{\frac{1}{2}} \Leftrightarrow \begin{bmatrix} \alpha \Sigma_\theta & \Sigma_\theta \\ \Sigma_\theta & \Phi_\theta \end{bmatrix} \succeq 0 \quad (6.5)$$

onde  $\alpha \in \mathbb{R}_+^*$  e  $I$  é a matriz identidade de dimensão apropriada.

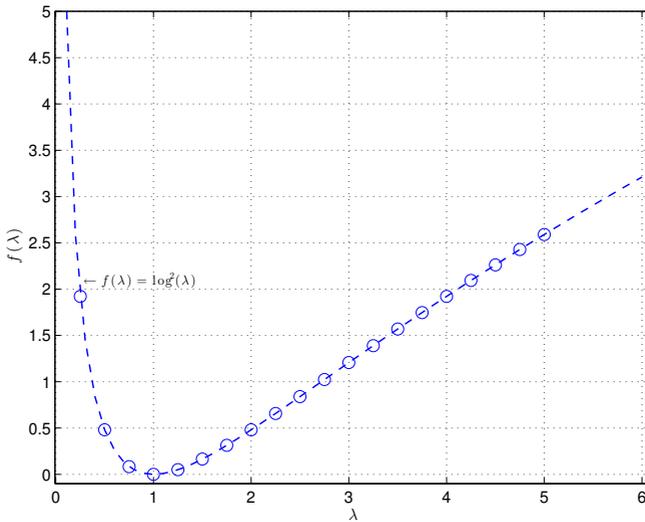
Note que a equivalência descrita em Eq. (6.5) foi obtida aplicando-se o artifício do complemento Schur descrito na seção 5.5.

O problema de otimização expresso na Eq. (6.2) pode ser re-escrito como segue,

$$\begin{aligned} \min_{\theta, \alpha} \quad & \alpha \\ \text{Sujeito à:} \quad & \text{Eq. (6.4), Eq. (6.5)} \end{aligned} \tag{6.6}$$

Observe que o menor valor teórico possível para  $\alpha$  no problema Eq. (6.6) com as restrições LMIs Eq. (6.4) e Eq. (6.5), é a unidade ( $\alpha \geq 1$ ), onde a igualdade, nesse caso, implica em  $\Phi_\theta = \Sigma_\theta$ .

Figura 20: Esboço da função logarítmica ao quadrado utilizada para o cálculo da distância de Riemann. No eixo das abcissas estão os autovalores ( $\lambda$ ). No eixo das ordenadas estão os valores da função ( $f(\lambda) = \log^2(\lambda)$ ).



Note que o problema de otimização Eq. (6.6) é quasi-convexo nas variáveis  $\theta$  e  $\alpha$ . No entanto, para um valor fixo de  $\alpha$  o problema de otimização torna-se convexo na variável  $\theta$  e pode ser resolvido de forma eficiente com ferramentas de otimização convexa. Ao se fixar a variável  $\alpha$ , o problema expresso na Eq. (6.6) torna-se um problema de factibilidade, onde para um dado  $\alpha$  fixo, deseja-se determinar o vetor  $\theta$  que satisfaz, simultaneamente, as restrições expressas em Eq. (6.4) e Eq. (6.5).

Nesse contexto, suponha que  $\alpha^*$  represente o valor ótimo de  $\alpha$  a ser encontrado ao se resolver o problema Eq. (6.6). Seja  $\alpha_*$  um valor de  $\alpha$  em que o problema Eq. (6.6) apresenta solução. Logo, conclui-se que o valor ótimo  $\alpha^*$  deve ser menor que  $\alpha_*$ , ou seja,  $\alpha^* \leq \alpha_*$ .

Uma estratégia simples para resolver o problema de factibilidade é adotar algoritmos iterativos de tal forma a ir reduzindo o valor de  $\alpha$  até que as restrições Eq. (6.4) e Eq. (6.5) não apresentem solução.

## 6.2.2 Conexão entre número de condição e Distância de Riemann

A distância de Riemann entre duas matrizes simétricas e positivas foi definida na seção 3.5.2. Ela expressa o comprimento da curva geodésica entre as duas matrizes e é calculada como uma função não-linear dos autovalores generalizados das duas matrizes, expresso em Eq. (3.28) ou Eq. (3.29).

Na equação Eq. (3.29) observa-se a não linearidade através da função logarítmica ao quadrado, ou seja,  $f(\lambda) = \log^2(\lambda)$ . Na Figura 20 tem-se um esboço dessa função não-linear.

Note que, por causa do mapeamento logarítmico presente no cálculo da distância de Riemann entre matrizes, Eq. (3.29), os autovalores próximos à unidade ( $\lambda \approx 1$ ) têm peso pequeno no somatório, veja Figura 20. Como consequência, os autovalores máximo e mínimo terão maior peso, no sentido de aumentar a distância entre as matrizes. Assim, reduzir o número de condição da matriz através problema de otimização expresso em Eq. (6.6) pode ajudar a reduzir a distância de Riemann entre as matrizes consideradas.

A minimização do número de condição, expresso no problema de otimização Eq. (6.6), atua no autovalor máximo e no autovalor mínimo de  $A_\theta$ . Observe que uma maneira de reduzir o número de condição é aumentar o autovalor mínimo e diminuir o autovalor máximo.

Considere, novamente, a matriz  $A_\theta \triangleq \Sigma_\theta^{-\frac{1}{2}} \Phi_\theta \Sigma_\theta^{-\frac{1}{2}}$ . A distância de Riemann entre a matriz  $A_\theta$  e uma matriz identidade com a mesma

dimensão de  $A_\theta$ , isto é,  $\delta_R(I, A_\theta) = \delta_R(I, \Sigma_\theta^{-\frac{1}{2}} \Phi_\theta \Sigma_\theta^{-\frac{1}{2}})$ , expressa uma medida, no sentido de Riemann, de o quanto a matriz  $A_\theta$  está distante (geodesicamente) da matriz identidade e simultaneamente, o quanto as matrizes  $\Phi_\theta$  e  $\Sigma_\theta$  estão distantes entre si, isto é,  $\delta_R(\Sigma_\theta, \Phi_\theta) = \delta_R(I, A_\theta)$ .

Com base nessas ideias, as matrizes de covariância dos sinais de EEG transformados são utilizadas para representar os sinais de EEG e as ferramentas de Riemann são utilizadas para a classificação dos sinais.

Os parâmetros da transformação serão ajustados para reduzir o número de condição entre cada matriz de covariância e a matriz de covariância média no espaço transformado. Com isso, espera-se reduzir a distância de Riemann entre as matrizes de covariância dos sinais e a matriz de covariância média da classe. Como formular esse problema será o assunto da próxima seção.

### 6.3 Formulação do problema de otimização para EEG

O problema descrito nesta seção consiste em aplicar as ideias de redução do número de condição de uma matriz linearmente dependente de uma variável. As matrizes em questão aqui, são as matrizes de covariância dos sinais de EEG transformados. Os parâmetros a serem ajustados consistem na parametrização da transformação e que estão linearmente expressos na matriz de covariância.

A ideia é ajustar os parâmetros de tal forma que o número de condição entre cada matriz de covariância dos sinais de EEG transformados e a matriz de covariância média entre essas matrizes seja reduzido, ou no mínimo, mantido dentro de uma faixa de valores pré-definidos.

Uma vez determinados os parâmetros, as ferramentas de Riemann são utilizadas para classificar os sinais a partir das matrizes de covariância dos sinais de EEG transformados.

Note que, se assume que a redução do número de condição

entre duas matrizes implica em aproximar essas matrizes no sentido de Riemann. Por conseguinte, com a redução do número de condição entre matrizes espera-se diminuir a dispersão em torno da matriz de covariância média e com isso melhorar a taxa de classificação correta utilizando as ferramentas de Riemann. É esse o problema descrito nesta seção.

### 6.3.1 O problema

Sejam  $\mathcal{A}, \mathcal{B}$  duas classes de experimentos tendo elementos do tipo  $C_i \triangleq \text{cov}\{\hat{Y}_i\} \in \mathbb{R}^{p \times p}$ , como expresso na Eq. (5.5), representando a matriz de covariância dos segmentos de sinais de EEG resultantes da projeção dos sinais na base senoidal. Suponha que  $n_a, n_b$  elementos são conhecidos para estas classes, respectivamente.

Note que a matriz de covariância dos sinais transformados  $C_i$  é linearmente dependente da matriz de parâmetros  $H_0$ . Assim, vamos começar definindo a matriz de covariância média de ambas as classes, definidas como  $C_a$  e  $C_b$ , respectivamente.

$$C_a = \frac{1}{n_a} \sum_{i=1}^{n_a} \text{cov}\{\hat{Y}_i\}, \quad C_b = \frac{1}{n_b} \sum_{i=1}^{n_b} \text{cov}\{\hat{Y}_i\} \quad (6.7)$$

Considere as desigualdades matriciais lineares que relacionam a matriz de covariância média com cada um dos elementos de cada classe. Para os elementos da classe  $\mathcal{A}$  tem-se,

$$\tau I \succ C_a^{-\frac{1}{2}} C_i C_a^{-\frac{1}{2}} \Leftrightarrow \tau C_a - C_i \succ 0 \quad , \quad C_i \in \mathcal{A} \quad (6.8)$$

$$\tau I \succ C_a^{\frac{1}{2}} C_i^{-1} C_a^{\frac{1}{2}} \Leftrightarrow \begin{bmatrix} \tau C_a & C_a \\ C_a & C_i \end{bmatrix} \succ 0 \quad , \quad C_i \in \mathcal{A} \quad (6.9)$$

e para os elementos da classe  $\mathcal{B}$ , tem-se,

$$\tau I \succ C_b^{-\frac{1}{2}} C_i C_b^{-\frac{1}{2}} \Leftrightarrow \tau C_b - C_i \succ 0 \quad , \quad C_i \in \mathcal{B} \quad (6.10)$$

$$\tau I \succ C_b^{\frac{1}{2}} C_i^{-1} C_b^{\frac{1}{2}} \Leftrightarrow \begin{bmatrix} \tau C_b & C_b \\ C_b & C_i \end{bmatrix} \succ 0 \quad , \quad C_i \in \mathcal{B} \quad (6.11)$$

onde  $\tau \in \mathbb{R}_+^*$ .

Observe que  $\tau$  é um limitante superior do número de condição para todas as matrizes das classes  $\mathcal{A}$  e  $\mathcal{B}$ , isto é,  $\tau \geq \kappa\{C_a^{-\frac{1}{2}}C_iC_a^{-\frac{1}{2}}\}$  para todo  $C_i \in \mathcal{A}$ , assim como  $\tau \geq \kappa\{C_b^{-\frac{1}{2}}C_iC_b^{-\frac{1}{2}}\}$  para  $C_i \in \mathcal{B}$ .

Observe que a relação existente entre as restrições Eq. (6.8) e Eq. (6.10):  $(C_a^{-\frac{1}{2}}C_iC_a^{-\frac{1}{2}})^{-1} \triangleq C_a^{\frac{1}{2}}C_i^{-1}C_a^{\frac{1}{2}}$ ; assim como a relação entre Eq. (6.9) e Eq. (6.11):  $(C_b^{-\frac{1}{2}}C_iC_b^{-\frac{1}{2}})^{-1} \triangleq C_b^{\frac{1}{2}}C_i^{-1}C_b^{\frac{1}{2}}$ .

As equivalências expressas em Eq. (6.10) e Eq. (6.11) foram obtidas utilizando-se o complemento Schur descrito na seção 5.5.

Para fins de melhorar a taxa de classificação dos sinais de EEG entre as classes  $\mathcal{A}$  e  $\mathcal{B}$  a partir da matriz de covariância dos sinais transformados, a matriz de parâmetros  $H_0$  pode ser determinada de tal forma que  $\tau$  que satisfaz as restrições Eq. (6.8), Eq. (6.10), Eq. (6.9) e Eq. (6.11) seja o menor valor possível.

O problema de determinar a matriz de parâmetros pode ser formulado como uma problema de otimização que consiste em minimizar  $\tau$  sujeito a restrições lineares, do tipo,

$$\min_{H_0, \tau} \quad \tau \tag{6.12}$$

Sujeito à:  $H_0 \succeq 0$ , Eq. (6.8), Eq. (6.9), Eq. (6.10), Eq. (6.11)

O problema de minimizar  $\tau$  é um problema quasi-convexo (BOYD et al., 1994) e pode ser resolvido de forma iterativa, onde em cada iteração o valor de  $\tau$  é mantido constante e a matriz de parâmetros é ajustada.

Como já mencionado anteriormente, como  $\tau$  é otimizado para todos os elementos de ambas as classes, ele pode ser visto, na luz de Eq. (6.4) e Eq. (6.5), como um limite superior para o pior caso de dispersão em torno da média de matrizes de covariância dos sinais transformados.

Tenha em mente que o limitante superior  $\tau$  deve ser tão pequeno quanto possível, indicando uma baixa dispersão em torno do

valor médio para cada uma das classes, mas é também desejável que os valores médios das classes devem ser afastados, então é possível manter o valor de  $\tau$  fixo, como um limitante superior para a dispersão e separar os valores médios maximizando uma função objetivo de interesse que representa a separação desejada, por exemplo,  $\text{tr}(C_a - C_b)$  ou  $\text{tr}(C_b - C_a)$ .

Por exemplo, o problema de otimização expresso na Eq. (6.12) pode ser reformulado com o objetivo de incluir uma função objetivo que afaste as matrizes de covariância médias das classes, como segue,

$$\min_{H_0, \tau} (1 - \xi)\text{tr}(C_a - C_b) + \xi\tau \quad (6.13)$$

Sujeito à:  $H_0 \succeq 0$ , Eq. (6.8), Eq. (6.9), Eq. (6.10), Eq. (6.11)

onde  $\xi \in [0, 1]$  e expressa o peso de cada uma das parcelas que constituem a função objetivo.

Nesse contexto, diferentes limitantes superiores (valores de  $\tau$ ) podem ser considerados para cada classe, com o custo de perder a propriedade de quasi-convexidade do problema global, embora para valores fixos de  $\tau$  o problema seja convexo.

Quando nenhuma função objetivo é considerada e o problema é minimizar  $\tau$ , o problema em questão é quasi-convexo e é expresso na Eq. (6.12)

Neste Capítulo, apenas o problema expresso na Eq. (6.12) é resolvido. Então resolve-se o problema de tal forma que o valor de  $\tau$  seja minimizado e obtém-se a matriz de parâmetros  $H_0$ . Em seguida, já com a matriz  $H_0$  dada, utiliza-se as ferramentas de Riemann para realizar a classificação dos sinais de EEG entre as classes  $\mathcal{A}$  e  $\mathcal{B}$ .

Os resultados numéricos dessa abordagem aplicada em um conjunto de dados público disponibilizado para uma competição de sistemas ICM é discutido na próxima seção.

## 6.4 Resultados Numéricos

Nesta seção, ilustramos os resultados do método proposto usando um conjunto de dados da IV competição ICM <sup>1</sup>. O problema de otimização Eq. (6.12) é resolvido em um notebook com processador *quadri-core* I7 com 16GB de RAM, executando o Matlab com o resolvidor SeDuMi LMI e a interface Yalmip (LOFBERG, 2004).

O subconjunto 2a de dados EEG da competição IV de sistemas de ICM foi utilizado, ele inclui sinais de EEG de 9 sujeitos, cada um realizou quatro tarefas cognitivas diferentes, correspondentes à imaginação motora das mãos esquerda (LH) e direita (RH), dos pés (FE) e da língua (TO). Os sinais EEG foram coletados de  $p = 22$  sensores amostrados a uma taxa de amostragem de  $f_s = 250Hz$ . Dois conjuntos de dados estavam disponíveis: um deles é utilizado para treinamento, e o outro para testes, ambos contendo 72 ensaios para cada uma das tarefas cognitivas citadas. Os dados também incluíram 3 sensores para eletro-oculografia (EOG), que não foram considerados neste trabalho.

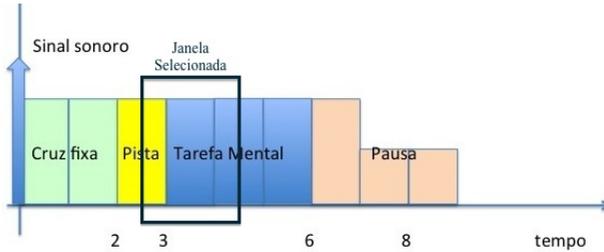
As janelas de sinal foram extraídas nos instantes de tempo entre 2,5 e 4,5 segundos após o sinal sonoro para cada tarefa, como em Lotte et al. (2007b) e indicado na Figura 21, produzindo um total de  $q = 500$  amostras por segmento e  $n_a + n_b = 144$  ensaios, com  $n_a = n_b$ . As frequências de corte inferior e superior foram definidas como  $f_l = 8Hz$  e  $f_u = 30Hz$ , respectivamente. O espaçamento entre duas frequências consecutivas foi  $\delta_f = 0,5$ , resultando em  $m = 45$  frequências distintas.

As janelas (segmentos de sinal) foram classificadas utilizando a estratégia um-contra-um, e por simplicidade, as classes foram escolhidas para ser LH  $\times$  RH para cada sujeito. O problema de otimização expresso na Eq. (6.12) foi resolvido para o conjunto de treinamento disponível, o que resulta em um par  $(\tau, H_0)$  para cada sujeito. Assim, um único ensaio foi classificado com base na sua distância riemanniana das matrizes de covariância média associadas às classes LH ou RH, con-

---

<sup>1</sup> <http://www.bbci.de/competition/iv/dataset2a>

Figura 21: Cronologia da seqüência de instruções apresentadas aos indivíduos sob coleta de sinais de EEG. Destaque para o intervalo de tempo onde os segmentos de sinais de EEG foram extraídos para ajuste e teste dos algoritmos.



forme descrito na seção 3.5.2. Os resultados de classificação mostrados na Tabela 3 foram obtidos para o conjunto de dados de teste.

Para comparar os resultados, o cenário descrito aqui foi o mesmo utilizado em (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013), onde os sinais foram filtrados na faixa  $\Delta f$  antes da estimativa da matriz de covariância. Em contraste, a abordagem proposta utiliza a projeção em uma base senoidal para obter uma aproximação do sinal filtrado na mesma faixa de frequência  $\Delta f$ . A diferença entre essas abordagens é ilustrada nas Figuras 15 a e 15 b da seção 4.3.

Além disso, duas situações da abordagem proposta foram analisadas e os resultados são ilustrados na Tabela 3. A primeira situação é ilustrar o impacto da utilização dos sinais projetados em vez dos sinais filtrados na classificação e corresponde à escolha  $H_0$  como matriz de identidade. A segunda situação foi considerada para ilustrar o impacto de ajustar a matriz  $H_0$  para melhorar a classificação. Neste caso consideramos a matriz  $H_0$  como uma matriz pentadiagonal<sup>2</sup> e é determinada resolvendo o problema de otimização expresso na Eq. (6.12).

O problema de otimização expresso na Eq. (6.12) foi resolvido

<sup>2</sup> A matriz pentadiagonal é uma matriz em que os únicos elementos não nulos estão na diagonal principal e as duas primeiras diagonais acima e abaixo dela.

para cada um dos sujeitos resultando em valores particulares para  $\tau$  e para a matriz  $H_0$  para cada um dos sujeitos.

Note que os resultados obtidos com a matriz  $H_0$  sendo igual a matriz identidade são similares aos resultados reportados em (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013). Essa similaridade era esperada, pois a projeção dos sinais na base de sinais senoidais é uma boa aproximação do sinal filtrado, veja a Figura 15. A vantagem neste caso é que a representação dos sinais projetados é compacta, como discutido no Capítulo 4 e além disso, permite que uma transformação linear aplicada no espaço das amostras dos sinais seja parametrizada com um número reduzido de parâmetros, e estes por final, aparecem linearmente na matriz de covariância dos sinais após a transformação, como discutido no Capítulo 5.

Como esperado os resultados de classificação dos sinais exibem uma variação significativa entre os indivíduos estudados. No entanto, note que para os indivíduos número 5 e número 7, os resultados obtidos em (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013) são resultados de classificação baixos, mas a acurácia de classificação foi relativamente mais alta quando o método de otimização proposto neste Capítulo foi aplicado.

Entretanto, para sujeitos em que taxas de classificação muito boas foram alcançadas por (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013), i.e, para os sujeitos 1, 3, 8 e 9, o método proposto não consegue gerar resultados melhores, exceto para o sujeito 9. Neste caso, as características estão bem discriminadas e não há muito espaço para se melhorar os resultados. A função objetivo em Eq. (6.12) parece falhar em gerar resultados melhores. Este aspecto indesejado poderia ser superado se uma função objetivo fosse melhor definida para lidar com os médios da classe.

Observe que as LMIs em Eq. (6.8), Eq. (6.9), Eq. (6.10) e Eq. (6.11) expressam a minimização da dispersão das matrizes de covariância dos sinais em torno da matriz de covariância média de cada

Tabela 3: Resultados obtidos para os dados IIa da Competição IV de sistemas ICM. São mostrados a acurácia de classificação para três abordagens. [ $H_0 = I$ ] Para a matriz  $H_0$  igual a matriz identidade (equivale a substituir a filtragem convencional pela projeção dos sinais na base senoidal). [ $H_0$  opt ] Para a matriz  $H_0$  obtida a partir do problema de otimização e [Refer] para os resultados extraídos da referência (BARACHANT et al., 2012; CONGEDO; BARACHANT; ANDREEV, 2013)

Sujeitos	1	2	3	4	5	6	7	8	9	Média
$H_0 = I$	93.67	<b>65.17</b>	91.55	72.22	59.50	70.44	65.28	96.53	91.64	78.44
$H_0$ opt	92.96	60.54	87.32	70.14	<b>66.45</b>	<b>73.23</b>	<b>79.17</b>	93.06	<b>93.06</b>	79.54
Refer	<b>93.75</b>	<b>63.19</b>	<b>94.44</b>	<b>75.00</b>	63.19	71.53	72.92	<b>96.55</b>	91.67	80.23

classe. Contudo, uma formulação de uma função objetivo que force os valores médios de cada classe estarem o mais afastado possível um do outro é ainda um trabalho em progresso.

É possível que incorporar uma melhor relação entre os valores médios de ambas as classes dentro da função objetivo pode melhorar os resultados da metodologia proposta. Estes resultados melhores podem ser buscados em trabalhos futuros.

Por fim, destaca-se que a matriz  $H_0$  foi considerada sendo pentadiagonal. Preliminarmente, os resultados sugerem que ao aumentar o número de sub-diagonais de  $H_0$  ajuda a decrescer a função objetivo, mas não melhora as taxas de classificação obtidas. Este fato também sugere que a função objetivo não está corretamente refletindo os propostos de classificação e melhores resultados podem ser obtidos considerando restrições adicionais ao problema ou até com mudanças na função objetivo.

## 6.5 Conclusões do Capítulo

Neste Capítulo, a matriz de covariância dos sinais transformados é utilizada para representar os sinais de EEG para fins de classificação. Os parâmetros que parametrizam a transformação linear aplicada

nos sinais são expressos linearmente na matriz de covariância dos sinais transformados. Esses parâmetros são organizados como elementos de uma matriz de parâmetros simétrica e positiva semi-definida e adicionam um grau de liberdade a mais para ser determinado.

A ideia é determinar os elementos da matriz de parâmetros com o propósito de melhorar o estágio de extração de características de um sistema interface cérebro-máquina. Um problema de otimização é então formulado com este propósito e condições LMI são então criadas para reduzir a dispersão entre as matrizes de covariância dos sinais transformados com respeito à média entre as matrizes de covariância para cada classe considerada.

Isto é alcançado ao se minimizar um limitante superior para o número de condição das matrizes  $C^{-\frac{1}{2}}C_iC^{-\frac{1}{2}}$ , onde  $C_i$  denota uma matriz de covariância dos sinais transformados da classe e  $C$  a matriz de covariância média da classe. Observe que a distância de Riemann expressa o quão longe, em sentido geodésico, a matriz  $C^{-\frac{1}{2}}C_iC^{-\frac{1}{2}}$  está da matriz identidade de mesma dimensão, enquanto que o número de condição da matriz  $C^{-\frac{1}{2}}C_iC^{-\frac{1}{2}}$  tende a unidade quando as matrizes satisfazem a relação:  $C \approx C_i$ .

O problema de otimização, expresso na Eq. (6.12), é formulado com base no número de condição entre as matrizes de covariância dos sinais transformados. No entanto, a etapa de classificação dos sinais é realizada com base na distância de Riemann, logo pode haver uma lacuna entre a solução ótima da classificação via ferramentas de Riemann e a solução ótima fornecida pelo problema de otimização expresso na Eq. (6.12). Isso fica claro, por exemplo, ao se analisar os resultados de classificação do indivíduo 1, onde o resultado otimizado para  $H_0$  foi inferior ao caso particular quando  $H_0$  é identidade. A escolha particular de  $H_0$ , apresentada na seção 5.4, também produz melhores resultados de classificação.

Uma possibilidade para superar esta lacuna é utilizar a distância de Riemann para quantificar a dispersão entre as matrizes de

covariância dos sinais transformados de uma mesma classe, mas neste caso as ferramentas usuais de desigualdades lineares matriciais (LMIs) não podem ser utilizadas para resolver o problema diretamente. Alternativamente, é possível utilizar o número de condição no passo de classificação, mas as propriedades desse classificador não são conhecidas e devem ser estudadas.

Destaca-se que o método proposto tem contribuições importantes. Primeiramente, informações sobre o espectro dos sinais de EEG podem ser facilmente estimadas a partir das colunas da matriz de coeficientes  $\Gamma$  e ficam disponíveis na matriz de covariância dos sinais transformados,  $cov\{\hat{Y}_i\} = \Gamma H_0 \Gamma^T$  como discutido nos capítulos 4 e 5. Infelizmente, os métodos convencionais de classificação usados neste trabalho (LDA e Riemann) não permitem explorar esse fato.

O segundo ponto a ser destacado é que uma redução considerável no tamanho do problema a ser resolvido é obtida com a parametrização, como discutido no Capítulo 4. E finalmente, a possibilidade de utilizar diferentes matrizes de parâmetros para cada uma das classes e determinar a matriz de parâmetros  $H_0$  para reduzir a distância entre os elementos das classes, enquanto mantém distantes as matrizes de covariância médias das classes, esta é uma característica importante que precisa de mais pesquisa. Especialmente, no que diz respeito a formulação do problema que permita com que o problema de otimização possa ser resolvido com as ferramentas numéricas disponíveis.

Os resultados apresentados nesse Capítulo foram ilustrados através de um problema padrão de sistemas ICM baseados em imaginação motora e resultados encorajadores foram encontrados em termos de acurácia de classificação.



# 7 Determinação da Matriz de Parâmetros: abordagem CSP-LDA

## 7.1 Introdução

No Capítulo 6 foi descrito como formular um problema de otimização convexa com restrições LMI e resolvê-lo para determinar a matriz de parâmetros  $H_0$  com o objetivo de reduzir a dispersão entre as matrizes de covariância dos sinais transformados. Uma vez que a matriz de parâmetros é determinada, as ferramentas de Riemann foram utilizadas para classificar os sinais.

Nesse Capítulo, o problema de otimização expresso na Eq. (5.26) é particularizado considerando aspectos do algoritmo CSP, descrito na seção 3.3. O filtro espacial CSP é utilizado com sucesso em conjunto com o classificador LDA, descrito na seção 8.4, em problemas de classificação de sinais de EEG aplicados em sistemas de interface cérebro-máquina - ICM.

O algoritmo CSP determina transformações espaciais com base na matriz de covariância dos sinais de EEG filtrados em uma faixa de frequências de interesse. Originalmente, o algoritmo CSP considera as matrizes de covariância dos sinais, no entanto, os sinais são filtrados e possuem média nula (nível DC nulo), então a matriz de covariância coincide com a matriz de covariância.

Uma aproximação dos sinais filtrados é obtida ao se realizar a projeção dos sinais brutos em uma base de sinais senoidais definida em uma faixa de frequências de interesse.

Os coeficientes resultantes da projeção dos sinais são utilizados como representações compactas dos sinais. Duas vantagens podem ser destacadas: a primeira é a redução de dimensão na nova representação

e, a segunda, é a informação espectral dos sinais pode ser estimada diretamente.

Os parâmetros de uma transformação linear aplicada nos sinais são expressos linearmente na matriz de covariância do sinais transformados. Assim adiciona-se um novo grau de liberdade na matriz de covariância dos sinais transformados através de uma matriz auxiliar, aqui denominada de matriz de parâmetros.

Uma vez que os parâmetros estão linearmente expressos na matriz de covariância dos sinais transformados, problemas de otimização convexa com restrições LMIs podem ser formulados com o propósito de determinar a matriz de parâmetros para ressaltar características dos sinais favoráveis à classificação.

Nesse sentido, neste capítulo é formulado um problema de otimização convexa, inspirado no algoritmo CSP atuando em conjunto com o algoritmo LDA, para se determinar a matriz de parâmetros. Uma vez a matriz de parâmetros determinada, filtros espaciais via algoritmo CSP são calculados e em seqüência, o algoritmo LDA é utilizado para classificar os sinais.

## 7.2 Análise preliminar

Os algoritmos CSP e LDA funcionam bem porque eles atuam de forma complementar. Para sinais filtrados de EEG com média nula e que pertençam a duas classes diferentes, o algoritmo CSP determina uma transformação espacial que quando aplicada nos sinais gera novos sinais, denominados de componentes dos sinais. As componentes apresentam máxima variância para uma classe e simultaneamente, mínima variância para a outra classe.

Após a transformação espacial via CSP, as componentes dos sinais apresentam contraste de variância para classes diferentes, e assim, usa-se a energia média das componentes dos sinais para formar um vetor de característica discriminativo das classes.

Destaca-se que o vetor de característica compacta a informação de energia das componentes e por conseguinte, carrega consigo a informação discriminativa de classe.

De posse do vetor de característica, o algoritmo LDA determina um hiperplano de separação entre as classes. Essencialmente, o que o LDA realiza é uma projeção dos vetores de característica em um subespaço em que os dados são linearmente separáveis. Como visto anteriormente, o pré-requisito mínimo para a existência de um hiperplano de separação entre os dados é que a diferença entre os valores médios dos vetores de característica para classes diferentes deve ser diferente de zero (não nulo).

Assim uma forma possível para melhorar a taxa de classificação é separar ao máximo os valores médios do vetor de característica de cada uma das classes,  $m_{\mathbf{a}}$  e  $m_{\mathbf{b}}$  e, simultaneamente, aproximar ao máximo possível cada um dos vetores de características de seu respectivo vetor médio. Essencialmente, é o que se faz ao maximizar a função custo na Eq. (3.22) e obtém-se o vetor na Eq. (3.21).

Além disso, observe que o vetor de característica é obtido a partir dos elementos da diagonal da matriz de covariância dos sinais filtrados após a transformação espacial via filtro CSP na Eq. (3.17), onde a matriz  $W \in \mathbb{R}^{r \times p}$  é a matriz de transformação com dimensões reduzidas, como descrito na seção 3.3.

No entanto, pode-se definir o vetor de característica como função da matriz de parâmetros  $H_0$ . Com esse fim, substitui-se a matriz de covariância em Eq. (3.17) pela matriz de covariância dos sinais transformados e linearmente dependente da matriz de parâmetros  $H_0$ , definida na Eq. (5.5). Obtem-se, assim, o seguinte vetor de característica

$$z = \text{diag}(W\Gamma H_0\Gamma^T W^T). \quad (7.1)$$

Observa-se na Eq. (7.1) que o vetor de característica é função tanto da matriz  $W$ , quanto da matriz de parâmetros  $H_0$ .

Assim, a matriz de parâmetros  $H_0$  pode ser determinada para

se atingir uma separação máxima entre os vetores de característica médios de cada classe e também, reduzir a dispersão entre cada vetor de característica e o médio de cada classe.

A partir das considerações acima, vários problemas de otimização podem ser formulados, dependendo de como o vetor de característica médio entre as classes é separado e de como a dispersão em torno do vetor médio das classes é minimizada.

### 7.3 Separar o vetor de característica médio

No algoritmo CSP descrito na seção 3.3, a matriz de transformação é composta pelo produto de duas matrizes, como definido na Eq. (3.13). A primeira delas, a matriz  $Q \triangleq \Lambda^{-\frac{1}{2}}U^T$  realiza uma normalização dos sinais de EEG filtrados. A segunda, projeta os sinais em direções com maior ou menor variância dependendo da classe a que pertence os sinais.

A matriz  $Q$  é definida a partir dos autovalores e autovetores da matriz  $M = M_{\mathbf{a}} + M_{\mathbf{b}}$ . Note ainda que a seguinte relação é válida:  $M^{-\frac{1}{2}}MM^{-\frac{1}{2}} = I$  assim como  $QMQ^T = I$ . Logo, a normalização dos sinais de EEG filtrados pode ser obtida a partir da raiz quadrada da inversa da matriz  $M$ , isto é, a matriz simétrica  $\tilde{Q} \triangleq M^{-\frac{1}{2}}$ .

Os vetores de característica médios das classes  $\mathcal{A}$  e  $\mathcal{B}$ , isto é,  $m_{\mathbf{a}}$  e  $m_{\mathbf{b}}$  são obtidos a partir dos autovalores das matrizes  $\tilde{Q}M_{\mathbf{a}}\tilde{Q}$ ,  $\tilde{Q}M_{\mathbf{b}}\tilde{Q}$  que satisfazem a equação Eq. (3.10), ou seja,

$$\tilde{Q}(M_{\mathbf{a}} + M_{\mathbf{b}})\tilde{Q} = I \quad (7.2)$$

As direções discriminativas, no sentido do filtro CSP, são obtidas dos autovetores da matriz  $\tilde{M}_{\rho} \triangleq \tilde{Q}M_{\rho}\tilde{Q} = V\tilde{\Lambda}_{\rho}V^T$ , onde  $V$  e  $\tilde{\Lambda}_{\rho}$  são as matrizes de autovetores e autovalores de  $\tilde{M}_{\rho}$ , respectivamente, onde  $\rho \in \{\mathbf{a}, \mathbf{b}\}$ . Note ainda que a matriz  $\tilde{\Lambda}_{\rho}$  é diagonal e os seus elementos estão no intervalo entre  $[0, 1]$ . Assim sendo, a matriz de transformação do algoritmo CSP é como segue,

$$\tilde{W} \triangleq V^T\tilde{Q} \quad (7.3)$$

onde os vetores de característica médios são:  $m_{\boldsymbol{\rho}} \triangleq \text{diag}\{\tilde{W}M_{\boldsymbol{\rho}}\tilde{W}^T\}$ , onde  $\boldsymbol{\rho} \in \{\mathbf{a}, \mathbf{b}\}$ . Note que  $m_{\mathbf{a}} = \text{diag}\{\tilde{W}M_{\mathbf{a}}\tilde{W}\} = \text{diag}\{\tilde{\Lambda}_{\mathbf{a}}\}$ , assim como  $m_{\mathbf{b}} = \text{diag}\{\tilde{\Lambda}_{\mathbf{b}}\}$ .

O maior número de direções discriminantes (colunas da matriz  $V$ ) é obtido do CSP quando os autovalores de  $\tilde{M}_{\mathbf{a}}$  estão todos longe de  $\frac{1}{2}$  indicando que são autovalores pequenos para a classe  $\mathcal{A}$  e grandes para a classe  $\mathcal{B}$  ou vice-versa.

À luz das considerações acima, considere a diferença normalizada entre as médias das matrizes de covariância dos sinais transformados das classes  $\mathcal{A}$  e  $\mathcal{B}$ , isto é, a matriz  $\tilde{Q}(M_{\mathbf{a}} - M_{\mathbf{b}})\tilde{Q}$ . Utilizando a igualdade definida na Eq. (7.2), pode-se verificar que,

$$\tilde{Q}M_{\mathbf{a}}\tilde{Q} - \frac{1}{2}I = \frac{1}{2}\tilde{Q}(M_{\mathbf{a}} - M_{\mathbf{b}})\tilde{Q} \quad (7.4)$$

Observe que cada autovalor de  $\tilde{Q}M_{\mathbf{a}}\tilde{Q}$  menor do que  $\frac{1}{2}$  corresponde a um autovalor negativo da matriz  $M_{\mathbf{a}} - M_{\mathbf{b}}$ . A rigor, como os autovalores da matriz  $\tilde{Q}M_{\mathbf{a}}\tilde{Q}$  estão entre  $[0, 1]$ , os autovalores de  $\frac{1}{2}\tilde{Q}(M_{\mathbf{a}} - M_{\mathbf{b}})\tilde{Q}$  estão no intervalo entre  $[-1, 1]$ .

Então uma maneira simples de separar os valores médios das classes  $\mathcal{A}$  e  $\mathcal{B}$  é maximizar (ou minimizar) a soma dos autovalores de  $M_{\mathbf{a}} - M_{\mathbf{b}}$ .

Isto nos dá a função objetivo  $J(H_0) = \text{tr}(M_{\mathbf{a}} - M_{\mathbf{b}})$  que deve ser minimizada. A maximização também é possível, neste caso maximizando os autovalores de  $\tilde{Q}M_{\mathbf{a}}\tilde{Q}$  para serem maiores do que  $\frac{1}{2}$ . Estas duas possibilidades devem ser testadas para um dado problema.

## 7.4 Reduzir a dispersão

Separar os valores médios do vetor de característica é algo desejável que pode ser alcançado com a determinação da matriz de parâmetros  $H_0$ . No entanto, reduzir a dispersão entre os elementos de cada classe em relação ao médio da classe deve ser outro objetivo quando se deseja classificar bem.

Uma possibilidade para se reduzir a dispersão entre os elementos de uma classe com relação ao valor médio da respectiva classe é o erro normalizado ao quadrado entre cada elemento e o médio. Sendo assim, seja  $C_i \triangleq \Gamma_i H_0 \Gamma_i^T$  e  $M_\rho$  é a matriz de covariância média dos sinais transformados da classe  $\rho$ , onde  $\rho \in \{\mathbf{a}, \mathbf{b}\}$  e indica a classe  $\mathcal{A}$  ou  $\mathcal{B}$ , respectivamente. A matriz de erro normalizado é definida como segue,

$$E_i = M^{-\frac{1}{2}}(M_\rho - C_i)M^{-\frac{1}{2}} \quad (7.5)$$

onde  $M = M_{\mathbf{a}} + M_{\mathbf{b}}$ . Note ainda que a matriz  $E_i$  é simétrica.

Pode-se utilizar o valor singular máximo da matriz de erro ao quadrado  $E_i^2$  para definir o pior caso de dispersão do erro ao quadrado entre cada elemento da classe e o valor médio de cada classe. Mais precisamente, estamos interessados em minimizar o escalar  $\tau$  tal que,

$$\tau I \succ (M^{-\frac{1}{2}}(M_{\mathbf{a}} - C_i)M^{-\frac{1}{2}})^2 \quad , \quad C_i \in \mathcal{A} \quad (7.6)$$

$$\tau I \succ (M^{-\frac{1}{2}}(M_{\mathbf{b}} - C_i)M^{-\frac{1}{2}})^2 \quad , \quad C_i \in \mathcal{B} \quad (7.7)$$

As desigualdades não-lineares acima podem ser re-escritas como LMIs como segue,

$$\begin{bmatrix} \tau M & M_{\mathbf{a}} - C_i \\ M_{\mathbf{a}} - C_i & M \end{bmatrix} \succ 0 \quad , \quad C_i \in \mathcal{A} \quad (7.8)$$

e

$$\begin{bmatrix} \tau M & M_{\mathbf{b}} - C_i \\ M_{\mathbf{b}} - C_i & M \end{bmatrix} \succ 0 \quad , \quad C_i \in \mathcal{B} \quad (7.9)$$

onde  $\tau$  indica o pior caso de dispersão entre os elementos de qualquer classe.

O problema de minimizar  $\tau$  é um problema quasi-convexo. Mesmo se a dispersão  $\tau$  deva ser tão pequena quanto possível, lembre-se de que estamos interessados em minimizar  $tr(M_{\mathbf{a}} - M_{\mathbf{b}})$ . Então a ideia é fixar o valor de  $\tau$  como um limitante máximo para a dispersão e otimizar a matriz de parâmetros  $H_0$  para minimizar a função objetivo:  $tr(M_{\mathbf{a}} - M_{\mathbf{b}})$ .

Nesse contexto, diferentes valores de  $\tau$  podem ser considerados para cada uma das classes, com a penalidade de perder a propriedade

de quasi-convexidade do problema global, mesmo que para valores fixos de  $\tau$  o problema seja convexo.

O problema de otimização acima pode ser representado como,

$$\min_{H_0} \quad tr(M_{\mathbf{a}} - M_{\mathbf{b}}) \quad (7.10)$$

Sujeito à:  $H_0 \succ 0$ , Eq. (7.8), Eq. (7.9)

Mesmo que o problema acima seja convexo (para um dado valor de  $\tau$ ), o tamanho do problema pode ser tornar muito grande dependendo do tamanho da matriz  $H_0$  a ser sintonizada, do tamanho das matrizes de covariância e do número de elementos das classes  $\mathcal{A}$  e  $\mathcal{B}$ . Quando este é o caso, uma maneira de contornar este problema computacional é impor uma estrutura diagonal para a matriz  $H_0$ . Isto reduzirá de forma significativa o esforço computacional para resolver o problema com o prejuízo de reduzir os graus de liberdade do método, sendo esperado resultados conservadores.

Uma vez que a matriz  $H_0$  é determinada ao se resolver o problema Eq. (7.10) para o conjunto de treinamento, os métodos clássicos CSP e LDA são aplicados para fins de classificação dos sinais utilizando o vetor de característica definido em Eq. (7.1).

## 7.5 Resultados numéricos

Na Tabela 4, os resultados de classificação alcançados com a metodologia proposta são apresentados para 9 sujeitos do conjunto de dados. As taxas de classificação alcançadas pela abordagem clássica utilizando CSP e LDA correspondem ao caso onde a matriz  $H_0$  é identidade e, nesse caso, não há a necessidade de se resolver o problema de otimização em Eq. (7.10).

Em ambos os casos, a Tabela 4 contém os resultados para o estágio de validação (conjunto de dados de teste). O parâmetro  $\tau$  indica a dispersão considerada em Eq. (7.8) e Eq. (7.9).

Tabela 4: Resultados em termos da acurácia de classificação obtidos para os dados IIA da competição IV de sistemas ICM.  $[\tau]$  Indica o pior caso de dispersão para as classes LH e RH. [proposto] Resultados obtidos para a matriz  $H_0$  é resultante da solução do problema de otimização. [clássico] Resultados para a matriz  $H_0$  igual a matriz identidade.

Sujeitos	1	2	3	4	5	6	7	8	9	Média
$\tau$	2.0	2.0	1.5	2.0	1.4	2.0	1.9	1.8	2.3	-
proposto	85.23	<b>61.43</b>	82.42	<b>73.61</b>	<b>62.70</b>	<b>70.40</b>	75.69	<b>96.53</b>	90.20	77.57
clássico	<b>88.06</b>	49.66	<b>97.18</b>	71.53	59.05	64.66	<b>77.78</b>	92.36	<b>93.05</b>	77.03

O método proposto corresponde a aplicar o método clássico CSP mais LDA usando o vetor de característica na Eq. (7.1), enquanto o método clássico consiste em aplicar o algoritmo CSP seguido do algoritmo LDA baseado no vetor de característica definido na Eq. (3.17) com a matriz de parâmetros  $H_0$  sendo a matriz identidade. Os métodos clássico e proposto foram sintonizados para cada um dos sujeitos.

Os resultados mostram que resultados consideravelmente melhores foram obtidos nos casos em que a abordagem clássica fornece resultados de classificação ruins. Este é o caso dos sujeitos 2,4,5 e 6. Os casos em que bons resultados foram alcançados com a abordagem clássica, sujeitos 1,3,8,9, o método proposto não foi capaz de superar o método clássico.

Para o sujeito 7, os resultados para ambos os métodos são simplesmente aceitáveis. Na média entre todos os sujeitos os resultados foram similares. Deve ser mencionado que os resultados obtidos com o método proposto são baseados em uma matriz de parâmetros  $H_0$  diagonal. Importantes graus de liberdade do método foram perdidos quando a estrutura diagonal foi imposta à matriz  $H_0$  para reduzir o custo computacional na hora de resolver as LMIs na Eq. (7.10).

O caso onde a restrição de matriz diagonal para a matriz  $H_0$  é removida, não poderia ser considerada para todo o conjunto de dados por causa da limitação do espaço de memória do computador utilizado.

Como mencionado anteriormente, as LMIs e a função objetivo do problema de otimização em Eq. (7.10) foram escolhidas para melhorar os métodos CSP seguido do LDA sintonizando apropriadamente a matriz de parâmetros  $H_0$ .

Diferentes possibilidades de função objetivo e restrições LMI podem ser consideradas com propósitos similares. Isto é um aspecto interessante da metodologia proposta no sentido de que há espaço para se obter melhores resultados. O que continua sendo o foco de investigação do grupo de pesquisa.

## 7.6 Considerações do Capítulo

Alguns aspectos interessantes do método proposto devem ser enfatizados. Primeiramente, para trazer mais graus de liberdade ao método, uma transformação nos sinais é aplicada e os parâmetros dessa transformação são definidos a partir de uma matriz  $H_0$  que aparece linearmente na matriz de covariância dos sinais. A ideia é sintonizar essa matriz para melhorar as taxas de classificação da abordagem utilizando o filtro espacial CSP seguido do classificador linear LDA em tarefas de classificação de sinais de EEG aplicados a sistemas ICM. O problema de determinar a melhor transformação é então formulado como um problema de otimização via LMI que pode ser resolvido eficientemente via ferramentas numéricas (LOFBERG, 2004; STURM, 1998; TOH et al., 1998).

O desempenho do método proposto, onde a matriz de parâmetros  $H_0$  é obtida como solução do problema expresso na Eq. (7.10), é comparado com o método clássico, onde a matriz  $H_0$  é igual a matriz identidade. Em ambas as abordagens, os algoritmos CSP e LDA são utilizados como em um problema de classificação padrão consistindo de um conjunto de dados obtidos de vários sujeitos onde tarefas de imagética motora são classificadas a partir de sinais de EEG. Os resultados sugerem um bom potencial para o método proposto nesse Capítulo.

O método ainda pode ser aperfeiçoado. A ideia é encontrar

mais restrições LMIs e a função objetivo do problema de otimização que determinem a matriz de parâmetros  $H_0$ . Várias possibilidades existem e mais pesquisa é necessária para encontrar a que forneça melhores resultados de classificação. Este ponto está sendo investigado.

Nesse sentido, no próximo Capítulo o conteúdo frequencial disponível explicitamente na representação compacta dos sinais é utilizado para compor sub-conjuntos de faixas de frequências (sub-bandas). Após aplicar os algoritmos CSP e LDA em cada sub-banda o resultado é utilizado para formar um vetor de característica aumentado. Assim, um algoritmo de classificação baseado em máquina de vetor suporte (SVM) é utilizado para classificar os sinais.

# 8 Sub-bandas com Projeção na Base Senoidal

## 8.1 Introdução

Em um método padrão de CSP-LDA (LOTTE; GUAN, 2011), os sinais são filtrados com um filtro passa-faixa, as matrizes de covariância desses sinais filtrados são calculadas e utilizadas como entrada para o algoritmo CSP que extrai características espaciais relevantes delas. Essas características são usadas por um classificador LDA que fornecerá a classe que um sinal pertence (o que significa que a tarefa cognitiva é realizada pelo usuário).

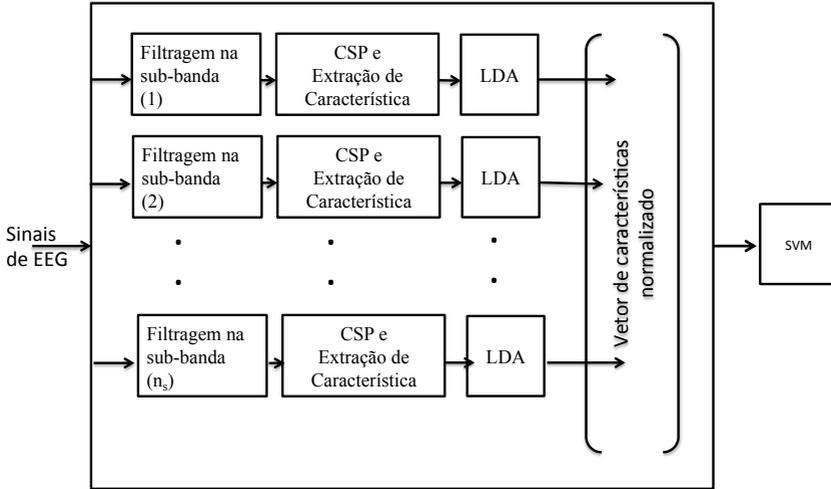
Em vez de usar apenas um filtro passa-faixa, a ideia neste Capítulo é dividir a faixa de frequência de interesse em um conjunto de sub-bandas e aplicar o método CSP-LDA para cada sub-banda e levar os resultados dos classificadores individuais para um classificador global e assim obter a classificação final. O esquema geral é representado na Figura 22, adaptado de (NOVI et al., 2007a).

Um aspecto interessante é notar que a informação frequencial dos sinais utilizada para formar o conjunto de sub-bandas pode ser obtida a partir da representação compacta dos sinais de EEG, discutido na seção 4.4.

Observe neste esquema que a cadeia CSP-LDA e um filtro passa-banda são substituídos por um banco de filtros resultando em um grupo de matrizes de covariância, uma para cada sub-banda considerada. Essas informações são processadas em paralelo, e são utilizadas múltiplas cadeias CSP-LDA, cada uma sendo treinada por sinais filtrados na faixa de frequência especificada.

Os resultados fornecidos pelo LDA aplicado às sub-bandas in-

Figura 22: Estrutura de um método baseado em Sub-bandas adaptado de (NOVI et al., 2007a).



dividuais não são utilizados como resultado final, como é o caso da abordagem padrão (LOTTE; GUAN, 2011). Cada LDA terá sua própria saída, que chamaremos pontuação. As pontuações obtidas em cada uma das sub-bandas são concatenadas em um único vetor por um bloco chamado "score fusion", e, em seguida, usadas como um vetor de características para treinar um classificador global baseado no método SVM.

Como foi descrito na seção 4.4, o foco neste Capítulo é a substituição do estágio de filtragem temporal por uma projeção dos sinais de EEG em uma base senoidal. Isto significa que a matriz de covariância usada para o algoritmo CSP não será calculada a partir do sinal no tempo, mas dos coeficientes que representam os sinais numa base senoidal e por isso carregam consigo informação frequencial. As operações restantes com os sinais são calculadas da mesma maneira como

proposto no método de sub-banda usual (NOVI et al., 2007a).

## 8.2 CSP via projeção na Base Senoidal

Em sistemas ICM baseado em imaginação motora, a informação espacial associada aos sinais é levada em conta usando o algoritmo de Padrões Espaciais Comuns (CSP) (MENG et al., 2013). O CSP procura uma transformação linear baseada nas matrizes de covariância média dos sinais EEG filtrados de tal forma que as características que definem cada classe são realçadas.

Para apresentar a versão do algoritmo CSP com base na matriz de coeficientes  $\Gamma$ , consideramos o algoritmo aplicado ao sinal projetado  $\hat{X}$  na faixa de frequência de interesse ao invés de aplicá-lo ao sinal filtrado  $X$  como é usualmente o caso. Observe que as matrizes de covariância e de correlação coincidem quando os sinais têm média zero, que é o caso do sinal filtrado  $X$ , e conseqüentemente, o sinal projetado  $\hat{X}$ .

Seja  $\mathcal{A}$ ,  $\mathcal{B}$  duas classes de experimentos, indicadas pelos rótulos  $\rho \in \{a, b\}$ , com  $n_e$  elementos do tipo  $cov_\ell(\hat{X}_i) \in \mathbb{R}^{p \times p}$  representando a matriz de covariância de um segmento de sinal de EEG na sub-banda  $\ell$ , como definido na Eq. (4.25), com  $p$  sensores. Suponha que os elementos  $n_a, n_b$  sejam conhecidos para essas classes, respectivamente e  $n_t = n_a + n_b$ .

As matrizes de covariância média de ambas as classes em cada sub-banda serão denotadas por  $C_a^{(\ell)}$  and  $C_b^{(\ell)}$ , respectivamente.

$$C_a^{(\ell)} = \frac{1}{n_a} \sum_{i=1}^{n_a} cov_\ell(\hat{X}_i), \quad \forall \hat{X}_i \in \mathcal{A} \quad (8.1)$$

e

$$C_b^{(\ell)} = \frac{1}{n_b} \sum_{i=1}^{n_b} cov_\ell(\hat{X}_i), \quad \forall \hat{X}_i \in \mathcal{B} \quad (8.2)$$

para  $\ell = 1, 2, \dots, n_s$ .

A transformação linear que define o filtro espacial CSP pode ser obtida a partir de um problema generalizado de autovalores (GEVP) (LOTTE; GUAN, 2011). Assim, obtemos

$$C_a^{(\ell)} v_j^{(\ell)} = \lambda_j^{(\ell)} (C_a^{(\ell)} + C_b^{(\ell)}) v_j^{(\ell)}, \forall j = 1, \dots, p \quad (8.3)$$

onde  $\{\lambda_j^{(\ell)}, v_j^{(\ell)}\}$  é o par autovalor e autovetor da matriz  $(C_a^{(\ell)} + C_b^{(\ell)})^{-1} C_a^{(\ell)}$ , com  $\lambda_j^{(\ell)} \in \mathbb{R}$  e  $v_j^{(\ell)} \in \mathbb{R}^p$ .

Suponha que os autovalores estejam organizados como  $\{\lambda_1^{(\ell)} > \lambda_2^{(\ell)}, \dots, > \lambda_p^{(\ell)}\}$  e observe que os valores de  $\lambda_j^{(\ell)}$  estão normalizados entre  $[0, 1]$ . Os filtros espaciais são escolhidos como sendo os autovetores de  $(C_a^{(\ell)} + C_b^{(\ell)})^{-1} C_a^{(\ell)}$ , que correspondem aos seus autovalores maiores e menores e é dada por

$$W_\ell \triangleq [v_1^{(\ell)} \ \dots \ v_r^{(\ell)} \ ; \ v_{p-r+1}^{(\ell)} \ \dots \ v_p^{(\ell)}] \in \mathbb{R}^{p \times r} \quad (8.4)$$

onde  $r$  pares de filtros espaciais foram selecionados para compor  $W_\ell$ .

### 8.3 Extração de Características via Sub-bandas

As características extraídas são usualmente tomadas como a energia de cada componente projetada, em cada sub-faixa de frequência considerada. Neste caso, o vetor de características  $\mathbf{q}_i$  de um dado elemento  $\hat{X}_i$  é escolhido como os elementos da diagonal da matriz de covariância projetada e pode ser expresso como :

$$\mathbf{q}_i^{(\ell)} = \frac{1}{2} \text{vdiag}(W_\ell^T \text{cor}_\ell(\hat{X}_i) W_\ell) \in \mathbb{R}^r \quad (8.5)$$

onde  $\text{vdiag}(\cdot)$  representa um vetor cujos os elementos são os elementos da diagonal da matriz em  $(\cdot)$  e  $i = 1, 2, \dots, n_t$ .

### 8.4 Pontuação das Sub-bandas

Depois de definir o vetor de características como Eq. (8.5) para cada sub-banda, o próximo passo é marcar essas sub-bandas para compor um vetor de características aumentadas. Para este propósito, usaremos o método de Análise Discriminante Linear de Fisher (LDA) que,

conforme relatado em trabalhos anteriores (LOTTE et al., 2007b), produz os melhores resultados, mantendo o baixo custo computacional e a complexidade matemática para dados linearmente separáveis.

O método LDA encontra a direção  $\phi^*$ , indicada abaixo, para projetar os dados resultando na melhor separabilidade no espaço projetado.

Essa direção ótima é aquela que maximiza o critério de Fisher, que é definido para cada sub-faixa como:

$$J^{(\ell)}(\phi) = \frac{\phi_\ell^T S_\beta^{(\ell)} \phi_\ell}{\phi_\ell^T S_\omega^{(\ell)} \phi_\ell} \quad (8.6)$$

onde as matrizes  $S_\beta^{(\ell)}$  e  $S_\omega^{(\ell)}$  são a matriz de dispersão entre classes e a dispersão dentro da classe, respectivamente (KEINOSUKE, 1990).

A solução ótima para Eq. (8.6) é,

$$\phi_\ell^* = (S_\omega^{(\ell)})^{-1}(\mu_a^{(\ell)} - \mu_b^{(\ell)}) \quad (8.7)$$

onde os vetores  $\mu_a$  e  $\mu_b$  representam os vetores médios para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , respectivamente.

A saída do LDA é  $s_i^{(\ell)}$ , a pontuação do exemplo  $i$  na sub-banda com índice  $\ell$  é computado como,

$$s_i^{(\ell)} = \phi_\ell^T \varrho_i^{(\ell)} \quad (8.8)$$

onde o vetor  $\varrho_i$ , como discutido anteriormente, é o vetor de característica.

## 8.5 Classificador Bayesiano

Uma maneira de implementar a fusão de recursos de pontuação é usando classificadores Bayesianos. Assumimos que a distribuição condicional das pontuações para cada classe e para cada sub-banda é uma distribuição normal, e estimamos suas médias e variâncias a partir do conjunto de treinamento:

$$\mu_a^{(\ell)} = \frac{1}{n_a} \sum_{i=1}^{n_a} s_i^{(\ell)}, \quad \forall s_i \in \mathcal{A} \quad (8.9)$$

$$\mu_b^{(\ell)} = \frac{1}{n_b} \sum_{i=1}^{n_b} s_i^{(\ell)}, \quad \forall s_i \in \mathcal{B} \quad (8.10)$$

onde  $\mu_\rho^{(\ell)}$ , para  $\rho \in \{a, b\}$  é o valor médio das classes  $\mathcal{A}$  ou  $\mathcal{B}$ , respectivamente. As variâncias para as classes  $\mathcal{A}$  e  $\mathcal{B}$  são estimadas como segue,

$$\sigma_a^{(\ell)} = \sqrt{\frac{1}{n_a} \sum_{i=1}^{n_a} \left( \mu_a^{(\ell)} - s_i^{(\ell)} \right)^2}, \quad \forall s_i \in \mathcal{A} \quad (8.11)$$

e

$$\sigma_b^{(\ell)} = \sqrt{\frac{1}{n_b} \sum_{i=1}^{n_b} \left( \mu_b^{(\ell)} - s_i^{(\ell)} \right)^2}, \quad \forall s_i \in \mathcal{B} \quad (8.12)$$

Definimos a saída de classificadores Bayesianos como uma meta-pontuação. É expresso como o logaritmo da razão de verossimilhança entre as duas classes:

$$x_i^{(\ell)} = \log \left( \frac{\mathcal{L}(\mathcal{A}|s_i^{(\ell)})}{\mathcal{L}(\mathcal{B}|s_i^{(\ell)})} \right). \quad (8.13)$$

Na prática, isso significa avaliar a função de densidade de probabilidade das distribuições normais obtidas com ambas as classes na pontuação dada pelo LDA:

$$\mathcal{L}(\mathcal{A}|s_i^{(\ell)}) = p_a^{(\ell)} \left( s_i^{(\ell)} \right) = \frac{1}{\sigma_a^{(\ell)} \sqrt{2\pi}} \exp \left( -\frac{(s_i^{(\ell)} - \mu_a^{(\ell)})^2}{2 \left( \sigma_a^{(\ell)} \right)^2} \right) \quad (8.14)$$

No entanto, de acordo com (NOVI et al., 2007a), esta solução não será ótima se as matrizes de covariância para as classes forem diferentes (DUDA; HART; STORK, 2012).

Portanto, usamos um classificador adicional, SVM no nosso caso, para ser colocado na saída do classificador Bayesiano. Este meta-classificador é proposto para compensar os erros produzidos por classificadores Bayesianos individuais, tendo em conta possíveis relações de alto nível. Este método evita o esquema iterativo na fase de treinamento que está presente no método RBE (GUYON et al., 2002), também descrito em (NOVI et al., 2007a).

Além disso, podemos agora escrever

$$\mathbf{x}_i = \left[ x_i^{(1)} \dots x_i^{(n_s)} \right]^T$$

Um vetor de tamanho  $n_s$ , como um vetor de característica que contém as meta-pontuações  $x_i^{(\ell)}$  para todas as sub-bandas.

## 8.6 Máquina de Vetor Suporte - SVM

Máquinas de vetor suporte (SVM, do inglês, para *Support Vector Machine*) é uma classe popular de classificadores que tentam maximizar uma margem de separação entre classes diferentes. Ele tem uma característica notável de generalizar bem em dados (BURGES, 1998).

A utilização anterior de SVMs em ICM inclui (LAL et al., 2004) (para seleção automática de canais em um sistema baseado em imagética motora) e (RAKOTOMAMONJY; GUIGUE, 2008) (onde um conjunto de SVMs é usado para resolver um problema P300, BCI).

Neste trabalho usamos uma implementação SVM linear fornecida por LIBSVM (CHANG; LIN, 2011) envolvida em um método *scikit-learn* (PEDREGOSA et al., 2011). O parâmetro de regularização  $C$ , o único grau de liberdade disponível, é ajustado por um algoritmo descrito na Seção 8.7.

Nós usamos  $\mathbf{x}_i$  como foi definido na seção 8.5 como vetores de característica, e os rótulos correspondem às classes  $\mathcal{A}$  e  $\mathcal{B}$ . A classe prevista deste SVM para um dado vetor de característica  $\mathbf{x}_i$  é a saída final do nosso sistema.

## 8.7 Método Adotado para o Ajuste do Algoritmo

O ajuste do algoritmo em cada passo do método SBCSP proposto neste trabalho, ou em qualquer tentativa de ajuste manual, pode afetar significativamente o resultado da classificação final. Os parâmetros que podem ser ajustados no método proposto são *e.g.* a banda de frequência total adotada, o número de sub-bandas dentro desta faixa,

o parâmetro de regularização SVM, o número de filtros espaciais no método CSP e a resolução de frequência usada para fazer a matriz de base. A estratégia convencional de ajustar esses parâmetros manualmente não seria viável, pois requer um grande número de ensaios para obter uma solução ótima conjunta. Assim, neste trabalho, incorporamos um método de otimização Bayesiana (SNOEK; LAROCHELLE; ADAMS, 2012) para encontrar esses parâmetros. Este algoritmo de otimização modela a precisão de generalização de um classificador como uma função dos parâmetros correspondentes e localiza os parâmetros ótimos que maximizam a precisão esperada dado o conjunto de dados observado. Se, por um lado, o uso deste algoritmo de otimização simplifica o problema, não é possível garantir soluções globais.

Em nossa implementação, usamos Hyperopt (BERGSTRA; YAMINS; COX, 2013), uma biblioteca Python para otimizar hiperparâmetros em algoritmos de aprendizado de máquina, com a árvore de Parzen Estimators (TPE) (BERGSTRA et al., 2011), um algoritmo que cai na classe de otimização baseada em modelo sequencial (SMBO) (HUTTER; HOOS; LEYTON-BROWN, 2011). O algoritmo TPE realiza a validação cruzada com os nossos conjuntos de dados e encontra um conjunto ideal de valores de parâmetros.

## 8.8 Experimentos usando os dados IVa da III Competição de Sistemas ICM

Nesta seção, apresentamos os resultados para os métodos propostos usando o conjunto de dados da competição III de BCI <sup>1</sup>, um problema *benchmark* em BCI discutido em (BLANKERTZ et al., 2006).

### 8.8.1 Descrição do conjunto de dados

O conjunto de dados IVa da competição III BCI inclui sinais de EEG obtidos de 5 sujeitos que realizaram duas tarefas mentais di-

<sup>1</sup> [http://www.bbci.de/competition/iii/desc\\_IVa.html](http://www.bbci.de/competition/iii/desc_IVa.html)

ferentes: imaginação do movimento da mão direita (R) e do pé direito (F). Os sinais foram registrados de  $p = 118$  canais amostrados a uma taxa de amostragem de  $100Hz$ . Cada indivíduo realizou ambas as tarefas 140 vezes, resultando em um total de 280 ensaios por sujeito, cada época tendo 4 segundos de duração.

### 8.8.2 Procedimento de avaliação adotado

Para avaliar o desempenho do método proposto, começamos por separar os dados por sujeito, emulando um cenário onde o sistema BCI seria treinado especificamente para um sujeito que seria o usuário. Para cada um dos sujeitos, realizamos uma validação cruzada de 10 subconjuntos, em todas as épocas e avaliamos o desempenho do sujeito pela média da taxa de erro no conjunto de teste de cada iteração. Para obter o desempenho global, a média em relação ao desempenho de cada sujeito foi obtida.

### 8.8.3 Pré-processamento dos sinais

Para todos os casos, as épocas foram extraídas do segmento de tempo entre 0,5 a 2,5 segundos após o rótulo de cada tarefa, resultando em um total de  $q = 200$  amostras no tempo por época.

### 8.8.4 Resultados e discussão

Realizamos um procedimento de seleção de modelo descrito em 8.7 para encontrar ajustes para os seguintes sistemas:

#### 8.8.4.1 Algoritmo CSP/LDA padrão como referência

Neste cenário, testar a nossa decomposição proposta usando um algoritmo padrão para BCI que não divide o sinal em múltiplas bandas. Aqui, os sinais serão classificados diretamente pelo LDA (LOTTE et al., 2007b; BLANKERTZ et al., 2008)

Tabela 5: Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao utilizar um algoritmo de referência baseado no CSP-LDA e no sistema equivalente mas utilizando uma abordagem de projeção, descrito no Capítulo 4.

Método	Abordagem	Banda Total	Bandas	Acurácia	Tempo (s)
CSP-LDA	Projeção	8-30	1	87.5%	6.2
CSP-LDA	Filtragem	8-30	1	89.71%	21.01

Tabela 6: Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao restringir o sistema a utilizar uma banda total entre 8 e 30 Hertz e ajustar o número de sub-bandas.

Método	Abordagem	Banda total	Bandas	Acurácia	Tempo (s)
SBCSP	Projeção	8-30	18	88.07%	51.15

Limitamo-nos a um total entre 8 e 30 Hertz, enquanto ajustamos o número de filtros espaciais utilizados no CSP, a resolução de frequência no caso de projeção e a ordem do filtro utilizado no caso de filtragem.

Na Tabela 5 são mostrados os resultados obtidos. Podemos ver aqui o impacto da substituição de uma fase de filtragem temporal por uma projeção linear: o tempo de computação é reduzido significativamente (quase 4 vezes) e a acurácia de classificação é equivalente.

#### 8.8.4.2 SBCSP com restrição no banco de filtros entre 8 and 30 Hz

Nestes testes pretendemos realizar uma comparação mais objetiva entre os métodos SBCSP e CSP-LDA utilizando a mesma faixa de frequências (entre 8 e 30 Hertz). Isso irá forçar as duas abordagens a usarem as mesmas informações.

Ajustamos o número de sub-bandas, o número de filtros espaciais usados no CSP, o parâmetro de regularização SVM e a resolução de frequência. Somente a abordagem de projeção foi utilizada para este cenário.

Os resultados são exibidos na Tabela 6. Comparando esse resultado com os da tabela 5, vemos que a acurácia de classificação é

Tabela 7: Resultados obtidos para o conjunto de dados IVa da Competição III de ICM ao restringir o sistema a utilizar uma banda total entre 0 e 50 Hertz com 24 sub-bandas.

Método	Abordagem	Banda total	Bandas	Acurácia	Tempo (s)
SBCSP	Projeção	0-50	24	89.75%	47.63
SBCSP	Filtragem	0-50	24	88.86%	847.83

equivalente, enquanto o custo computacional para executar o algoritmo é muito alto comparando a um simples algoritmo CSP-LDA.

Estes resultados sugerem que não há benefícios adicionais, em termos de acurácia de classificação, simplesmente usando a banda de frequência usual e dividindo-as em sub-bandas para serem analisadas em paralelo. No entanto, com a abordagem de sub-bandas, podemos adicionar faixas de frequência que normalmente não são usadas no sistema CSP-LDA para uma quantidade maior de informações, sem comprometer os resultados, pois o sistema rejeitará automaticamente bandas que não são úteis. Isso nos leva ao nosso próximo cenário.

#### 8.8.4.3 SBCSP com um banco de filtro com restrições de 24 bandas entre 0 e 50 Hz

Neste cenário usamos um ajuste semelhante ao usado em (NOVI et al., 2007a). O artigo usou toda a banda disponível (0-50 Hz) e dividiu-a em 24 sub-bandas com uma sobreposição de 50 %. O uso de toda a banda adiciona flexibilidade ao algoritmo SBCSP, e ajusta automaticamente as sub-bandas relevantes e mais importantes, rejeitando aquelas que não estão contribuindo para a discriminação das classes.

Com as informações adicionais contidas nas frequências baixas e altas, também podemos assumir que poderiam ser encontrados melhores resultados comparando esse cenário com o anterior, no qual os resultados foram mostrados na Tabela 6.

Aqui ajustamos o número de filtros espaciais no estágio CSP, o parâmetro de regularização SVM, a resolução de frequência na abordagem de projeção e a ordem do filtro na abordagem de filtragem.

Tabela 8: Resultados obtidos para os dados Iva da competição III de ICM ao ajustar automaticamente a sua banda total e o número de sub-bandas.

Método	Abordagem	Banda Total	Bandas	Acurácia	Tempo (s)
SBCSP	Projeção	5-35	10	89.71%	34.4
SBCSP	Filtragem	3-46	6	90.64%	142.73

Na Tabela 7 são mostrados os resultados obtidos para este cenário. Em primeiro lugar, podemos ver que ambos os métodos, baseados em ideias de filtragem e projeção, superaram ligeiramente o cenário que usa uma faixa de frequência limitada em termos de acurácia de classificação. Em segundo lugar, e mais importante, podemos ver que o custo computacional das etapas de treinamento e teste, medido pelo tempo, foi muito menor na abordagem baseada em projeção, por uma ordem de magnitude ainda maior do que no caso CSP-LDA.

Uma pergunta que podemos fazer é se o ajuste utilizando a faixa de frequência total possível para o algoritmo e 24 sub-bandas seria o melhor possível. Isso nos leva ao último cenário.

#### 8.8.4.4 SBCSP com o banco de filtros sem restrições

O objetivo deste último cenário é explorar o espaço de parâmetros do método SBCSP e encontrar novos ajustes ótimos, principalmente para o tamanho da faixa total e o número de sub-bandas utilizadas, testando as abordagens de filtragem temporal e de projeção.

Na Tabela 8 são mostradas as configurações que obtiveram os melhores desempenhos. O ajuste encontrado pelo nosso algoritmo de seleção de modelo foi em algum grau semelhante aos números classicamente utilizados para o problema, sem que se tenham feito melhorias significativas em comparação com a faixa total e o número de bandas utilizadas em (NOVI et al., 2007a). No entanto, pudemos encontrar ajustes que se comportaram de forma semelhante em termos de acurácia e tiveram um número reduzido de sub-bandas, reduzindo o custo computacional do treinamento e aplicando o algoritmo.

## 8.9 Experimentos usando dados do conjunto 2a da competição IV de ICM

Nesta seção, ilustramos os resultados propostos usando o conjunto de dados da competição IV de ICM <sup>2</sup>, uma referência para a competição ICM discutida, por exemplo, em (TANGERMANN et al., 2012).

### 8.9.1 Descrição dos dados

O conjunto de dados IIa inclui sinais de EEG de 9 sujeitos que realizaram quatro tarefas mentais diferentes: imaginação do movimento da mão esquerda (LH) e da mão direita (RH), dos pés (FE) e da língua (TO). Os sinais foram registrados de  $n = 22$  canais amostrados a uma taxa de amostragem de  $250Hz$ . Dois conjuntos de dados estavam disponíveis para treinamento e testes, contendo cada um 72 ensaios por sujeito para cada uma das tarefas mentais mencionadas acima, sendo cada ensaio com 4 segundos de duração. Neste conjunto de dados também estão incluídas 3 eletrodos para registro da atividade de eletro-oculografia (EOG), que não foram consideradas neste trabalho.

### 8.9.2 Procedimento de avaliação do sistema

Novamente, as seções de treinamento e teste foram separadas por sujeito, da mesma forma que foi feito para o conjunto de dados anterior. Os ensaios foram classificados utilizando a estratégia um-contrum (LH  $\times$  RH, por exemplo) para cada indivíduo. Como há uma divisão entre dados de treinamento e dados de teste, não realizamos validação cruzada. Para obter o desempenho final de um sistema, medimos a acurácia de classificação obtida no conjunto de teste de todos os sujeitos e todas as combinações de tarefas.

---

<sup>2</sup> <http://www.bbc.de/competition/iv/#dataset2a>

Tabela 9: Resultados obtidos para o conjunto de dados 2a da Competição IV de ICM. A banda total e o número de filtros foram encontrados por um algoritmo de otimização de hiper-parâmetros no caso SBCSP. A coluna "tempo de computação" refere-se ao tempo de treinamento e teste de um sujeito em uma combinação de tarefas mentais.

Método	Abordagem	Banda total	Bandas	Acurácia	Tempo (s)
CSP-LDA	Filtragem	8-30	1	80.36%	0.55
SBCSP	Filtragem	0-51	33	79.75%	12.5
SBCSP	Projeção	0-51	33	84.87%	0.93

### 8.9.3 Pré-processamento

As épocas foram extraídas do segmento de tempo entre 2,5 e 4,5 segundos após o rótulo de cada tarefa, produzindo um total de  $q = 500$  amostras por época.

### 8.9.4 Resultados e discussões

Realizou-se um procedimento de seleção de modelo para encontrar ajustes para os seguintes sistemas:

- Algoritmo padrão CSP / LDA como linha de base
- SBCSP com um banco de filtros sem restrições, com a abordagem de decomposição

Depois, utilizamos o mesmo ajuste encontrado para o SBCSP com a abordagem de decomposição para testar a versão de filtragem temporal também. Os resultados são exibidos na Tabela 9.

A segunda e a terceira observações feitas na seção 8.8 se mantêm válidas para este conjunto de dados. Aqui, não só uma redução significativa do custo computacional é conseguida graças às ideias de projeção, mas também à acurácia geral do método que é melhorada por aproximadamente 5 % graças às ideias de sub-banda.

Também apresentamos os resultados nas Tabelas 10 e 11 por sujeito e tarefa para o algoritmo clássico CSP-LDA e SBCSP com as

Tabela 10: Porcentagem de acurácia por sujeito (indicado pela primeira linha) e tarefa (indicada pela primeira coluna) do algoritmo CSP-LDA clássico.

	1	2	3	4	5	6	7	8	9	
LR	82.64	63.89	92.36	67.36	56.94	64.58	77.78	93.05	91.67	76.7
LF	97.92	59.72	93.06	84.03	66.67	65.97	98.61	84.72	94.44	82.79
LT	96.53	63.19	93.06	88.89	70.83	70.14	96.53	90.28	95.83	85.03
RF	100	77.08	92.36	87.5	54.86	63.89	95.14	86.81	81.94	82.18
RT	100	63.89	97.22	82.64	61.11	70.14	95.14	88.89	81.94	82.33
FT	71.53	72.92	74.31	60.42	61.11	65.97	81.94	84.72	85.42	73.15
	91.44	66.78	90.39	78.47	61.92	66.78	90.86	88.08	88.54	<b>80.36</b>

Tabela 11: Percentual de acurácia por sujeito (indicado pela primeira linha) e tarefa (indicada pela primeira coluna) do algoritmo SBCSP com as alterações propostas.

	1	2	3	4	5	6	7	8	9	
LR	90.28	50.69	93.75	68.06	82.64	64.58	90.97	95.83	94.44	81.25
LF	97.92	84.72	95.83	83.33	65.97	75	98.61	94.44	95.83	87.96
LT	97.92	70.14	95.14	92.36	80.56	65.28	95.83	95.14	97.92	87.81
RF	98.61	79.17	97.22	80.56	68.75	70.14	99.31	88.19	83.33	85.03
RT	100	69.44	98.61	86.11	82.64	62.5	97.22	89.58	91.67	86.42
FT	80.56	81.25	86.11	75	65.97	78.47	86.11	82.64	90.97	80.79
	94.21	72.57	94.44	80.9	74.42	69.33	94.68	94.68	90.97	<b>84.87</b>

alterações propostas. Pode-se notar que não só a acurácia global do sistema é aumentada, mas no pior cenário (tanto para sujeitos ruins e tarefas difíceis, por exemplo, sujeitos 2 e 5 ou tarefas LR e FT) a acurácia de classificação é significativamente melhorada no sistema baseado em projeção, indicando uma robustez característica das projeções conjuntas de projeção e sub-banda propostas nesta tese.

## 8.10 Conclusões do Capítulo

Neste capítulo, é apresentada uma abordagem alternativa ao método de Padrões Espaciais Comuns por Sub-banda (SBCSP) para sistemas ICM. O método baseia-se numa decomposição dos sinais EEG numa base senoidal / cossenoidal. Essa abordagem permite uma representação mais compacta dos dados, resultando em um custo computacional reduzido para treinamento, teste e aplicação do método em um

ambiente real.

O sistema foi comparado com a sua contra parte de filtragem temporal e com o algoritmo CSP-LDA clássico em dois conjuntos de dados diferentes e verificamos que, de fato, o custo computacional é reduzido, enquanto que a acurácia geral pode ser melhorada quando comparada com o caso usual de uma sub-banda.

Um trabalho mais aprofundado na mesma direção seria adaptar esse método a um problema de várias classes.

## 9 Considerações Finais

Esta tese de Doutorado teve como objetivo desenvolver novas metodologias para o processamento de sinais de EEG aplicados a sistemas de interface cérebro-máquina baseados em imagética motora. As metodologias propostas tiveram como objetivo melhorar, principalmente, a etapa de extração de características dos sinais de EEG e como consequência reduzir o erro de classificação dos sinais.

A estratégia adotada nesse trabalho consistiu, primeiramente, em obter uma representação compacta dos sinais de EEG filtrados em uma faixa de frequências de interesse. Além da compactação, a nova representação apresenta outras propriedades interessantes. Ela conserva a informação espacial da representação original, intrínseca da distribuição dos sensores de EEG no couro cabeludo e também, expressa o conteúdo frequencial dos sinais diretamente.

A partir da representação compacta dos sinais de EEG e de uma base senoidal, o sinal de EEG filtrado, na mesma faixa de frequências em que a base senoidal é definida, pode ser obtido com erro de aproximação pequeno que depende do número de elementos da base. Além disso, a partir de um sub-conjunto de termos da representação compacta e uma sub-banda de frequências, caracterizada por um sub-conjunto de sinais consecutivos da base senoidal, podem ser obtidos sinais que aproximam os sinais de EEG filtrados na mesma sub-banda de frequências correspondente a nova base de sinais senoidais.

A fim de introduzir graus de liberdade que pudessem ser ajustados para melhorar a classificação dos sinais, uma transformação linear é aplicada na aproximação dos sinais de EEG obtida a partir da base de senoides e da representação compacta. Esses graus de liberdade são então expressos como parâmetros de uma matriz auxiliar com dimensões reduzidas e que aparecem linearmente na matriz de covariância

dos sinais transformados. Portanto, duas abordagens de processamento de sinais baseados em matriz de covariância, como o CSP-LDA e os que utilizam ferramentas da geometria de Riemann, puderam ter suas expressões rearranjadas em termos de restrições LMI e determinadas a fim de otimizar a classificação dos sinais.

Destaca-se ainda que, com o uso da representação compacta dos sinais de EEG, a dimensão da matriz auxiliar, denominada de matriz de parâmetros, foi reduzida sem perda de informação relevante. Assim a transformação linear, contendo os graus de liberdade do método, pôde ser obtida através da determinação da matriz de parâmetros via solução de problemas de otimização.

Uma vez que a matriz de parâmetros é linearmente expressa na matriz de covariância dos sinais de EEG após a transformação, problemas de otimização convexa com restrições LMI puderam ser formulados para determiná-la.

A vantagem em se determinar a matriz de parâmetros como solução de um problema de otimização convexo com restrições LMI é a disponibilidade de ferramentas numéricas capazes de resolvê-lo de forma eficiente. Para isto, basta, definir uma função objetivo e um conjunto de restrições LMI que traduzam, por exemplo, a tarefa de classificação dos sinais de EEG.

Com esse propósito, duas abordagens foram propostas. A primeira abordagem foi motivada pela expressão para o cálculo da distância de Riemann entre duas matrizes positivas. Essa expressão utiliza a soma do logaritmo ao quadrado dos autovalores generalizados entre duas matrizes positivas para quantificar a distância geodésica entre essas matrizes. Assim, um conjunto de restrições LMI foi criado com o objetivo de aproximar, o máximo possível, os autovalores extremos entre cada matriz de covariância dos sinais transformados e a média entre as matrizes de covariância dos transformados. O problema de otimização formulado é quasi-convexo. Os resultados alcançados com essa abordagem demonstraram que, ao adicionar mais graus de liberdade ao

problema e determiná-los de forma ótima, os resultados de classificação podem ser melhorados.

A matriz de parâmetros, nesse caso, como grau de liberdade extra foi determinada para reduzir o número de condição das matrizes e com isso, aproximar cada matriz de covariância dos sinais transformados da média entre as matrizes de covariância.

Na segunda abordagem, o método CSP seguido do classificador LDA foi utilizado como fonte inspiradora. O filtro espacial obtido via método CSP é determinado com base na média entre as matrizes de covariância para cada uma das classes. Com isso, supõe-se que o sucesso da transformação espacial é dependente de o quão cada matriz de covariância é semelhante (ou está próxima) da média das matrizes de covariância. Assim como o classificador LDA é dependente da diferença entre os vetores de característica médios. Diante dessas observações, o conjunto de restrições do problema de otimização foi formulado para reduzir a dispersão entre as matrizes de covariância em torno da matriz de covariância média e a função objetivo com base na função traço de matrizes foi utilizada para aumentar o afastamento entre os vetores de característica. Os resultados obtidos foram promissores, principalmente para indivíduos em que os métodos convencionais não funcionam bem. No entanto, a média da taxa de classificação correta, considerando todos os indivíduos estudados, não apresentou melhora significativa em relação aos métodos clássicos.

Nas duas abordagens, problemas de otimização foram formulados e resolvidos via LMIs para a determinação dos parâmetros livres e as taxa de classificação foram melhoradas, principalmente nos casos onde os métodos clássicos apresentam fraco desempenho.

Uma terceira abordagem foi proposta, nela a matriz de parâmetros não está envolvida. Aproveitou-se do fato de que a projeção dos sinais de EEG numa base senoidal pode ser utilizada para emular um banco de filtros. A ideia consistiu, basicamente, em aplicar um filtro espacial seguido de um classificador LDA para um conjunto de sub-

faixas de frequências. Em seguida, formar um vetor de característica aumentado e utilizar um classificador SVM. Mostrou-se que a etapa de filtragem convencional via filtros FIR ou IIR pode ser substituída, sem perda significativa de informação, pela projeção dos sinais na base senoidal. A principal vantagem é observada durante o treinamento do algoritmo, onde um número menor de operações é necessária para ajustar os algoritmos.

Apesar dos resultados promissores, é importante ressaltar alguns aspectos da metodologia proposta. Muito embora o objetivo seja otimizar os classificadores, através dos parâmetros livres oriundos da transformação, não foi possível até o momento encontrar condições convexas que expressem os classificadores sem conservadorismo. Outro ponto interessante a ser ressaltado é que muito embora a representação compacta dos sinais aqui proposta traga de forma explícita informações sobre o conteúdo frequencial dos sinais, os métodos de classificação aqui utilizados (CSP/LDA e Riemann) não exploram essas informações diretamente já que ficam restritas ao estágio inicial de filtragem dos sinais. Todas essas limitações da técnica aqui proposta são passíveis de serem melhoradas em trabalhos futuros.

Nesse sentido, seguem algumas abordagens sugeridas para trabalhos futuros.

1. Investigar novas funções objetivo e conjuntos de restrições, com o objetivo de afastar os valores médios foi investigada a função traço de matrizes. Uma outra função que pode ser investigada é a função convexa  $\log(\det(A))$  que pode ser útil na abordagem por Riemann. A proposta é estudar uma classe de problemas de otimização conhecida como maximização do determinante (VAN-DENBERGHE; BOYD; WU, 1998) no qual programação semi-definida e programação linear são casos especiais. O problema de maximização do determinante é definido como

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + \log(\det(G(\mathbf{x})^{-1})) \quad (9.1)$$

$$\text{sujeito à: } G(\mathbf{x}) \triangleq G_0 + x_1 G_1 + \cdots + x_k G_k \succ 0 \quad (9.2)$$

$$\text{sujeito à: } F(\mathbf{x}) \triangleq F_0 + x_1 F_1 + \cdots + x_k F_k \succ 0 \quad (9.3)$$

onde  $\mathbf{x} \in \mathbb{R}^k$  é a variável de decisão,  $\mathbf{c} \in \mathbb{R}^k$  é constante,  $G_i \in \mathbb{R}^{k \times k} / G_i = G_i^T$  e  $F_i \in \mathbb{R}^{k \times k} / F_i = F_i^T$ . O problema de otimização de maximização do determinante é convexo e pode ser resolvido eficientemente por meio do método do ponto interior (SIVALINGAM et al., 2014).

2. Motivado pela simplicidade e pelos bons resultados obtidos com o método apresentado na seção 5.4, seria interessante melhor estudá-lo em trabalhos futuros para poder extrair propriedades que expliquem esses resultados e permitam que o método seja melhorado.
3. Incorporar ao problema de otimização convexa as transformações espaciais e frequenciais simultaneamente, através, por exemplo, da seguinte formulação,

$$\min_{U, W, R} \text{tr}\{U\} \quad (9.4)$$

Sujeito à:

$$\begin{bmatrix} U & W^T \Gamma_i \\ \Gamma_i^T W & R \end{bmatrix} \succeq 0 \quad \forall i, \quad W, R \in \mathcal{W} \quad (9.5)$$

onde a matriz  $W$  é representa uma transformação espacial, a matriz  $R$  está associada à transformação frequencial, cuja inversa corresponde à matriz  $H_0$  apresentada nos capítulos anteriores, e  $\mathcal{W}$  é o conjunto de restrições do problema. Note que, isto é possível, pois a matriz de coeficientes  $\Gamma$  contém informações espaciais e frequenciais.

4. Investir a relação existente entre a projeção dos sinais na base de sinais senoidais com a transformada discreta de Fourier (DFT - *Discrete Fourier Transform*). Caso haja uma relação, espera-se

que a etapa de projeção dos sinais na base poderia ser substituída por algoritmos de obtenção da DFT, como por exemplo, o algoritmo da transformada rápida de Fourier (FFT - *Fast Fourier Transform*).

5. Finalmente, seria interessante encontrar uma maneira de incorporar o classificador diretamente na função objetivo do problema, evitando assim o conservadorismo da abordagem. Por exemplo, utilizando as ideias de classificação apresentadas no apêndice A. Encontrar formas de otimizar esse novo classificador via LMIs poderia gerar resultados interessantes.
-

# Referências

- AGHAEI, A. S.; MAHANTA, M. S.; PLATANIOTIS, K. N. Separable common spatio-spectral patterns for motor imagery bci systems. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 63, n. 1, p. 15, 2016. [29](#), [103](#)
- AMIRI, S.; FAZEL-REZAI, R.; ASADPOUR, V. A review of hybrid brain-computer interface systems. *Advances in Human-Computer Interaction*, n. 187024, 2013. [28](#)
- BALAKRISHNAN, V.; VANDENBERGHE, L. Linear Matrix Inequalities for signal processing: An overview. In: *Proceedings of the 32nd Annual Conference on Information Sciences and Systems*. Princeton, NJ: [s.n.], 1998. [105](#)
- BARACHANT, A.; BONNET, S.; CONGEDO, M.; JUTTEN, C. Common spatial pattern revisited by riemannian geometry. *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, p. 472–476, 2010. [65](#)
- BARACHANT, A.; BONNET, S.; CONGEDO, M.; JUTTEN, C. Multiclass brain-computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, v. 59, n. 4, p. 920–928, 2012. Disponível em: <http://ieeexplore.ieee.org/document/6046114/>. [15](#), [29](#), [66](#), [67](#), [70](#), [71](#), [73](#), [74](#), [123](#), [124](#), [125](#)
- BASHASHATI, A.; FATOURECHI, M.; WARD, R. K.; BIRCH, G. E. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, v. 4, n. 2, p. R32, 2007. [28](#), [57](#)
- BERGSTRA, J.; YAMINS, D.; COX, D. D. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In: CITESEER. *Proceedings of the 12th Python in Science Conference*. [S.l.], 2013. p. 13–20. [146](#)
- BERGSTRA, J. S.; BARDENET, R.; BENGIO, Y.; KÉGL, B. Algorithms for hyper-parameter optimization. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2011. p. 2546–2554. [146](#)

- BHATIA, R. The riemannian mean of positive matrices. In: F, N.; R, B. (Ed.). *Matrix Information Geometry*. [S.l.]: Springer, 2013. 66, 70, 71
- BLANKERTZ, B.; MULLER, K.-R.; KRUSIENSKI, D. J.; SCHALK, G.; WOLPAW, J. R.; SCHLOGL, A.; PFURTSCHELLER, G.; MILLAN, J. R.; SCHRODER, M.; BIRBAUMER, N. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE transactions on neural systems and rehabilitation engineering*, IEEE, v. 14, n. 2, p. 153–159, 2006. 146
- BLANKERTZ, B.; TANGERMANN, M.; VIDAURRE, C.; DICKHAUS, T.; SANNELLI, C.; POPESCU, F.; FAZLI, S.; DANOCZY, M.; CURIO, G.; MULLER, K.-R. Detecting mental states by machine learning techniques: the berlin brain-computer interface. In: GRAIMANN, B.; ALLISON, B.; PFURTSCHELLER, G. (Ed.). *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. [S.l.]: Springer Berlin Heidelberg, 2010. p. 113–135. 103, 104
- BLANKERTZ, B.; TOMIOKA, R.; LEMM, S.; KAWANABE, M.; MULLER, K. robert. Optimizing spatial filters for robust eeg single-trial analysis. In: *EEE Signal Proc. Magazine*. [S.l.: s.n.], 2008. p. 581–607. 29, 66, 147
- BLINOWSKA, K.; DURKA, P. Electroencephalography (eeg). In: \_\_\_\_\_. *Wiley Encyclopedia of Biomedical Engineering*. John Wiley & Sons, Inc., 2006. ISBN 9780471740360. Disponível em: <<http://dx.doi.org/10.1002/9780471740360.ebs0418>>. 43
- BOYD, S.; GHAOUI, L. E.; FERON, E.; BALAKRISHNAN, V. *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, EUA: SIAM, 1994. 30, 106, 120
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, Springer, v. 2, n. 2, p. 121–167, 1998. 145
- CHANG, C.-C.; LIN, C.-J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, v. 2, n. 3, p. 27, 2011. 145
- CONGEDO, M.; BARACHANT, A.; ANDREEV, A. A new generation of brain-computer interface based on riemannian geometry. *CoRR*, abs/1310.8115, 2013. Disponível em: <<http://arxiv.org/abs/1310.8115>>. 15, 30, 102, 123, 124, 125

- DORNHEGE, G.; BLANKERTZ, B.; KRAULEDAT, M.; LOSCH, F.; CURIO, G.; MULLER, K. R. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 53, n. 11, p. 2274, 2006. 28, 64
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012. 144
- DUDA, R. O.; HART, P. E.; STORK, D. G.; DUDA, C. R. O.; HART, P. E.; STORK, D. G. *Pattern Classification, 2nd Ed.* 2001. 64, 69
- DURKA, P. *Matching Pursuit and Unification in EEG Analysis*. [S.l.]: Artech House, inc., 2007. 83, 85, 90
- EBRAHIMI, T. Recent advances in brain-computer interfaces. In: *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*. 10.1109/MMSP.2007.4412807: [s.n.], 2007. p. 17–17. 27
- FATOURECHI, M.; BASHASHATI, A.; WARD, R. K.; BIRCH, G. E. {EMG} and {EOG} artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, v. 118, n. 3, p. 480 – 494, 2007. ISSN 1388-2457. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1388245706015124>>. 49, 50, 57
- FELZER, T. *On the Possibility of Developing a Brain-Computer Interface (BCI)*. Alexanderstr. 10, D-64283 Darmstadt, Germany, E-Mail: felzer@informatik.tu-darmstadt.de, 2001. 45
- Index. In: GANDHI, V. (Ed.). *Brain-Computer Interfacing for Assistive Robotics*. San Diego: Academic Press, 2015. p. 231 – 236. ISBN 978-0-12-801543-8. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780128015438000211>>. 28
- GROSSE-WENTRUP, M.; SCHOLKOPF, B.; HILL, J. Causal influence of gamma oscillations on the sensorimotor rhythm. *NeuroImage*, v. 56, p. 837–842, 2011. 45
- GUYON, I.; WESTON, J.; BARNHILL, S.; VAPNIK, V. Gene selection for cancer classification using support vector machines. *Machine learning*, Springer, v. 46, n. 1-3, p. 389–422, 2002. 144

- HERMAN, P.; PRASAD, G.; MCGINNITY, T. M.; COYLE, D. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*, v. 16, n. 4, p. 317, 2008. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4536580>>. 65
- HIGHAM, D. J. Condition numbers and their condition numbers. *Linear Algebra and its Applications*, v. 214, p. 193 – 213, 1995. ISSN 0024-3795. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0024379593000669>>. 114
- HUTTER, F.; HOOS, H. H.; LEYTON-BROWN, K. Sequential model-based optimization for general algorithm configuration. In: SPRINGER. *International Conference on Learning and Intelligent Optimization*. [S.l.], 2011. p. 507–523. 146
- IFEACHOR, E. C.; JERVIS, B. W. *Digital Signal Processing: A Practical Approach*. 2nd. ed. [S.l.]: Pearson Education, 2002. ISBN 0201596199. 58
- JASPER, H. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, v. 10, p. 371 – 375, 1958. 45, 46
- JUNIOR, G. F. M.; GIRALDI, G. *Aprendizado de Variedades: Aspectos Geométricos e Aplicações*. [S.l.], 2012. 71
- KEINOSUKE, F. *Introduction to Statistical Pattern Recognition (2Nd Ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990. ISBN 0-12-269851-7. 59, 60, 68, 69, 143
- KRUSIENSKI, D. J.; GROSSE-WENTRUP, M.; GALÁN, F.; COYLE, D.; MILLER, K. J.; FORNEY, E.; ANDERSON, C. W. Critical issues in state-of-the-art brain–computer interface signal processing. *Journal of neural engineering*, v. 8, n. 2, p. 025002–025002, 04 2011. 28
- LAFLEUR, K.; CASSADY, K.; DOUD, A.; SHADES K.AND ROGIN, E.; HE, B. Quadcopter control in three-dimensional space using a noninvasive motor imagery based brain-computer interface. *Journal of Neural Engineering*, v. 10 (4), 2013. 27
- LAL, T. N.; SCHRODER, M.; HINTERBERGER, T.; WESTON, J.; BOGDAN, M.; BIRBAUMER, N.; SCHOLKOPF, B. Support

vector channel selection in BCI. *IEEE transactions on biomedical engineering*, IEEE, v. 51, n. 6, p. 1003–1010, 2004. 145

LEEB, R.; LEE, F.; KEINRATH, C.; SCHERER, R.; BISCHOF, H.; PFURTSCHELLER, G. Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, v. 15, n. 4, p. 473–482, 2007. 54

LEMM, S.; BLANKERTZ, B.; CURIO, G.; MULLER, K.-R. Spatio spectral filters for improving the classification of single trial eeg. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 52, n. 9, p. 1541, 2005. 64

LI, Y.; ANG, K. K.; GUAN, C. Digital signal processing and machine learning. In: GRAIMANN, B.; ALLISON, B.; PFURTSCHELLER, G. (Ed.). *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. [S.l.]: Springer Berlin Heidelberg, 2010. p. 305–330. 11, 50, 51, 69

LOFBERG, J. Yalmip : A toolbox for modeling and optimization in matlab. In: *Proceedings of the CACSD Conference*. Taipei, Taiwan: [s.n.], 2004. Disponível em: <<http://users.isy.liu.se/johanl/yalmip>>. 106, 107, 122, 137

LOTTE, F.; CONGEDO, M.; LECUYER, A.; LAMARCHE, F.; ARNALDI, B. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, v. 4, n. 2, p. R1, 2007. 29, 30

LOTTE, F.; CONGEDO, M.; LÉCUYER, A.; LAMARCHE, F.; ARNALDI, B. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, v. 4, n. 2, p. R1, 2007. Disponível em: <<http://stacks.iop.org/1741-2552/4/i=2/a=R01>>. 12, 50, 67, 68, 70, 122, 143, 147

LOTTE, F.; GUAN, C. Spatially regularized common spatial patterns for eeg classification. In: *Pattern Recognition (ICPR), 2010 20th International Conference on*. 10.1109/ICPR.2010.904: [s.n.], 2010. (1051-4651), p. 3712–3715. 62

LOTTE, F.; GUAN, C. Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 58, n. 2, p. 355, 2011. 29, 64, 66, 139, 140, 142

- LU, Z.; PONG, T. K. Minimizing condition number via convex programming. *SIAM Journal on Matrix Analysis and Applications*, v. 32, n. 4, p. 1193–1211, 2011. 115
- MAKEIG, S.; KOTHE, C.; MULLEN, T.; BIGDELY-SHAMLO, N.; ZHANG, Z.; KREUTZ-DELGADO, K. Evolving signal processing for brain–computer interfaces. *Proceedings of the IEEE*, v. 100, p. 1567–1584, 2012. Disponível em: <[https://scn.ucsd.edu/~scott/pdf/Makeig\\_ProcIEEE12.pdf](https://scn.ucsd.edu/~scott/pdf/Makeig_ProcIEEE12.pdf)>. 46
- MATLAB. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010. 106
- MELLINGER, J.; SCHALK, G.; BRAUN, C.; PREISSEL, H.; ROSENSTIEL, W.; BIRBAUMER, N.; KÜBLER, A. An meg-based brain-computer interface (bci). *NeuroImage*, v. 36, n. 3, p. 581–593, 2007. 48
- MENG, J.; HUANG, G.; ZHANG, D.; ZHU, X. Optimizing spatial spectral patterns jointly with channel configuration for brain–computer interface. *Neurocomputing*, v. 104, p. 115 – 126, 2013. 141
- MILLAN, J. R.; RENKENS, F.; MOURINO, J.; GERSTNER, W. Noninvasive brain-actuated control of a mobile robot by human eeg. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 51, 2004. 27
- MILLER, K. J.; DENNIJS, M.; SHENOY, P.; MILLER, J. W.; RAO, R. P.; OJEMANN, J. G. Real-time functional brain mapping using electrocorticography. *NeuroImage*, v. 37, n. 2, p. 504 – 507, 2007. ISSN 1053-8119. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S105381190700417X>>. 47
- MOAKHER, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, v. 26, n. 3, p. 735–747, 2005. 71, 73, 74, 189
- MOWLA, M. R.; NG, S.-C.; ZILANY, M. S. A.; PARAMESRAN, R. Single-trial evoked potential estimation using iterative principal component analysis. *IEEE SENSORS JOURNAL*, v. 16, n. 18, p. 6955, 2016. 66

NAGAOKA, T.; SAKATANI, K.; AWANO, T.; YOKOSE, N.; HOSHINO, T.; MURATA, Y.; KATAYAMA, Y.; ISHIKAWA, A.; EDA, H. Development of a new rehabilitation system based on a brain-computer interface using near-infrared spectroscopy. In: \_\_\_\_\_. *Oxygen Transport to Tissue XXXI*. Boston, MA: Springer US, 2010. p. 497–503. ISBN 978-1-4419-1241-1. Disponível em: <[http://dx.doi.org/10.1007/978-1-4419-1241-1\\_72](http://dx.doi.org/10.1007/978-1-4419-1241-1_72)>. 48

NORCIA, A. M.; APPELBAUM, L. G.; ALES, J. M.; COTTEREAU, B. R.; ROSSION, B. The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, Journal of Vision, v. 15, n. 6, p. 1–46, 2015. 49

NOVI, Q.; GUAN, C.; DAT, T. H.; XUE, P. Sub-band common spatial pattern (SBCSP) for brain-computer interface. In: IEEE. *2007 3rd International IEEE/EMBS Conference on Neural Engineering*. [S.l.], 2007. p. 204–207. 14, 29, 139, 140, 141, 144, 149, 150

NOVI, Q.; GUAN, C.; DAT, T. H.; XUE, P. Sub-band common spatial pattern (sbcsp) for brain-computer interface. *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, 2007. 64

PALANIAPPAN, R. Brain computer interface design using band powers extracted during mental tasks. In: *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*. 10.1109/CNE.2005.1419622: [s.n.], 2005. p. 321–324. 28

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011. 145

PFURTSCHELLER, G.; NEUPER, C. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, v. 89, p. 1123 – 1134, 2001. 28

PFURTSCHELLER, G.; NEUPER, C.; BRUNNER, C.; SILVA, F. L. da. Beta rebound after different types of motor imagery in man. *Neuroscience Letters*, v. 378, n. 3, p. 156 – 159, 2005. 57

PFURTSCHELLER, G.; SILVA, F. H. Lopes da. Event related eeg or meg synchronization and desynchronization basic principles. *Clinical Neurophysiology*, v. 110, n. 11, p. 1842–1857, October 1999. 28

- RAKOTOMAMONJY, A.; GUIGUE, V. BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. *IEEE transactions on biomedical engineering*, IEEE, v. 55, n. 3, p. 1147–1154, 2008. 145
- RAMOSER, H.; GERKING, J. M.; PFURTSCHELLER, G. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE TRANSACTIONS ON REHABILITATION ENGINEERING*, v. 8, n. 4, p. 441, 2000. 29, 60
- RAO, R.; SCHERER, R. Statistical pattern recognition and machine learning in brain-computer interfaces. In: \_\_\_\_\_. *Statistical Signal Processing for Neuroscience and Neurotechnology*. [S.l.]: Elsevier Inc., 2010. p. 335–367. ISBN 9780123750273. 75, 76, 77
- RIVET, B.; SOULOUMIAC, A.; ATTINA, V.; GIBERT, G. xdawn algorithm to enhance evoked potentials: Application to brain-computer interface. *IEEE transactions on biomedical engineering*, IEEE, v. 56, n. 8, p. 2035–2043, 2009. 11, 50
- RODRÍGUEZ-BERMÚDEZ, G.; GARCÍA-LAENCINA, P. J.; ROCA-GONZÁLEZ, J.; ROCA-DORDA, J. Efficient feature selection and linear discrimination of {EEG} signals. *Neurocomputing*, v. 115, p. 161–165, 2013. 28
- SANEI, S.; CHAMBERS, J. A. Eeg signal processing. In: \_\_\_\_\_. [S.l.]: John Wiley & Sons, Inc., 2007. 11, 38, 40, 41, 42, 43, 44, 45, 46, 47, 58
- SCHOLKOPF, B.; SMOLA, A. *Kernel Methods and Support Vector Machines*. [S.l.], 2003. 32, 181
- SHIH, J. J.; KRUSIENSKI, D. J.; WOLPAW, J. R. Brain-computer interfaces in medicine. *Mayo Clinic Proceedings*, v. 87 (3), p. 268–279, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497935/>>. 27
- SILVA, C.; DUARTE, R.; GOULART, R.; TROFINO, A. Towards a lmi approach to feature extraction improvements and classification by riemann distance. *2016 12th IEEE International Conference on Control and Automation (ICCA)*, p. 990 – 995, 2016. 30, 176
- SILVA, C.; DUARTE, R.; TROFINO, A. Feature extraction improvements using an lmi approach and riemannian geometry tools: An application to bci. *2016 IEEE Conference on Control Applications (CCA)*, p. 966 – 971, 2016. 175

- SILVA, C.; MAFRA, G.; TROFINO, A. Frequency decomposition approach for sub-band common spatial pattern methods for brain-computer interfaces. 2017. 177
- SILVA, E. N. F. L. da. *Eletroencephalography: basic principles, clinical applications, and related fields*. [S.l.]: Lippincott Williams and Wilkins, 2005. 40, 41, 42
- SILVERTHORN, D. U. *Fisiologia Humana: uma abordagem integrada*. [S.l.]: Artmed, 2010. 11, 38, 39
- SITARAM, R.; CARIA, A.; VEIT, R.; GABER, T.; ROTA, G.; KUEBLER, A.; BIRBAUMER, N. fmri brain-computer interface: A tool for neuroscientific research and treatment. *Computational Intelligence and Neuroscience*, p. 25487, 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233807/>>. 48
- SIVALINGAM, R.; BOLEY, D.; MORELLAS, V.; PAPANIKOLOPOULOS, N. Tensor sparse coding for positive definite matrices. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 36, n. 3, p. 592–605, mar. 2014. ISSN 0162-8828. Disponível em: <<http://dx.doi.org/10.1109/TPAMI.2013.143>>. 159
- SNOEK, J.; LAROCHELLE, H.; ADAMS, R. P. Practical bayesian optimization of machine learning algorithms. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 2951–2959. 146
- STURM, J. F. *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*. 1998. 106, 107, 137
- TANGERMANN, M.; MÜLLER, K.-R.; AERTSEN, A.; BIRBAUMER, N.; BRAUN, C.; BRUNNER, C.; LEEB, R.; MEHRING, C.; MILLER, K. J.; MUELLER-PUTZ, G. et al. Review of the BCI competition IV. *Frontiers in neuroscience*, Frontiers, v. 6, p. 55, 2012. 151
- TEPLAN, M. Fundamentals of eeg measurements. In: *In: Measurement Science Review, Volume 2, Section*. [S.l.: s.n.], 2002. p. 2002. 28, 56
- TOH, K. C.; TODD, M.; TütüNCü, R.; TUTUNCU, R. H. Sdpt3 - a matlab software package for semidefinite programming. *Optimization Methods and Software*, v. 11, p. 545–581, 1998. 106, 107, 137

VANDENBERGHE, L.; BOYD, S.; WU, S. Determinant maximization with linear matrix inequality constraints. *SIAM journal on matrix analysis and applications*, v. 19, p. 499–533, April 1998. [158](#)

WOLPAW, J.; WOLPAW, E. W. *Brain–Computer Interfaces: Principles and Practice*. 1. ed. [S.l.]: Oxford Scholarship, 2012. [11](#), [28](#), [46](#), [47](#), [54](#)

WOLPAW, J. R.; BIRBAUMER, N.; MCFARLAND, D. J.; PFURTSCHELLER, G.; VAUGHAN, T. M. Brain computer interfaces for communication and control. *Clinical Neurophysiology*, v. 113, n. 6, p. 767–791, 2015/10/13 2002. [27](#)

YGER, F. A review of kernels on covariance matrices for bci applications. *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, 2013. [12](#), [72](#)

# Apêndices



# APÊNDICE A – Otimização para o Classificador

Seja o produto interno entre duas matrizes  $A, B \in \mathbb{R}^{p \times p}$  definido como segue,

$$\langle A, B \rangle \triangleq \text{tr}\{A^T B\} \quad (\text{A.1})$$

e dizemos que as matrizes  $A$  e  $B$  são ortogonais quando o produto interno entre elas é nulo, isto é,  $\langle A, B \rangle = \text{tr}\{A^T B\} = 0$ .

Suponha a matriz formada pela combinação linear convexa entre as matrizes de covariância médias de cada classe, por exemplo, para as classes  $\mathcal{A}$  e  $\mathcal{B}$ , isto é,  $C_a$  e  $C_b$ , respectivamente.

$$\hat{C}_i = \alpha_i C_a + (1 - \alpha_i) C_b \quad (\text{A.2})$$

Considere agora que  $\hat{C}_i$  é a projeção ortogonal da matriz  $C_i$  na reta que passa entre as matrizes de covariância médias, como é ilustrado na Figura 23 e, observe que a matriz  $\hat{C}_i$  é completamente determinada pela constante  $\alpha_i \in \mathbb{R}$ . Defina a matriz de erro entre  $C_i$  e  $\hat{C}_i$  como  $E_i \triangleq C_i - \hat{C}_i$ .

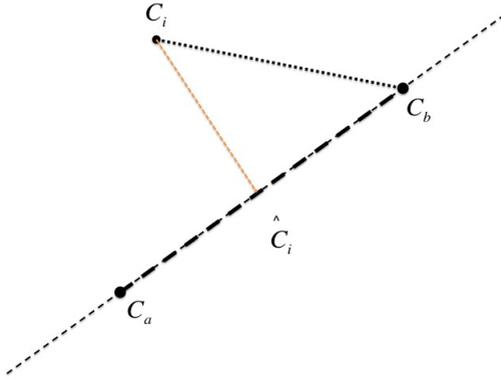
Deseja-se determinar a constante  $\alpha_i$  para cada  $i$ , tal que a matriz erro  $E_i$  e a matriz de erro entre as matrizes médias, isto é,  $C_a - C_b$  apresentem produto interno nulo, ou seja, sejam ortogonais. Então,  $\langle E_i, C_a - C_b \rangle = 0$ , o que resulta em,

$$\alpha_i = \frac{\text{tr}\{(C_i - C_b)^T (C_a - C_b)\}}{\text{tr}\{(C_a - C_b)^T (C_a - C_b)\}} \quad (\text{A.3})$$

Observe na equação A.3 que quando a matriz  $C_i$  se aproxima da matriz  $C_b$ , a constante  $\alpha_i$  se aproxima de 0. Assim como quando a matriz  $C_i$  se aproxima da matriz  $C_a$ , a constante  $\alpha_i$  se aproxima de 1.

Ao introduzirmos a transformação de sinais, como discutida no Capítulo 3, a constante  $\alpha_i$  é uma função de parâmetros estudados

Figura 23: Representação gráfica da projeção ortogonal da matriz  $C_i$  na reta que une as matrizes médias  $C_a$  e  $C_b$ .



nessa tese. Por exemplo,  $\alpha_i(W)$  é função da matriz  $W$  que realiza uma transformação espacial nos sinais. Ou ainda,  $\alpha_i(H_0)$  é função da matriz de parâmetros proposta nessa tese, quando a representação compacta discutida no Capítulo 4 é utilizada.

Para um caso ou outro, a equação A.3 é interpretada como o próprio classificador para os sinais. Nesse sentido, formas de como formular o problema de otimização para determinar, ou a matriz  $W$  ou a matriz  $H_0$  tal que a função  $\alpha(\cdot)$  tenda a 1 para elementos que pertencem a classe  $\mathcal{A}$  e tenda a 0 caso contrário são atividades a serem desenvolvidas como trabalhos futuros.

# APÊNDICE B – Resumos dos Artigos Científicos Produzidos ao longo do trabalho

Como resultados das pesquisas desenvolvidas ao longo do trabalho, três artigos científicos foram redigidos, dos quais dois foram aceitos e apresentados em eventos internacionais e um se encontra submetido para revista internacional. Na sequência apresentamos os resumos desses artigos.

## **Resumo do artigo 01:**

**Aperfeiçoamento de Extração de Características utilizando abordagem LMI e Ferramentas da Geometria de Riemann: uma aplicação a sistemas ICM.** [Silva, Duarte e Trofino \(2016\)](#)

Neste artigo, é proposto uma técnica para melhorar os estágios de extração e classificação de características em Sistemas de Interface Cérebro-Computador (BCI) baseados em sinais de EEG. O problema pode ser formulado como Desigualdades Matriciais Lineares (LMIs) e, portanto, ser resolvido através de ferramentas computacionais robustas. A idéia é representar os sinais EEG usando uma base de sinal senoidal em uma dada faixa de frequência, e introduzindo uma transformação para o sinal, nesta nova representação, parametrizamos a matriz de covariância do sinal transformado. A parametrização é realizada em termos de uma matriz auxiliar relacionada à transformação. O interesse dessas idéias é triplo. Em primeiro lugar, permite uma representação

compacta dos sinais onde as matrizes que representam os sinais têm dimensões consideravelmente menores. A segunda característica interessante é que as informações sobre o espectro de frequência dos sinais aparecem explicitamente na matriz de covariância, o que pode ajudar a extração de características com base em matrizes de covariância. Finalmente, os parâmetros da transformação aparecem linearmente na matriz de covariância e as técnicas padrão de LMI podem ser usadas para determinar a melhor transformação para o problema de classificação, dependendo de como o problema é formulado. A classificação é realizada aproveitando as métricas de Riemann que oferecem uma noção mais conveniente de distância no espaço das matrizes de covariância. Um exemplo numérico baseado em um conjunto de dados de EEG disponibilizados para uma competição de sistemas BCI é usado para ilustrar a metodologia proposta.

### **Resumo do artigo 02:**

**Em direção a uma abordagem LMI para aperfeiçoar a Extração de Características e Classificação por Distância de Riemann** [Silva et al. \(2016\)](#)

Neste trabalho é proposto um método para melhorar a extração de características e taxas de classificação de métodos clássicos baseados em algoritmos padrões espaciais comuns - CSP e algoritmos LDA aplicados a interface cérebro-computador (BCI). A melhora é obtida ajustando os parâmetros de um filtro através da otimização "LMI"(Desigualdade Matriz Linear). Inspiradas nos métodos CSP e LDA, as LMIs são formuladas de forma a reduzir a dispersão dos vetores de características entre os elementos da mesma classe, ao mesmo tempo em que separam seus valores médios de classe (etapa de treinamento). Para abordar os problemas computacionais envolvidos na resolução do problema de otimização, é introduzido uma representação compacta dos sinais EEG em termos de uma matriz de coeficientes e uma base senoidal. São três as principais contribuições nas ideias propostas. Em primeiro lugar, permite uma representação compacta dos

sinais, uma vez que as matrizes que representam os sinais têm dimensões muito menores. Em segundo lugar, as informações sobre o espectro de frequência dos sinais aparecem explicitamente na matriz de covariância, o que pode ajudar a extração de características com base em matrizes de covariância. Finalmente, os parâmetros da transformação aparecem linearmente na matriz de covariância e as técnicas padrão de LMI podem ser usadas para determinar a melhor transformação no problema de classificação, dependendo de como o problema é formulado. Para ilustrar o potencial do método proposto, utilizou-se uma base de dados EEG pública e obtiveram-se taxas de classificação mais elevadas quando comparadas com os resultados clássicos.

### **Resumo do artigo 03:**

**Abordagem por decomposição em frequência para o método Padrões Espaciais Comuns por sub-bandas para interface cérebro-computador** [Silva, Mafra e Trofino \(2017\)](#)

Neste artigo é proposto um método para reduzir o custo computacional do estágio de filtragem de métodos clássicos baseados em Padrões Espaciais Comuns *Sub-band* para Interface Cérebro-Computador sem comprometer sua acurácia geral. A melhoria é obtida representando os sinais EEG em termos de uma matriz de coeficientes e uma base senoidal. Esta representação é mais compacta do que o sinal original. Além disso, as informações de sinal relevantes para discriminação de classes são mantidas. Assim, esta nova representação tornou possível a determinação dos parâmetros para realizar a classificação de uma forma semelhante à que seria aplicada ao sinal bruto. As principais contribuições do trabalho proposto são duas. Em primeiro lugar, permitem uma representação compacta dos sinais, uma vez que as matrizes que representam os sinais têm dimensões muito menores. Em segundo lugar, o custo computacional para ajustar os parâmetros do sistema, bem como aplicá-lo em um cenário de mundo real é consideravelmente reduzido. Para ilustrar o potencial do método proposto, foram utilizados dois conjuntos de dados EEG públicos. O custo computacional foi

reduzido e as taxas de classificação equivalentes foram obtidas quando comparadas aos resultados clássicos.

### **Colaboração com trabalho de Mestrado**

Vale ressaltar que concomitante ao desenvolvimento de trabalho de doutorado houve colaboração nas etapas de idealização, proposição e desenvolvimento do trabalho de mestrado do discente Rafael Mendes Duarte, cujo título é: **Interface Cérebro-Máquina para controle de AR Drone 2.0**. Este trabalho de mestrado tem como objetivo permitir o controle de navegação de um veículo aéreo não tripulado através de comandos gerados a partir dos sinais de EEG do usuário. Este trabalho encontra-se em fase de escrita do documento final e ao longo do desenvolvimento foram realizados diversos ensaios para a coleta de sinais de EEG utilizando um sistema de aquisição de baixo custo. Destaca-se também, que um ambiente para o processamento *online* dos sinais de EEG foi desenvolvido assim como a realização de experimentos em conformidade com os protocolos utilizados em sistemas ICM. Os resultados obtidos ao longo desse trabalho deverão submetidos para publicação em periódicos e/ou revistas científicas em breve.

# Anexos



# ANEXO A – Classificador baseado em Hiperplanos

Com o propósito de deixar o trabalho mais autocontido, apresenta-se a seguir alguns conceitos sobre separação através de hiperplanos e idéias básicas sobre o método SVM extraídas de (SCHOLKOPF; SMOLA, 2003)

Considere  $\mathcal{X}$  como sendo o espaço vetorial dos vetores de característica de entrada, e  $\mathcal{H}$  como sendo o espaço vetorial das características no qual o produto interno é definido e, portanto, podemos utilizá-lo como uma medida de similaridade  $k(x, x')$  entre vetores (o produto interno pode ser utilizado tanto para calcular o ângulo entre dois vetores quanto a norma - ou distância - de um vetor em relação a origem). Assumir também que podemos mapear os vetores de características em  $\mathcal{X}$  para novos vetores em  $\mathcal{H}$  utilizando uma função  $\phi$ , possivelmente não-linear. A dimensão de  $\mathcal{H}$  não é necessariamente igual a de  $\mathcal{X}$ , sendo normalmente maior que esta última.

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

$$x \mapsto x : \phi(x)$$

Este mapeamento apresenta as seguintes vantagens:

1. Permite definir uma medida de similaridade entre vetores de característica utilizando o produto interno, que é definido em  $\mathcal{H}$ , mas não necessariamente em  $\mathcal{X}$ .
2. Permite tratar vetores de característica geometricamente, utilizando álgebra linear e geometria analítica.

3. A liberdade na escolha do mapeamento  $\phi$  nos permite projetar uma variedade de métricas de similaridade e algoritmos de aprendizado.
4. Elementos de classes linearmente inseparáveis em  $\mathcal{X}$  podem apresentar-se linearmente separáveis no espaço  $\mathcal{H}$ .

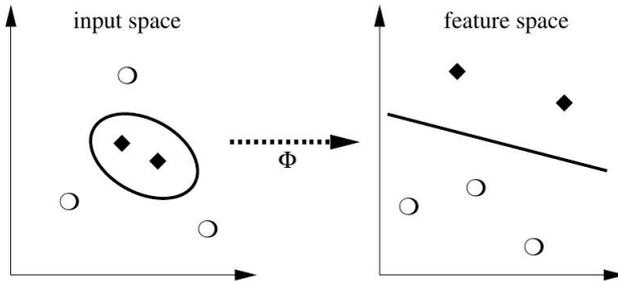


Figura 24: Mapeamento do espaço de entrada  $\mathcal{X}$  para o espaço de *features*  $\mathcal{H}$ . A superfície de decisão correspondente a um hiperplano em  $\mathcal{H}$  não necessariamente é um hiperplano em  $\mathcal{X}$ .

Consideramos que cada padrão  $x_i \in \mathcal{X}$  pertence a uma determinada classe  $y_i \in \mathcal{Y}$ , com  $y_i \in \{-1, +1\}$ . Nosso objetivo é, dado um conjunto de padrões de exemplo  $(x_1, y_1), \dots, (x_{n_e}, y_{n_e})$  já classificados (conjunto de treinamento, ou *training data*), encontrar uma forma de atribuir a classe certa a todos os padrões  $x_i$ , ou seja, queremos que  $y_i f(x_i) > 0, \forall i = 1, \dots, n_e$ , onde  $f(x)$  é a função de decisão que recebe como entrada um padrão  $x_i$  e retorna a classe a que pertence este padrão. Além disso, gostaríamos de fazer com que novos padrões  $x_i$ , nunca antes vistos (conjunto de teste, ou *test data*), fossem também classificados corretamente.

## A.1 Máquina de Vetor Suporte - SVM

Uma abordagem possível é supor que os padrões de entrada, quando mapeados do espaço  $\mathcal{X}$  para o espaço das features  $\mathcal{H}$ , se distribuem neste novo espaço de forma tal que os elementos de diferentes

classes podem ser separados por um hiperplano (uma superfície linear que divide o espaço em dois). Nosso problema se reduz então a encontrar os parâmetros  $w \in \mathcal{H}$  e  $b \in \mathbb{R}$  do hiperplano  $\langle w, x \rangle + b = 0$  que separe as classes adequadamente. Existem infinitas soluções para este problema, então vamos restringir a busca ao hiperplano que, além de satisfazer esta condição, possua a máxima margem de separação medida do hiperplano ao padrão de treino mais próximo. Podemos escrever o problema da seguinte forma:

$$\underset{w \in \mathcal{H}, b \in \mathbb{R}}{\text{maximizar}} \left( \min\{\|x - x_i\| \mid \langle w, x \rangle + b = 0, \forall i = 1, \dots, n_e\} \right) \quad (\text{A.1})$$

$$\text{sujeito a: } y_i(\langle w, x_i \rangle + b) > 0 \quad \forall i = 1, \dots, n_e \quad (\text{A.2})$$

Existe um único hiperplano que satisfaz estas duas condições, sendo portanto chamado de hiperplano ótimo.

Vamos agora multiplicar os parâmetros  $w$  e  $b$  do hiperplano ótimo por um fator tal que os pontos mais próximos do mesmo (pontos situados exatamente na margem) satisfaçam  $|\langle w, x \rangle + b| = 1$ , obtendo assim a chamada forma canônica do hiperplano. Este novo hiperplano faz com todos os pontos/padrões  $x_i$  satisfaçam  $y_i(\langle w, x_i \rangle + b) \geq 1$ . Isso ocorre pois, sendo válida a suposição de que as classes são separáveis, todos estes pontos/padrões estão agora além da margem (definida pela distância do hiperplano ao ponto mais próximo), conforme mostra a Figura 25.

Para obter uma expressão para a margem do hiperplano na forma canônica, consideraremos dois pontos  $x_1$  e  $x_2$  de lados opostos a superfície de decisão e tais que  $\langle w, x_1 \rangle + b = +1$ , e  $\langle w, x_2 \rangle + b = -1$ . Isso significa que  $x_1$  e  $x_2$  estão localizados um de cada lado da superfície definida pelo hiperplano e a uma distância exatamente igual a margem  $m$  deste. Pontos situados a esta distância do hiperplano serão posteriormente chamados de vetores de suporte.

Sejam  $x_1^{(p)}$  e  $x_2^{(p)}$  as projeções de  $x_1$  e  $x_2$  sobre o vetor unitário normal ao plano  $w/\|w\|$ , cujas normas são dadas respectivamente por

$$\|x_1^{(p)}\| = \frac{\langle x_1, w \rangle}{\|w\|} \quad \|x_2^{(p)}\| = \frac{\langle x_2, w \rangle}{\|w\|}$$

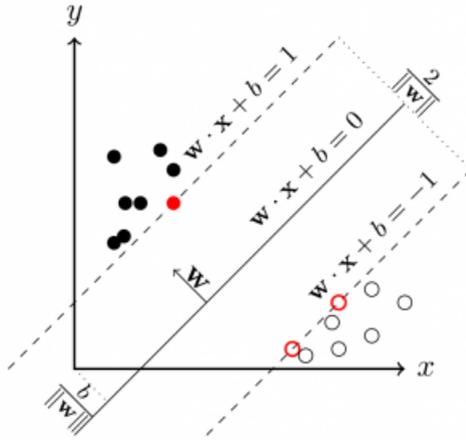


Figura 25: Forma canônica do hiperplano de separação. Destaca-se a margem de separação, linha pontilhada.

Podemos escrever a margem  $m$  como sendo

$$m = \frac{\|x_1^{(p)}\| - \|x_2^{(p)}\|}{2} = \frac{\langle x_1, w \rangle - \langle x_2, w \rangle}{2\|w\|}$$

Usando as equações  $\langle w, x_1 \rangle + b = +1$ , e  $\langle w, x_2 \rangle + b = -1$  e substituindo-as na equação acima, obtemos

$$m = \frac{(1 - b) - (-1 - b)}{2\|w\|}$$

$$m = \frac{1}{\|w\|}$$

Conhecendo a expressão para a margem do hiperplano ótimo na forma canônica, podemos re-escrever o problema da seguinte forma:

$$\underset{w \in \mathcal{H}, b \in \mathbb{R}}{\text{minimizar}} (\tau(w) = \frac{1}{2} \|w\|^2)$$

sujeito a

$$y_i (\langle w, x_i \rangle + b) \geq 1 \quad \forall i = 1, \dots, m$$

A função  $\tau(w)$  é chamada de função objetivo, e a inequação a qual o hiperplano deve estar sujeito é conhecida como de inequação de

restrição. Juntas elas caracterizam um problema de otimização restrita (*constrained optimization*). Cujas soluções se dá através da introdução de multiplicadores de Lagrange  $\alpha_i \geq 0$  e do Lagrangeano  $L(w, b, \alpha)$ , e da minimização deste último. A regra para a construção do Lagrangeano a partir de um problema de otimização e de uma restrição do tipo  $r_i \geq 0$  é multiplicar as equações de restrição por multiplicadores de Lagrange e subtraí-las da função objetivo. Para que esta regra seja aplicável, devemos reescrever a restrição  $y_i(\langle w, x_i \rangle + b) \geq 1$  como  $y_i(\langle w, x_i \rangle + b) - 1 \geq 0$ . Assim, obtemos o Lagrangeano, dado por

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n_e} \alpha_i (y_i(\langle w, x_i \rangle + b) - 1)$$

O Lagrangeano deve ser minimizado em relação a  $w$  e  $b$  (primal variables) e maximizado em relação a  $\alpha_i$  (dual variables). Isso pode ser interpretado como a busca por um ponto de sela. Se uma restrição for violada (algum exemplo for classificado incorretamente), então  $y_i(\langle w, x_i \rangle + b) - 1 < 0$ , e os parâmetros  $w$  e  $b$  deverão variar de forma a fazer com que o valor de  $L$  diminua. Para evitar que  $\alpha_i (y_i(\langle w, x_i \rangle + b) - 1)$  se torne um número negativo arbitrariamente grande, as variações em  $w$  e  $b$  garantirão que a restrição seja eventualmente satisfeita (considerando que o problema é separável). De forma similar, para todas as restrições que não são cumpridas precisamente (para as quais  $(y_i(\langle w, x_i \rangle + b) - 1) > 0$ , o  $\alpha_i$  correspondente deve ser zero. Este é o valor de  $\alpha_i$  que maximiza  $L$ .

Temos agora um problema de otimização convexa (a função objetivo era originalmente convexa, e os pontos que satisfazem a restrição também formam um conjunto convexo, uma vez que qualquer restrição linear define um conjunto convexo, e um conjunto de restrições lineares define a interseção de conjuntos convexos, que é também convexa). Podemos portanto igualar a zero as derivadas de  $L$  em relação a  $w$  e  $b$ , obtendo:

$$w = \sum_i \alpha_i y_i x_i$$

$$\sum_i \alpha_i y_i = 0$$

A solução  $w$  é uma expansão em termos de um subconjunto dos padrões de treino, mais especificamente daqueles com  $\alpha_i$  não nulo, chamados de vetores de suporte. Estes padrões são aqueles que estão localizados sobre a margem. Todos os demais exemplos de treino não precisariam nem ter sido apresentados.

## A.2 Sobreposição entre exemplos de classes diferentes

Na prática, entretanto, as classes podem estar parcialmente sobrepostas mesmo no espaço de *features*, por exemplo em situações em que há muito ruído. Nestes casos, não existe um hiperplano que as separe perfeitamente, e faz-se necessário introduzir novas variáveis  $\xi_i \geq 0 \forall i = 1, \dots, n_e$ , chamadas de variáveis de *slack*. Antes queríamos que o hiperplano ótimo (na forma canônica) fosse capaz de fazer com que  $y_i(\langle w, x_i \rangle + b) \geq 1 \forall i = 1, \dots, n_e$ , ou seja, classificar corretamente todos os padrões  $x_i$  do conjunto de treino. Agora porém, assumindo que isso possa não mais ser possível, nos contentamos em fazer com que  $y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \forall i = 1, \dots, n_e$ , abrindo a restrição imposta anteriormente. As variáveis de *slack* permitem que os pontos estejam situados a uma distância do hiperplano menor do que a margem, permitindo inclusive que estejam do lado correspondente a outra classe. Como consequência deste relaxamento da restrição, e do controle tanto das magnitudes de  $\|w\|$  quanto de  $\sum \xi_i$  obtém-se um classificador com boa generalização para padrões de teste. Uma possível formulação de um classificador como o descrito acima é:

$$\tau(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i$$

sujeito as restrições

$$\xi_i \geq 0$$

e

$$y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \forall i = 1, \dots, n_e$$

A constante  $C$  estabelece um compromisso entre a maximização da margem e a minimização do erro de treino. Da mesma forma que foi feito na seção anterior, podemos escrever este problema utilizando multiplicadores de Lagrange, e resolvê-lo por otimização convexa.



# ANEXO B – Algoritmo para a estimativa da média geométrica em Matrizes Positivas Definidas

A média geométrica entre matrizes simétricas e positivas pode ser estimada, de forma iterativa, pelo algoritmo descrito em [B.1](#).

Para simplificar a notação, considere a seguinte nomenclatura e defina,

$$\text{Log}_{M_g}\{\Omega\} \triangleq M_g^{\frac{1}{2}} \ln(M_g^{-\frac{1}{2}} \Omega M_g^{-\frac{1}{2}}) M_g^{\frac{1}{2}}$$

e

$$\text{Exp}_{M_g}\{S\} \triangleq M_g^{\frac{1}{2}} \exp(M_g^{-\frac{1}{2}} S M_g^{-\frac{1}{2}}) M_g^{\frac{1}{2}}$$

**Algoritmo B.1.** *Algoritmo da Média de Riemann. Adaptado de* ([MOAKHER, 2005](#))

Entradas:  $n_r$  matrizes SPD  $\Omega_i \in \mathcal{P}^n$ , um limiar  $\epsilon > 0$  e o número de iterações  $n_t$

Saída: Média estimada  $M_g$

1: Inicializa  $M_g \leftarrow \frac{1}{n_r} \sum_{i=1}^{n_r} \Omega_i$  e  $t = 1$

2: Repete

3:  $S \leftarrow \frac{1}{n_r} \sum_{i=1}^{n_r} \text{Log}_{M_g^{(t)}}\{\Omega_i\}$

4:  $M_g^{(t+1)} \leftarrow \text{Exp}_{M_g^{(t)}}\{S\}$

5:  $t \leftarrow t + 1$

6: Até  $\|S\|_F < \epsilon$  ou  $t \geq n_t$