
Ondaletas e modelagem estatística aplicados na redução de ruído em sinais

Regis Nunes Vargas



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Uberlândia, 10 de outubro de 2018.

Regis Nunes Vargas¹

**Ondaletas e modelagem estatística
aplicados na redução
de ruído em sinais**

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Processamento Digital de Sinais.

Orientador:

Dr. Antônio Cláudio Paschoarelli Veiga

Banca Examinadora:

Dr. Gilberto Arantes Carrijo

Dr. Lorenzo Santos Vasconcelos

Dr. Luciano Xavier Medeiros

Dr. Marcio Valk

Dr. Tiago Moreira Vargas

Uberlândia, 10 de outubro de 2018.

¹ Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

V297o
2018 Vargas, Regis Nunes, 1985-
 Ondaletas e modelagem estatística aplicados na redução de ruído em
 sinais [recurso eletrônico] / Regis Nunes Vargas. - 2018.

Orientador: Antônio Cláudio Paschoarelli Veiga.

Tese (Doutorado) - Universidade Federal de Uberlândia, Programa
de Pós-Graduação em Engenharia Elétrica.

Modo de acesso: Internet.

Disponível em: <http://dx.doi.org/10.14393/ufu.te.2018.812>

Inclui bibliografia.

Inclui ilustrações.

1. Engenharia elétrica. 2. Ruído eletrônico. 3. Eletrocardiografia -
Controle de ruído. 4. Processamento de sinais. I. Veiga, Antônio Cláudio
Paschoarelli, 1963- (Orient.) II. Universidade Federal de Uberlândia.
Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 621.3

Maria Salete de Freitas Pinheiro - CRB6/1262

Agradecimentos

Gostaria de agradecer a Deus pela oportunidade de vivenciar este momento. À minha esposa Raquel Romes Linhares pelo apoio incondicional durante esta caminhada. Ao meu orientador Antônio Cláudio Paschoarelli Veiga pela atenção e tempo dedicados a este trabalho ao longo destes quatro anos. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro. Aos membros da banca pela dedicação ao analisar este trabalho e por enriquecê-lo com suas pertinentes observações.

Um agradecimento especial a minha mãe Sirlei e às minhas tias Lia e Iara por terem me ajudado e me incentivado desde a graduação. Ao meu pai pelo exemplo de humildade e respeito ao próximo.

Dedico este trabalho à minha esposa Raquel e aos meus filhos Rafael e Roger que me ensinaram o significado do amor incondicional.

Para concluir, gostaria de agradecer a todos que contribuíram direta ou indiretamente para realização deste trabalho.

Resumo

Neste trabalho são propostos novos métodos para redução de ruído em sinais. Estes métodos são baseados na transformada de ondaletas discreta e na modelagem estatística dos coeficientes obtidos a partir da aplicação desta transformada. Serão realizadas simulações de Monte Carlo considerando os sinais de testes de Donoho & Johnstone, sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH e traços sísmicos obtidos do pacote R *astsa*. As simulações serão realizadas considerando diferentes relações sinal-ruído e os novos métodos propostos serão comparados com outros métodos já bastante conhecidos na literatura.

Palavras-chave: Modelagem Estatística. Modelos Ocultos de Markov. Ondaletas. Redução de ruído.

Abstract

In this work new methods are proposed to reduce noise in signals. These methods are based on the discrete wavelet transform and on the statistical modeling of the coefficients obtained from the application of this transform. Monte Carlo simulations will be performed considering the Donoho & Johnstone test signals, electrocardiogram signals obtained from the MIT-BIH database and seismic traces obtained from the R *astsa* package. The simulations will be performed considering different signal-to-noise ratios and the proposed new methods will be compared with other methods already well known in the literature.

Keywords: Statistical Modeling. Hidden Markov Models. Wavelets. Noise Reduction.

Lista de ilustrações

Figura 1 – Representação gráfica da transformada de ondaletas discreta	18
Figura 2 – Exemplo de função soft-thresholding considerando um limiar de 3. . . .	21
Figura 3 – Exemplo de função hard-thresholding considerando um limiar de 3. . .	22
Figura 4 – Exemplo para função \mathbf{h} considerando um limiar de 3.	22
Figura 5 – Representação gráfica do primeiro, segundo e terceiro passos do método RRVO.	43
Figura 6 – Representação gráfica do quarto e quinto passos do método RRVO. . .	43
Figura 7 – Fluxograma do método RRSCST	46
Figura 8 – Representação gráfica do método SureShrinkWin	48
Figura 9 – Representação gráfica do método RRDEL.	51
Figura 10 – Nuvem de palavras representado o número de vezes que cada método obteve a melhor performance	61
Figura 11 – Representação gráfica dos resultados da tabela 1 - SNR de saída	62
Figura 12 – Representação gráfica dos resultados da tabela 1 - PRD	63
Figura 13 – Representação gráfica dos resultados da tabela 2 - SNR de saída	64
Figura 14 – Representação gráfica dos resultados da tabela 2 - PRD	65
Figura 15 – Representação gráfica dos resultados da tabela 3 - SNR de saída	66
Figura 16 – Representação gráfica dos resultados da tabela 3 - PRD	67
Figura 17 – Nuvem de palavras representado o número de vezes que cada método obteve a melhor performance	69
Figura 18 – Representação gráfica dos resultados da tabela 4 - SNR de saída	73
Figura 19 – Representação gráfica dos resultados da tabela 4 - PRD	74
Figura 20 – Representação gráfica dos resultados da tabela 5 - SNR de saída	75
Figura 21 – Representação gráfica dos resultados da tabela 5 - PRD	76
Figura 22 – Representação gráfica dos resultados da tabela 6 - SNR de saída	77
Figura 23 – Representação gráfica dos resultados da tabela 6 - PRD	78

Lista de tabelas

Tabela 1	– Resultados da redução de ruído Gaussiano nos sinais de teste de Donoho & Johnstone, as melhores performances estão apresentadas em negrito.	58
Tabela 2	– Resultados da redução de ruído Gaussiano nos sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH, as melhores performances estão apresentadas em negrito.	59
Tabela 3	– Resultados da redução de ruído Gaussiano nos traços sísmicos obtidos a partir do pacote R <i>astsa</i> , as melhores performances estão apresentadas em negrito.	60
Tabela 4	– Resultados da redução de ruído Rayleigh nos sinais de teste de Donoho & Johnstone, as melhores performances estão apresentadas em negrito.	70
Tabela 5	– Resultados da redução de ruído Rayleigh nos sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH, as melhores performances estão apresentadas em negrito,	71
Tabela 6	– Resultados da redução de ruído Rayleigh nos traços sísmicos obtidos a partir do pacote R <i>astsa</i> , as melhores performances estão apresentadas em negrito,	72

Lista de abreviaturas e siglas

DMA	mediana dos desvios absolutos da mediana
RRSCST	redução de ruído em sinal por clusterização e soft-thresholding
HTLU	hard-thresholding por meio do limiar universal
ITOD	inversa da transformada de ondaletas discreta
ME	maximização de expectativas
MOM	modelos ocultos de Markov
NLOA	nova limiarização de ondaletas aperfeiçoada
RRDEL	redução de ruído por meio de dupla estimativa de limiares
RAGB	ruído aditivo gaussiano branco
RAR	ruído aditivo de Rayleigh
RRVO	redução de ruído por meio do algoritmo de Viterbi e ondaletas
SURE	método SureShrink
SureShrinkWin	método SureShrink suavizado
RRVT	redução de ruído por variação total
TOD	transformada de ondaletas discreta
FVC	função de verossimilhança completa

Sumário

Introdução	12
0.1 Relevância e justificativa	12
0.2 Estrutura do trabalho	13
1 TRANSFORMADA DE ONDALETAS DISCRETA	16
1.1 Encolhimento de ondaletas	19
1.1.1 Método SureShrink	22
1.2 Conclusões	24
2 MODELOS OCULTOS DE MARKOV	25
2.1 Distribuição univariada	27
2.2 Função de verossimilhança	28
2.3 Probabilidades do tipo “forward” e “backward”	29
2.4 O algoritmo de maximização de expectativas	31
2.5 Algoritmo de Viterbi	37
2.6 Conclusões	40
3 REDUÇÃO DE RUÍDO POR MEIO DO ALGORITMO DE VITERBI E ONDALETAS	41
3.1 Conclusões	44
4 REDUÇÃO DE RUÍDO EM SINAL POR CLUSTERIZAÇÃO E SOFT-THRESHOLDING	45
4.1 Conclusões	46
5 REDUÇÃO DE RUÍDO POR MEIO DO MÉTODO SURESH- RINK SUAVIZADO	47
5.1 Conclusões	49

6	REDUÇÃO DE RUÍDO POR MEIO DE DUPLA ESTIMA- TIVA DE LIMIARES	50
6.1	Conclusões	54
7	APLICAÇÕES PARA SINAIS CORROMPIDOS POR RUÍDO ADITIVO GAUSSIANO BRANCO	55
7.1	Conclusões	61
8	APLICAÇÕES PARA SINAIS CORROMPIDOS POR RUÍDO ADITIVO DE RAYLEIGH	68
8.1	Conclusões	69
9	CONCLUSÕES	79
	REFERÊNCIAS	81

Introdução

Este capítulo apresenta a relevância, justificativa e a estrutura deste trabalho. Começaremos apresentando o conceito de sinal. Sinais são funções de variáveis aleatórias independentes que contém informações sobre o comportamento ou natureza de um fenômeno. É possível utilizar sistemas, que são entidades que respondem à aplicação de sinais, para extrair suas informações e traduzi-las em atributos ou artefatos. A interpretação destes atributos ou artefatos permite delinear conclusões a respeito de uma determinada análise ou diagnóstico.

0.1 Relevância e justificativa

Grande parte dos sinais são obtidos na presença de ruído e tentar interpretar estes sinais ruidosos pode levar a conclusões equivocadas. Sendo assim, a proposição de novos métodos de redução de ruído em sinais tem sido uma preocupação constante em artigos recentes (HAN; XU, 2016). A forma tradicional de se reduzir ruídos em sinais consiste na utilização de filtros do tipo FIR e IIR. Porém, em decorrência do grande desafio representado pela redução de ruído de sinais não estacionários, uma série de outros métodos têm sido propostos.

Um dos primeiros métodos a serem propostos foi o *SureShrink* (SURE) (DONOHO; JOHNSTONE, 1995), desde então este método tem sido aplicado em diversas áreas. Por exemplo, a combinação do *SureShrink* com um algoritmo de reconstrução (KOLACZYK, 1996) mostra que imagens tomográficas podem ser obtidas a partir da sobreposição de duas imagens, a primeira resultante da transformada inversa sobre os coeficientes da escala mais grossa e a segunda resultante da transformada inversa sobre os coeficientes da escala mais fina após estes terem sido submetidos ao método *SureShrink*.

Um outro exemplo, que mostra a superioridade do método *SureShrink* sobre métodos baseados na transformada de Fourier no que tange a redução de ruídos em espectros astronômicos é dado em (FLIGGE; SOLANKI, 1997). No ano seguinte, (YOUNG; SOHER; MAUDSLEY, 1998) utiliza o método *SureShrink* na identificação de linha de base em si-

nais de ressonância magnética. Já (RAMSEY, 1999) indica aplicações na análise de dados econômicos e financeiros.

Em (FENG et al., 2000) é apresentada uma implementação do método SureShrink por meio de um algoritmo iterativo que permite a redução de ruído em tempo real, (LOTRIČ; DOBNIKAR, 2001) combina redes neurais do tipo MLP com o método SureShrink para estabelecer um modelo preditivo de séries temporais.

Já (MOYO; BROWNJOHN, 2002) apresenta uma aplicação na identificação de eventos e mudanças no estado estrutural de pontes durante e após sua construção. A aplicação do SureShrink na redução de ruído em sinais de radar pode ser verificada em (WEI; WU, 2003). Em (BENAZZA-BENYAHIA; PESQUET, 2004) a ideia do estimador não viciado de Stein é estendida para a redução de ruído em imagens de satélite.

A aplicação do método SureShrink para redução de ruído em sinais de eletrocardiograma (ECGs) pode ser conferida em (KOZAKEVICIUS et al., 2005), neste artigo, além da redução de ruído por meio do método SureShrink também é apresentado um algoritmo para a detecção de curvas do tipo QRS.

Outro método de redução de ruído bastante popular é o *hard-thresholding por meio do limiar universal* (HTLU) (PERCIVAL; WALDEN, 2006), como aplicação no campo da economia este método foi utilizado para propor um estimador de inflação subjacente (ANDERSSON, 2008). Além disso, este método também já foi aplicado na redução de ruído em sinais de ECG (SADASIVAN; MUKHERJEE; SEELAMANTULA, 2014) e redução de ruído em imagens (SINGH; RANADE; SINGH, 2017).

Pode-se destacar também a *redução de ruído por variação total* (RRVT) (CONDAT, 2013). A qual têm sido aplicada em traços sísmicos (LIU et al., 2016), redução de ruído em ECG (NING; SELESNICK, 2013) e na recuperação de imagens nítidas a partir de imagens desfocadas (PERRONE; FAVARO, 2016). Merece destaque também a *nova limiarização de ondaletas aperfeiçoada* (NLOA), a qual é proposta e aplicada na redução de ruídos de ECG em (HAN; XU, 2016).

0.2 Estrutura do trabalho

Na Seção anterior vimos a diversidade de métodos de redução de ruído bem como as diversas aplicações decorrentes de cada um, ainda assim é perceptível a escassez de métodos na literatura que:

- proponham um processamento estatístico para os coeficientes;
- proponham a clusterização de coeficientes em grupos representativos de informação de ruído e informação de sinal;
- proponham a aplicação de janelamento durante a redução de ruído;

□ proponham a dupla estimativa de limiares.

Estes são os itens norteadores desta pesquisa e é com base neles que são propostos novos métodos de redução de ruído. A base teórica necessária para o entendimento das características inerentes a estes métodos consiste na transformada de ondaletas discreta, apresentada no capítulo 1, e nos modelos ocultos de Markov, apresentado no capítulo 2.

A transformada de ondaletas discreta mapeia dados que estão no domínio do tempo para o domínio das ondaletas. A partir de sua aplicação é possível encontrar coeficientes e localizar cada um deles em um intervalo de tempo e frequência. Analisando a magnitude destes coeficientes é possível determinar o quão significativo este coeficiente é para determinação dos valores do sinal. É esta característica que permite a transformada de ondaletas discreta ter ampla aplicação na redução de ruídos em sinais.

Já um modelo oculto de Markov consiste na modelagem de dados observacionais, associando-os com determinadas distribuições que, por sua vez, estão relacionadas a estados gerados a partir de uma cadeia de Markov oculta ao observador. No caso deste trabalho, utilizaremos uma cadeia de Markov a dois estados, será mostrado que por meio do algoritmo de maximização de expectativas (algoritmo ME) e do algoritmo de Viterbi é possível aplicar os modelos ocultos de Markov aos coeficientes de ondaletas de detalhe e clusterizá-los em dois grupos distintos (um representativo de sinal e outro de ruído) utilizando a sequência mais provável de estados ocultos. O processo decisório que permite saber qual grupo é representativo de ruído e qual grupo é representativo de sinal será feito tomando como base o erro quadrático médio ou *mean squared error* (MSE).

O método de redução de ruído por meio do algoritmo de Viterbi e ondaletas (RRVO), apresentado no Capítulo 3, utiliza o algoritmo de Viterbi para decidir qual coeficiente de ondaletas é representante de ruído e qual é representante de sinal. O Capítulo 4, apresenta o método denominado por Redução de ruído em sinal por clusterização e soft-thresholding (RRSCST) que, além de fazer o agrupamento anterior, aplica uma técnica de limiarização no conjunto representate de ruído.

O Capítulo 5 apresenta suavizado (SureShrinkWin), o qual utiliza o método SureShrink considerando uma técnica de janelamento. A redução de ruído por meio de dupla estimativa de limiares (RRDEL), apresentada no Capítulo 6, propõe uma dupla estimativa de limiares e uma nova função de limiarização.

Para verificar a performance dos métodos propostos serão realizadas nos capítulos 7 e 8 simulações de Monte Carlo (MOONEY, 1997). A simulação de Monte Carlo baseia-se na sucessiva aplicação dos métodos em diferentes sinais ruidosos de maneira iterativa. Os parâmetros de performance obtidos em cada iteração serão separados por método, sinal e relação sinal-ruído e terão sua média calculada. Quanto maior o número de iterações mais a média amostral se aproxima do parâmetro populacional. Neste trabalho serão consideradas 100 iterações para cada grupo descrito acima.

As simulações serão realizadas considerando sinais ruidosos com diferentes relações sinal-ruído e os novos métodos propostos serão comparados com os métodos supracitados, os quais já são bastante conhecidos na literatura. Os grupos de sinais escolhidos para a realização das simulações são: os sinais de testes de Donoho & Johnstone, os sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH e traços sísmicos obtidos do pacote R *astsa*. Todos os sinais possuem um tamanho de 1024, este tamanho é aceito e utilizado na maior parte das simulações presentes na literatura (REMÉNYI; VIDAKOVIC, 2015).

Os sinais de teste de Donoho & Johnstone possuem características como descontinuidades e variações de frequência que os tornam amplamente utilizados na avaliação da performance de métodos de redução de ruído, esta larga utilização na literatura é motivação suficiente para sua escolha na avaliação de novos métodos. Os quatro sinais de teste de Donoho & Johnstone são denominados por: Blocks, Bumps, Doppler e Heavisine.

O sinal Blocks é uma sequência de funções escada variando em largura e altura. O sinal Bumps se caracteriza por ser um sinal com picos repentinos. O sinal Doppler é um sinal que possui variação gradativa de frequência ao longo do tempo. O sinal Heavisine é uma senoide com dois saltos.

Adicionalmente, os erros de diagnóstico (no caso dos sinais de eletrocardiograma) ou de análises (no caso de traços sísmicos) obtidos a partir da interpretação de artefatos ou atributos provenientes de sinais ruidosos faz com que a redução de ruído em ECG e em traços sísmicos seja uma preocupação recorrente na literatura associada a este tema. Por isso a escolha dos sinais provenientes da base de dados MIT-BIH e do pacote R *astsa*.

Os sinais de ECG obtidos a partir da base MIT-BIH estão subcategorizados como *normal sinus rhythm* e sua escolha foi baseada no fato destes sinais serem amplamente utilizados na literatura (CHANDRAKAR; KOWAR, 2012). Já os traços sísmicos obtidos do pacote *astsa* dividem-se em sinais provenientes de terremotos “EQ” e sinais provenientes de explosão “EX”, estes sinais também são amplamente aceitos na comunidade científica (SHUMWAY; STOFFER, 2010).

A partir da análise dos dados obtidos nos capítulos 7 e 8 é possível delinear conclusões a respeito de cada um dos métodos apresentados. Estas conclusões podem ser conferidas no capítulo 9.

Transformada de ondaletas discreta

Neste capítulo será apresentada a teoria básica de ondaletas, sem entrar em detalhes profundos a respeito dos algoritmos matemáticos e numéricos. O fato básico a respeito de ondaletas é sua localização temporal e a ideia da análise de ondaletas consiste em aproximar sinais por uma combinação linear de ondaletas. Estas ondaletas geram uma família de funções que podem reconstruir o sinal. A ondaleta mãe $\psi(\cdot)$ é uma função real $\psi : \mathbb{R} \rightarrow \mathbb{R}$ tal que $\int_{\mathbb{R}} \psi(t) dt = 0$. E a função pai (ou função escala) $\phi(\cdot)$ é uma função real $\phi : \mathbb{R} \rightarrow \mathbb{R}$ tal que $\int_{\mathbb{R}} \phi(t) dt = 1$ ambas satisfazendo a condição de integrabilidade $\psi, \phi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. Tipicamente, exige-se que a ondaleta mãe seja limitada, centrada na origem e com limite igual a zero quando $|t| \rightarrow \infty$. Considerando $j, k \in \mathbb{Z}$ é possível relacionar estas duas funções pelas equações de expansão.

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \quad (1)$$

$$\phi(t) = \sqrt{2} \sum_k g_k \phi(2t - k). \quad (2)$$

A partir da família (ψ, ϕ) , construímos a sequência

$$\begin{aligned} \psi_{j,k}(t) &= 2^{\frac{j}{2}} \psi(2^j t - k), \\ \phi_{j,k}(t) &= 2^{\frac{j}{2}} \phi(2^j t - k). \end{aligned}$$

Os valores g_k e h_k são, respectivamente, filtros passa-baixa e passa-alta satisfazendo $h_k = (-1)^k g_{1-k}$.

A transformada de ondaletas discreta mapeia dados do domínio do tempo para o domínio de ondaletas como é possível ver na Definição 1.

Definição 1 (Transformada de Ondaletas Discreta). *Seja $\mathbf{y} = (y_0, y_1, \dots, y_{N-1})'$ uma amostra aleatória i.i.d., com $N = 2^J$, $J \in \mathbb{N}$. A Transformada de Ondaletas Discreta*

(TOD) de \mathbf{y} , com respeito à ondaleta mãe $\psi(\cdot)$, é definida como

$$d_{j,k} = \sum_{t=0}^{N-1} y_t \psi_{j,k}(t/N), \quad (3)$$

para todo $j = 0, 1, 2, \dots, J-1$ e $k = 0, 1, 2, \dots, 2^j - 1$. Podemos reescrever a transformada (3) na forma de matriz por

$$\mathbf{d} = \mathbf{WY}, \quad (4)$$

supondo-se condições de fronteira apropriada, a transformada é ortogonal e pode ser obtida a Inversa da Transformada de Ondaletas Discreta (ITOD) dada por

$$\mathbf{Y} = \mathbf{W}'\mathbf{d},$$

onde \mathbf{W}' denota a transposta de \mathbf{W} .

O cálculo da TOD não é feito por essa multiplicação matricial, mas por um algoritmo piramidal (MEYER, 1993) de complexidade $O(N)$, o qual consiste da aplicação iterativa e sequencial de filtros passa-alta e passa-baixa.

Dado um sinal ruído \mathbf{y} , este algoritmo retorna dois conjuntos de coeficientes, os coeficientes de detalhe denotados por $\{d_{j,k}\}$ e os coeficientes de aproximação denotados por $\{s_{j,k}\}$, onde $j \in \{0, \dots, J-1\}$ e $k \in \{0, \dots, 2^j - 1\}$, veja as Equações (5), (6), (7) e (8).

$$s_{J-1,k} = \sum_{n=0}^{N-1} g_{2k-n} y_n. \quad (5)$$

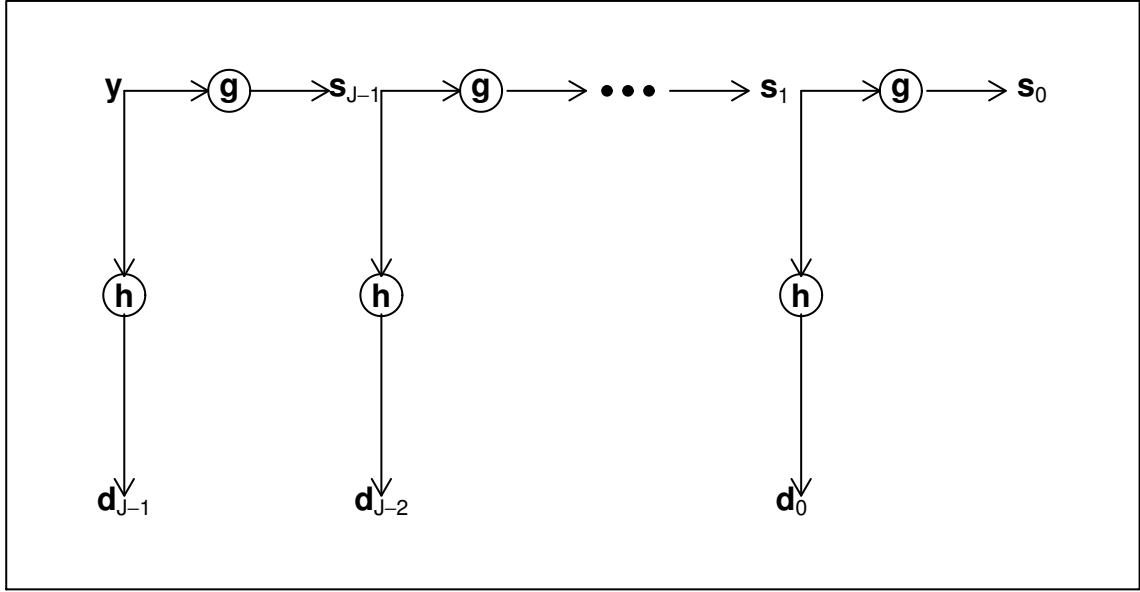
$$s_{j,k} = \sum_{n=0}^{2^j-1} g_{2k-n} s_{j+1,n}, \text{ se } j < J-1. \quad (6)$$

$$d_{J-1,k} = \sum_{n=0}^{N-1} h_{2k-n} y_n. \quad (7)$$

$$d_{j,k} = \sum_{n=0}^{2^j-1} h_{2k-n} s_{j+1,n}, \text{ se } j < J-1. \quad (8)$$

Os coeficientes de ondaletas podem ser agrupados em níveis de acordo com o valor de j , os coeficientes pertencentes ao nível $J-1$, são ditos coeficientes da escala mais fina, enquanto que os coeficientes pertencentes ao nível 0 são ditos os coeficientes da escala mais grossa. Denotando $\mathbf{y} = (y_0, \dots, y_{N-1})$ e os valores representados pelo lado esquerdo de cada uma das equações acima como sendo respectivas coordenadas dos vetores \mathbf{s}_{J-1} , \mathbf{s}_j , \mathbf{d}_{J-1} e \mathbf{d}_j pode-se denotar este processo pela figura 1.

Figura 1 – Representação gráfica da transformada de ondaletas discreta



Um exemplo bastante tradicional é a ondaleta Haar, neste caso a ondaleta mãe é dada pela Equação (9), enquanto que a ondaleta pai é dada pela Equação (10).

$$\psi^{(H)}(t) = \begin{cases} 1, & \text{se } 0 \leq t < \frac{1}{2}, \\ -1, & \text{se } \frac{1}{2} \leq t < 1, \\ 0, & \text{caso contrário.} \end{cases} \quad (9)$$

$$\phi^{(H)}(t) = \begin{cases} 1, & \text{se } 0 \leq t \leq 1, \\ 0, & \text{caso contrário.} \end{cases} \quad (10)$$

É possível obter os filtros da ondaleta Haar utilizando-se a Equação (2).

Primeiramente note que $\phi(t) = \sqrt{(2)}(g_0\phi(2t) + g_1\phi(2t - 1))$.

Substituindo-se t por 0, tem-se que $g_0 = \frac{1}{\sqrt{2}}$.

Substituindo-se t por 1, tem-se que $g_1 = \frac{1}{\sqrt{2}}$.

Lembrando que $h_k = (-1)^k g_{1-k}$, tem-se que $h_0 = g_1 = \frac{1}{\sqrt{2}}$ e que $h_1 = -g_0 = \frac{-1}{\sqrt{2}}$.

Aplicando estes filtros por meio das Equações (5), (6), (7) e (8) em um sinal, como por exemplo, $y = \{1; 2; 3; \dots; 14; 15; 16\}$, obtém-se, considerando a ordem de apresentação

da escala mais fina para a mais grossa, os coeficientes de detalhe

$$\mathbf{d} = \{-0,70; -0,70; -0,70; -0,70; -0,70; -0,70; -0,70; -0,70; -2; -2; -2; -2; -5,65; -5,65; -16\},$$

e os coeficientes de aproximação

$$\mathbf{s} = \{2, 12; 4, 94; 7, 77; 10, 60; 13, 43; 16, 26; 19, 09; 21, 92; 5, 13; 21, 29; 12, 72; 35, 35; 34\}.$$

Veja a seguir como encontrar estes valores utilizando o pacote *wavethresh* no software R (R Core Team, 2016).

```

1  require(wavethresh)
2  y=1:16#carregando sinal y
3  DWT=wd(y,filter.number = 1,family = "DaubExPhase")#aplicando transformada de
   ondaleta Haar no sinal y
4  DWT$filter #mostrando as especificações do filtro
5  $H
6  [1] 0.7071068 0.7071068
7
8  $G
9  NULL
10
11 $name
12 [1] "Haar wavelet"
13
14 $family
15 [1] "DaubExPhase"
16
17 $filter.number
18 [1] 1
19
20 s=DWT$C#carregando os coeficientes suaves
21 d=DWT$D#carregando os coeficientes de detalhe
22 s#apresentando os coeficientes suaves
23      2.121320  4.949747  7.778175 10.606602 13.435029 16.263456 19.091883
24      21.920310  5.00      13.00      21.00      29.00      12.727922 35.355339 34.00
25 d#apresentando os coeficientes de detalhe
26      -0.707106 -0.707106 -0.707106 -0.707106 -0.707106 -0.707106 -0.707106
27      -0.707106 -2.00      -2.00      -2.00      -2.00      -5.6568542 -5.6568542 -16.00

```

Um caso particular de aplicação da TOD é a redução de ruído feita por meio do encolhimento de ondaletas, assunto introduzido pela seção 1.1.

1.1 Encolhimento de ondaletas

De início, é necessário formalizar o conceito de Ruído Aditivo Gaussiano Branco (RAGB) (Definição 2).

Definição 2. *Seja \mathbf{x} um sinal de tamanho N . Se e_0, e_1, \dots, e_{N-1} é uma amostra aleatória independente e igualmente distribuída com distribuição normal $\mathcal{N}(0, \sigma^2)$, então $\mathbf{e} = (e_0, \dots, e_{N-1})$ é um RAGB e $\mathbf{y} = \mathbf{x} + \mathbf{e}$ é um sinal ruído.*

Um outro ruído bastante comum é o Ruído de Rayleigh (veja definições 3 e 4).

Definição 3. *Seja X uma variável aleatória contínua. Diz-se que X segue uma distribuição de Rayleigh se sua função densidade de probabilidade f é dada por*

$$f(x) = \frac{x}{\sigma^2} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\}$$

Sendo assim pode-se definir o ruído aditivo de Rayleigh (RAR) pela Definição 4.

Definição 4. *Seja \mathbf{x} um sinal de tamanho N . Se e_0, e_1, \dots, e_{N-1} é uma amostra aleatória independente e igualmente distribuída com distribuição Rayleigh(σ), então $\mathbf{e} = (e_0, \dots, e_{N-1})$ é um RAR e $\mathbf{y} = \mathbf{x} + \mathbf{e}$ é um sinal ruído.*

O encolhimento de ondaletas usualmente se refere as construções obtidas pela diminuição da magnitude dos coeficientes de ondaletas com o objetivo de reduzir (ou até mesmo remover) o ruído presente no sinal. Os coeficientes de ondaletas $\{d_{j,k}\}$ (Definição 1) são suavizados por funções como, por exemplo, a *soft-thresholding* (Definição 5) e a *hard-thresholding* (Definição 6).

Definição 5. *Seja: \mathbf{x} , \mathbf{y} e \mathbf{e} dados na Definição 2; $\{d_{j,k}\}$ os coeficientes de ondaletas de detalhe obtidos a partir da aplicação da TOD no sinal ruído \mathbf{y} ; $\lambda \in \mathbb{R}$ um limiar.*

Então, a função soft-thresholding é dada por

$$\eta_S(d_{j,k}, \lambda) = \begin{cases} \text{sgn}(d_{j,k})(|d_{j,k}| - \lambda), & \text{se } |d_{j,k}| > \lambda, \\ 0, & \text{caso contrário.} \end{cases} \quad (11)$$

Definição 6. *Seja \mathbf{x} , \mathbf{y} e \mathbf{e} como na Definição 2; $\{d_{j,k}\}$ os coeficientes de ondaletas de detalhe obtidos a partir da aplicação da TOD no sinal ruído \mathbf{y} ; $\lambda \in \mathbb{R}$ um limiar.*

Então, a função hard-thresholding é dada por

$$\eta_H(d_{j,k}, \lambda) = \begin{cases} d_{j,k}, & \text{se } |d_{j,k}| > \lambda, \\ 0, & \text{caso contrário.} \end{cases} \quad (12)$$

Um exemplo de gráfico para a função *soft-thresholding* pode ser conferido na figura 2. Enquanto que a figura 3, mostra um exemplo de gráfico para a função *hard-thresholding*. Existem diversas regras para selecionar o limiar λ . Algumas destas regras são motivadas pela teoria estatística e oferecem uma variedade de maneiras de reduzir os coeficientes. Dentre as diferentes formas de estimar este limiar estão o limiar universal (Definição 7) e o limiar Sureshrink (subseção 1.1.1).

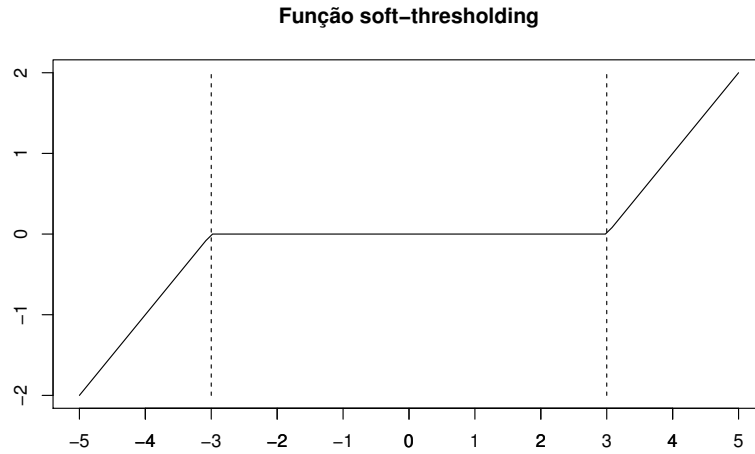
Definição 7. O Limiar Universal λ é definido por

$$\lambda = \hat{\sigma} \sqrt{2 \log N}, \quad (13)$$

onde N é o tamanho do sinal e $\hat{\sigma}$ é a estimativa para o desvio padrão do ruído dada por $\hat{\sigma} = DMA(\{d_{J-1,k} : k = 0, 1, \dots, 2^{J-1} - 1\})$. Aqui, DMA representa o desvio médio absoluto.

Finalmente, a estimativa de \mathbf{x} , denotada por $\hat{\mathbf{x}}$, é obtida pelo método de encolhimento de ondaletas aplicando a ITOD considerando a função de redução de ruído (*soft-thresholding* ou *hard-thresholding*) ao invés de $\{d_{j,k}\}$ (Definição 1) para todos os valores de j e k . No caso do limiar universal utiliza-se a função *hard-thresholding*, de fato, este método é denominado de *Hard-thresholding por meio do Limiar Universal* (HTLU) (PERCIVAL; WALDEN, 2006).

Figura 2 – Exemplo de função *soft-thresholding* considerando um limiar de 3.



Para um melhor entendimento a respeito de como pode ser feita a estimativa do limiar pelo método SureShrink é necessário introduzir o conceito de função fracamente diferenciável (Definição 8), para maiores detalhes consulte (WASSERMANN, 2006).

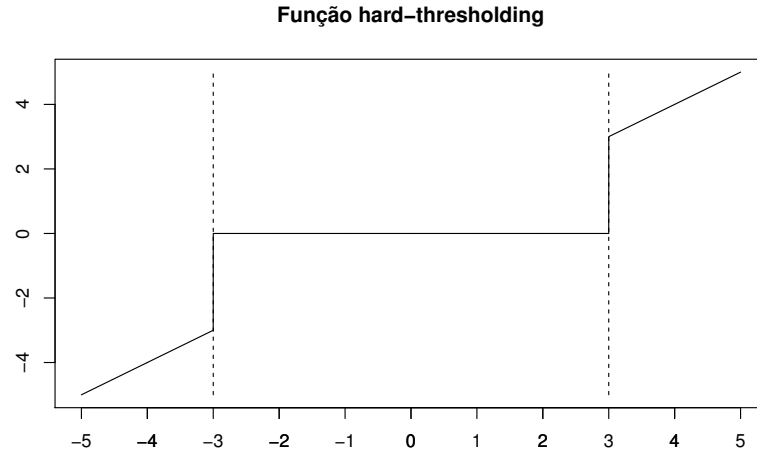
Definição 8. Seja $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}$ uma função integrável em todo intervalo limitado. A função \mathbf{f} é fracamente diferenciável se: existe uma função \mathbf{f}' que é integrável em todo intervalo limitado, tal que

$$\int_x^y \mathbf{f}'(s) ds = \mathbf{f}(y) - \mathbf{f}(x)$$

para todo $x \leq y$. A função \mathbf{f}' é chamada a derivada fraca de \mathbf{f} .

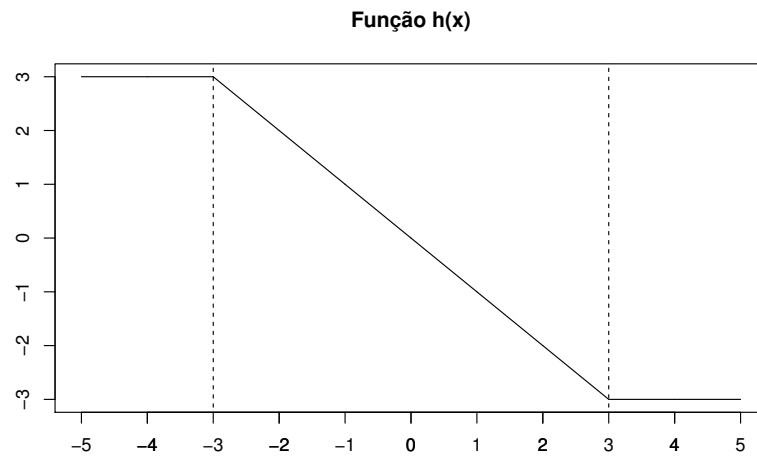
Agora serão introduzidas algumas notações que auxiliarão no entendimento do método SureShrink. Denotando $g(x) = \eta_S(d_{j,k}, \lambda)$, nota-se que $g(x)$ é fracamente diferenciável (veja Figura 2) com derivada fraca g' dada por

Figura 3 – Exemplo de função hard-thresholding considerando um limiar de 3.



$$g'(x) = \begin{cases} 1, & \text{se } |x| > \lambda, \\ 0, & \text{caso contrário.} \end{cases} \quad (14)$$

Defina também a função $h(x) = g(x) - x$, a qual também é fracamente diferenciável. Um exemplo de gráfico para função h é apresentado na Figura 4.

Figura 4 – Exemplo para função **h** considerando um limiar de 3.

1.1.1 Método SureShrink

O método SureShrink é uma forma de estimar o limiar λ baseado no seguinte resultado (STEIN, 1981).

Teorema 1. *Seja:*

1. $\boldsymbol{\mu} \in \mathbb{R}^{N-1}$ um parâmetro a ser estimado;
2. \mathbf{y} uma realização aleatória normal de tamanho $N - 1$ tal que $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $\sigma \in \mathbb{R}$ $i \in \{1, 2, 3, \dots, N - 1\}$;
3. $\mathbf{g} : \mathbb{R}^{N-1} \rightarrow \mathbb{R}^{N-1}$ tal que $\mathbf{g}(\mathbf{y})$ estima $\boldsymbol{\mu} \in \mathbb{R}^{N-1}$;
4. $\mathbf{h} : \mathbb{R}^{N-1} \rightarrow \mathbb{R}^{N-1}$ tal que $\mathbf{h}(\mathbf{y})$ é fracamente diferenciável e $\mathbf{h}(\mathbf{y}) = \mathbf{g}(\mathbf{y}) - \mathbf{y}$.

Então

$$E_{\mu} \|\mathbf{g}(\mathbf{y}) - \boldsymbol{\mu}\|^2 = (N - 1)\sigma^2 + \|\mathbf{h}(\mathbf{y})\|^2 + 2\sigma^2 \sum_{i=1}^{N-1} \frac{\partial h_i}{\partial y_i}. \quad (15)$$

O método SureShrink (DONOHO; JOHNSTONE, 1995) consiste em selecionar um limiar λ que minimiza a Equação (15) considerando $\mathbf{g}(\mathbf{y})$ a função soft-thresholding dada pela Equação (11). Para contextualizar este teorema com o trabalho em questão é necessário entender $\boldsymbol{\mu}$ como os coeficientes ondaletas de detalhe do sinal limpo e \mathbf{y} como os coeficientes ondaletas de detalhe do sinal ruidoso, ambos obtidos a partir da aplicação da TOD em sinais de tamanho $N = 2^J, J \in \mathbb{Z}$. Adicionalmente considera-se σ como dado na Equação (13). A seguir, será detalhado passo a passo como a Equação (15) pode ser colocada em função de λ quando cada componente de $\mathbf{g}(\mathbf{y})$ é substituída pela função soft-thresholding.

Primeiramente, é possível expandir os valores absolutos da função soft-thresholding de modo que cada componente da função $\mathbf{g}(\mathbf{y})$, denotado por g_i , possa ser escrito como:

$$g_i(\mathbf{y}) = \begin{cases} y_i + \lambda, & \text{se } y_i < -\lambda, \\ 0, & \text{se } -\lambda \leq y_i \leq \lambda, \\ y_i - \lambda, & \text{se } y_i > \lambda. \end{cases}$$

Assim, cada componente da função $\mathbf{h}(\mathbf{y})$, denotado por h_i , pode ser escrito como:

$$h_i(\mathbf{y}) = \begin{cases} \lambda, & \text{se } y_i < -\lambda, \\ -y_i, & \text{se } -\lambda \leq y_i \leq \lambda, \\ -\lambda, & \text{se } y_i > \lambda. \end{cases}$$

Assim, $h_i^2 = \lambda^2 \mathbb{I}(|y_i| > \lambda) + y_i^2 \mathbb{I}(|y_i| \leq \lambda) = \min\{\lambda^2, y_i^2\}$. Onde \mathbb{I} é a função indicadora. E, conseqüentemente

$$\|\mathbf{h}(\mathbf{y})\|^2 = \sum_{i=1}^{N-1} \min\{\lambda^2, y_i^2\}. \quad (16)$$

Também é possível notar que

$$\frac{\partial h_i}{\partial y_i} = -\mathbb{I}(|y_i| \leq \lambda). \quad (17)$$

Substituindo os valores dados pelas Equações (13), (16) e (17) na Equação (15), tem-se que:

$$E_\mu \|\mathbf{g} - \boldsymbol{\mu}\|^2 = (N-1)\hat{\sigma}^2 + \sum_{i=1}^{N-1} \min\{\lambda^2, y_i^2\} - 2\hat{\sigma}^2 \sum_{i=1}^{N-1} \mathbb{I}(|y_i| \leq \lambda). \quad (18)$$

Assim, o método SureShrink resume-se a encontrar o valor de λ que minimiza a Equação (18).

1.2 Conclusões

Neste capítulo formalizamos a Transformada de Ondaletas Discreta (TOD) de um sinal. Apresentamos o seu cálculo com respeito a ondaleta mãe, bem como introduzimos o algoritmo piramidal. Como exemplo temos a aplicação da TOD utilizando a base Haar. Ainda vimos como é realizado o processamento tradicional de encolhimento de ondaletas apresentando adicionalmente o método SureShrink e formalizando o conceito de ruído.

Modelos ocultos de Markov

Neste capítulo serão introduzidos o Modelo Oculto de Markov (MOM), o algoritmo de Maximização de Expectativa (ME) e o algoritmo de Viterbi. Os resultados apresentados consideram o caso discreto, mas, se as probabilidades forem interpretadas como densidades, é possível aplicar estes mesmos resultados para o caso contínuo. Para definir o modelo oculto de Markov é necessário compreender o que é uma Cadeia de Markov, então iremos abordar este tópico de forma breve, para maiores detalhes consulte (ZUCCHINI; MACDONALD, 2009).

Existem duas partes envolvidas neste tipo de modelo: a parte não observável satisfazendo a propriedade de Markov, e a parte observável $\{C_t\}_{t \in \mathbb{N}^*}$. Cada valor C_t , no modelo $\{C_t\}_{t \in \mathbb{N}^*}$, depende do respectivo estado Z_t , na cadeia Markov $\{Z_t\}_{t \in \mathbb{N}^*}$ como podemos ver na Definição 9.

Definição 9. *Considerando:*

1. $\{C_t\}_{t \in \mathbb{N}^*}$ um modelo;
2. $\mathbf{C}^{(t)} = (C_1, \dots, C_t)$ e $\mathbf{Z}^{(t)} = (Z_1, \dots, Z_t)$ representando as histórias do tempo 1 até o tempo t .

Diz-se que $\{C_t\}_{t \in \mathbb{N}^*}$ é um modelo oculto de Markov se as seguintes condições forem satisfeitas

$$\mathbb{P}(Z_t = z_t | \mathbf{Z}^{(t-1)} = \mathbf{z}^{(t-1)}) = \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}), \quad (19)$$

$$t = 1, 2, \dots,$$

$$\mathbb{P}(C_t = c_t | \mathbf{C}^{(t-1)} = \mathbf{c}^{(t-1)}, \mathbf{Z}^{(t)} = \mathbf{z}^{(t)}) = \mathbb{P}(C_t = c_t | Z_t = z_t), \quad (20)$$

$$t = 1, 2, \dots,$$

onde $\mathbf{c}^{(t)} = (c_1, \dots, c_t)$ é a sequência de observações para o modelo, e $\mathbf{z}^{(t)} = (z_1, \dots, z_t)$ é a sequência de estados ocultos.

O MOM é utilizado quando existe a necessidade de se estimar processos ocultos ao observador que geram os dados observáveis. No caso deste trabalho os dados observáveis serão os coeficientes de ondaletas de detalhe obtidos a partir da aplicação de um TOD em um sinal ruidoso. Assim, a tentativa será a de identificar se o coeficiente em questão é proveniente de um processo gerador de ruído ou de um processo gerador de sinal.

Se a cadeia de Markov $\{Z_t\}_{t \in \mathbb{N}}$ tem m estados diz-se que $\{C_t\}_{t \in \mathbb{N}}$ é um MOM de m estados e se o estado $Z_t = i$ for conhecido, então é possível associar uma distribuição a C_t . Nesta pesquisa é considerada a distribuição Gaussiana e $i \in \{1, 2\}$, isto é, um MOM Gaussiano de 2 estados. Serão introduzidas agora algumas notações que cobrem tanto o caso de observações discretas quanto o caso de observações contínuas.

Definição 10. No caso de distribuições discretas define-se para $i \in \mathbb{N}^*$

$$p_i(c) = \mathbb{P}(C_t = c | Z_t = i). \quad (21)$$

Onde $p_i(c)$ é a função massa de probabilidade de C_t se a cadeia de Markov está no estado i no tempo t . O caso contínuo é tratado similarmente, neste caso, considera-se p_i como sendo uma função densidade de probabilidade.

Uma vez que o MOM está, por definição, associado a uma cadeia de Markov $\{Z_t\}_{t \in \mathbb{N}}$, torna-se importante abordar alguns conceitos a respeito desta teoria, como a definição de probabilidade de transição dada na Definição 11.

Definição 11. Dada uma cadeia de Markov $\{Z_t\}_{t \in \mathbb{N}}$ define-se a probabilidade de transição p_{ij} por

$$p_{ij} = \mathbb{P}(Z_{t+1} = j | Z_t = i). \quad (22)$$

A partir da Definição 11, é possível construir a matriz de probabilidade de transição (m.p.t.) \mathcal{P} . Neste caso, como estão sendo considerados 2 estados, a m.p.t. é dada por

$$\mathcal{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}.$$

Também é possível introduzir o vetor linha $\mathbf{u}(t)$ dado na Definição 12.

Definição 12. Dada a probabilidade não condicional $\mathbb{P}(Z_t = i)$ define-se $\mathbf{u}(t)$, considerando 2 estados, por

$$\mathbf{u}(t) = \begin{pmatrix} u_1(t) & u_2(t) \end{pmatrix},$$

onde

$$u_i(t) = \mathbb{P}(Z_t = i), i \in \{1, 2\}.$$

A Proposição 1 apresenta uma importante propriedade a respeito do vetor linha $\mathbf{u}(t)$.

Proposição 1. *Seja o vetor linha \mathbf{u} como na Definição 12. Tem-se as seguintes equações*

$$\mathbf{u}(t+1) = \mathbf{u}(t)\mathcal{P}, \quad (23)$$

$$\mathbf{u}(t) = \mathbf{u}(1)\mathcal{P}^{t-1}. \quad (24)$$

Demonstração. Por indução em t . Base de indução, $t = 2$. Usando a Equação (23), tem-se que

$$\mathbf{u}(2) = \mathbf{u}(1)\mathcal{P}^{2-1}.$$

O que mostra que a Equação (24) é válida para $t = 2$. Passo de indução, se $\mathbf{u}(n) = \mathbf{u}(1)\mathcal{P}^{n-1}$, $n \in \mathbb{N}$, então $\mathbf{u}(n+1) = \mathbf{u}(1)\mathcal{P}^n$.

Pela Equação (23) e pela hipótese de indução tem-se que

$$\begin{aligned} \mathbf{u}(n+1) &= \mathbf{u}(n)\mathcal{P}, \text{ pela Equação (23), e} \\ &= \mathbf{u}(1)\mathcal{P}^{n-1}\mathcal{P}, \text{ pela hipótese de indução} \\ &= \mathbf{u}(1)\mathcal{P}^n \end{aligned}$$

□

Com estas definições é possível encontrar a distribuição univariada C_t . O que será feito na próxima seção.

2.1 Distribuição univariada

A variável aleatória C_t tem uma distribuição univariada dada por

$$\begin{aligned} \mathbb{P}(C_t = c) &= \sum_{i=1}^2 \mathbb{P}(C_t = c, Z_t = i) \\ &= \mathbb{P}(C_t = c | Z_t = 1)\mathbb{P}(Z_t = 1) + \mathbb{P}(C_t = c | Z_t = 2)\mathbb{P}(Z_t = 2) \\ &= \sum_{i=1}^2 u_i(t)p_i(c). \end{aligned}$$

É possível escrever esta equação de uma forma mais sucinta por meio da seguinte notação matricial

$$\mathbb{P}(C_t = c) = \mathbf{u}(t)\mathcal{D}(c)\mathbf{1}' \quad (25)$$

onde $\mathcal{D}(c) = \begin{pmatrix} p_1(c) & 0 \\ 0 & p_2(c) \end{pmatrix}$ e $\mathbf{1}' = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. E, pela Equação (24), a Equação (25) pode ser reescrita como

$$\mathbb{P}(C_t = c) = \mathbf{u}(1)\mathcal{P}^{t-1}\mathcal{D}(c)\mathbf{1}' \quad (26)$$

Então para estimar a distribuição de C_t é necessário estimar $\mathbf{u}(1)$, \mathcal{P} e $\mathcal{D}(c)$. Em outras palavras precisa-se estimar a distribuição inicial $\mathbb{P}(C_1 = i)$, a probabilidade de transição p_{ij} e os parâmetros gaussianos (μ_i, σ_i) . Onde, $1 \leq i, j \leq 2$. A próxima seção lançará uma luz a respeito destas estimativas por meio da introdução da função de verossimilhança.

2.2 Função de verossimilhança

Nesta seção é introduzida a função verossimilhança para um MOM gaussiano a dois estados (veja Definição 13).

Definição 13. *Dada uma sequência de observações $\mathbf{c}^{(N)} = (c_1, \dots, c_N)$ gerada por um MOM $(C_t)_{t \in \mathbb{N}}$, a função de verossimilhança é dada por*

$$\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}).$$

Desenvolvimento esta equação, tem-se

$$\begin{aligned} \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \sum_{z_1, \dots, z_N=1}^2 \mathbb{P}(Z_1 = z_1) \mathbb{P}(Z_2 = z_2 | Z_1 = z_1) \\ &\quad \times \mathbb{P}(Z_3 = z_3 | \mathbf{Z}^{(2)} = \mathbf{z}^{(2)}) \times \dots \\ &\quad \times \mathbb{P}(Z_N = z_N | \mathbf{Z}^{(N-1)} = \mathbf{z}^{(N-1)}) \times \dots \\ &\quad \times \mathbb{P}(C_1 = c_1 | \mathbf{Z}^{(N)} = \mathbf{z}^{(N)}) \\ &\quad \times \mathbb{P}(C_2 = c_2 | C_1 = c_1, \mathbf{Z}^{(N)} = \mathbf{z}^{(N)}) \times \dots \\ &\quad \times \mathbb{P}(C_N = c_N | \mathbf{C}^{(N-1)} = \mathbf{c}^{(N-1)}, \mathbf{Z}^{(N)} = \mathbf{z}^{(N)}) \end{aligned}$$

Por meio das Equações (19) e (20) na Definição 9 tem-se que

$$\begin{aligned} \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \sum_{z_1, \dots, z_N=1}^2 \mathbb{P}(Z_1 = z_1) \mathbb{P}(Z_2 = z_2 | Z_1 = z_1) \\ &\quad \times \mathbb{P}(Z_3 = z_3 | Z_2 = z_2) \times \dots \\ &\quad \times \mathbb{P}(Z_N = z_N | Z_{N-1} = z_{N-1}) \times \dots \\ &\quad \times \mathbb{P}(C_1 = c_1 | Z_1 = z_1) \\ &\quad \times \mathbb{P}(C_2 = c_2 | Z_2 = z_2) \\ &\quad \times \mathbb{P}(C_N = c_N | Z_N = z_N). \end{aligned}$$

Ou, equivalentemente,

$$\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) = \sum_{z_1, \dots, z_N=1}^2 u_{z_1}(1) p_{z_1 z_2} \cdots p_{z_{N-1} z_N} p_{z_1}(c_1) \cdots p_{z_N}(c_N).$$

O que é equivalente ao seguinte produto matricial

$$\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) = \mathbf{u}(1) \mathcal{D}(c_1) \mathcal{PD}(c_2) \cdots \mathcal{PD}(c_N) \mathbf{1}'. \quad (27)$$

Com o objetivo de ajustar o cálculo da função de verossimilhança na forma de um algoritmo, define-se o vetor α_t , $1 \leq t \leq N$, $N \in \mathbb{N}$ a seguir.

Definição 14. *Considerando a Equação (27). Define-se o vetor de probabilidades do tipo “forward” como*

$$\alpha_t = \mathbf{u}(1) \mathcal{D}(c_1) \prod_{s=2}^t \mathcal{PD}(c_s), 1 \leq t \leq N, N \in \mathbb{N}. \quad (28)$$

E então, é possível calcular a função de verossimilhança de uma forma recorrente, como mostrado abaixo.

$$\begin{aligned} \alpha_1 &= \mathbf{u}(1) \mathcal{D}(c_1) \\ \alpha_t &= \alpha_{t-1} \mathcal{PD}(c_t), t = 2, 3, \dots, N; \\ \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \alpha_N \mathbf{1}' \end{aligned}$$

Na próxima seção serão aprofundados os conceitos de probabilidade do tipo “forward” e introduzido a probabilidade do tipo “backward”.

2.3 Probabilidades do tipo “forward” e “backward”

Na Definição 14 tem-se, para $1 \leq t \leq N$, $N \in \mathbb{N}$,

$$\alpha_t = \mathbf{u}(1) \mathcal{D}(c_1) \prod_{s=2}^t \mathcal{PD}(c_s), 1 \leq t \leq N, N \in \mathbb{N}.$$

Onde $\mathbf{u}(1)$ denota a distribuição inicial da Cadeia de Markov $\{Z_t\}_{t \in \mathbb{N}}$, veja Definição 9.

É possível demonstrar que $\alpha_t(j)$, $1 \leq j \leq 2$ é uma função densidade de probabilidade, para maior detalhes veja (ZUCCHINI; MACDONALD, 2009). Na verdade,

$$\alpha_t(j) = \mathbb{P}(\mathbf{C}^{(t)} = \mathbf{c}^{(t)}, Z_t = j). \quad (29)$$

É também necessário definir o seguinte vetor de probabilidades.

Definição 15. Considerando a função de verossimilhança como na Equação (27). Para $1 \leq t \leq N - 1$, define-se o vetor de probabilidades do tipo “backward” como

$$\beta_t = \left(\prod_{s=t+1}^N \mathcal{PD}(c_s) \mathbf{1}' \right) \quad (30)$$

Por convenção, $\beta_N = \mathcal{I}$. O valor $\beta_t(j)$, $1 \leq j \leq 2$ é uma função densidade de probabilidade condicional. De fato,

$$\beta_t(j) = \mathbb{P}(\mathbf{C}_{t+1}^{(N)} = \mathbf{c}_{t+1}^{(N)} | Z_t = j) \quad (31)$$

onde $\mathbf{C}_{t+1}^{(N)} = (C_{t+1}, \dots, C_N)$, é a história da série temporal do tempo $t + 1$ ao tempo N . Então,

$$\alpha_t(j)\beta_t(j) = \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}, Z_t = j) \quad (32)$$

Como está sendo considerado um MOM a dois estados, define-se

$$\alpha_t = \begin{pmatrix} \alpha_t(1) & \alpha_t(2) \end{pmatrix}$$

e

$$\beta_t = \begin{pmatrix} \beta_t(1) \\ \beta_t(2) \end{pmatrix}.$$

Então, pode-se notar que

$$\begin{aligned} \alpha_t \beta_t &= \alpha_t(1)\beta_t(1) + \alpha_t(2)\beta_t(2) \\ &= \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}, Z_t = 1) + \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}, Z_t = 2) \\ &= \mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \end{aligned} \quad (33)$$

Isto é, $\alpha_t \beta_t$ é equivalente à função de verossimilhança dada na Definição 13.

As duas propriedades seguintes, demonstradas em (ZUCCHINI; MACDONALD, 2009), são necessárias para a introdução do algoritmo de Maximização de Expectativas (algoritmo ME):

Proposição 2. Para $t = 1, \dots, N$ e $j = 1, 2$ temos que

$$\mathbb{P}(Z_t = j | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) = \frac{\alpha_t(j)\beta_t(j)}{\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)})}. \quad (34)$$

E, para $t = 2, \dots, N$

$$\mathbb{P}(Z_{t-1} = j, Z_t = k | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) = \frac{\alpha_{t-1}(j)p_{jk}p_k(c_t)\beta_t(k)}{\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)})}. \quad (35)$$

Com estes resultados é possível introduzir o algoritmo de maximização de expectativas na próxima seção.

2.4 O algoritmo de maximização de expectativas

O algoritmo de maximização de expectativas, ou o algoritmo ME, é um algoritmo que permite maximizar a função de verossimilhança. Este tipo de abordagem é normalmente usada quando algum dos dados estão ausentes. Como é o caso dos Modelos Ocultos de Markov (MOM).

Como foi apresentado na Definição 9, os MOM possuem uma parte observável $\{C_t\}_{t \in \mathbb{N}}$ e uma parte oculta $\{Z_t\}_{t \in \mathbb{N}}$. Ou seja, considera-se neste caso que $\{Z_t\}_{t \in \mathbb{N}}$ está ausente. Sendo assim, o algoritmo ME é uma forma de estimar todos os parâmetros no problema. Os parâmetros são inicializados de forma aleatória e, de forma iterativa, será executado o passo E (Expectativa) e o passo M (Maximização) até que uma condição de convergência previamente estabelecida esteja satisfeita.

Para começar o entendimento sobre o algoritmo ME considera-se a sequência de observações $\mathbf{c}^{(N)}$, e o MOM $\{C_t\}_{t \in \mathbb{N}^*}$ associado a cadeia de Markov $\{Z_t\}_{t \in \mathbb{N}^*}$ como na Definição 9, 10 e 11. Neste caso, como já se sabe, precisa-se estimar:

1. a distribuição inicial $\mathbf{u}(1)$;
2. a matriz de probabilidades de transição \mathcal{P} ;
3. e, a matriz Diagonal $\mathcal{D}(\mathbf{c})$.

É importante lembrar que está sendo considerado um MOM gaussiano de dois estados. Então, $\mathbf{u}(1)$, \mathcal{P} e $\mathcal{D}(\mathbf{c})$ tem dimensões, respectivamente, dadas por 1×2 , 2×2 e 2×2 . Também é necessário notar que para estimar $\mathcal{D}(\mathbf{c})$ é necessário estimar os parâmetros gaussianos (μ_i, σ_i) , $1 \leq i \leq 2$.

Agora que está claro quais os parâmetros a serem estimados introduziremos o passo E. Primeiramente, introduziremos a nova representação dos estados ocultos z_1, \dots, z_n usando as seguintes variáveis zero um.

$$\mathcal{U}_j(t) = \begin{cases} 1, & \text{se } z_t = j, \\ 0, & \text{caso contrário.} \end{cases}$$

$$\mathcal{V}_{jk}(t) = \begin{cases} 1, & \text{se } z_{t-1} = j \text{ e } z_t = k, \\ 0, & \text{caso contrário.} \end{cases}$$

O objetivo é maximizar a função de verossimilhança completa (FVC)

$$\log \mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)}).$$

Trabalhando a parte interna deste logaritmo tem-se o seguinte:

$$\begin{aligned}
\mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \mathbb{P}(Z_1 = z_1) \mathbb{P}(Z_2 = z_2 | Z_1 = z_1) \times \cdots \\
&\quad \times \mathbb{P}(Z_N = z_N | Z_{N-1} = z_{N-1}) \times \cdots \\
&\quad \times \mathbb{P}(C_1 = c_1 | Z_1 = z_1) \times \cdots \\
&\quad \times \mathbb{P}(C_N = c_N | Z_N = z_N) \\
&= u_{z_1}(1) \prod_{t=2}^N p_{z_{t-1}, z_t} \prod_{t=1}^N p_{z_t}(c_t)
\end{aligned} \tag{36}$$

Sem perda de generalidade, pode-se aplicar o logaritmo em ambos os lados desta equação. E, formalmente, definir o seguinte.

Definição 16. *Considerando um MOM a dois estados, define-se a FVC como a seguir*

$$\begin{aligned}
\log \mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \log u_{z_1}(1) + \left(\sum_{t=2}^N \log p_{z_{t-1}, z_t} \right) \\
&\quad + \left(\sum_{t=1}^N \log p_{z_t}(c_t) \right).
\end{aligned}$$

Ou, equivalentemente

$$\begin{aligned}
\log \mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) &= \sum_{j=1}^2 \mathcal{U}_j(1) \log u_j(1) \\
&\quad + \sum_{j=1}^2 \sum_{k=1}^2 \left(\sum_{t=2}^N \mathcal{V}_{jk}(t) \right) \log p_{jk} \\
&\quad + \sum_{j=1}^2 \sum_{t=1}^N \mathcal{U}_j(t) \log p_j(c_t) \\
&= A + B + C.
\end{aligned} \tag{37}$$

Denota-se $\hat{\mathcal{V}}_{jk}(t)$ e $\hat{\mathcal{U}}_j(t)$ as seguintes esperanças condicionais

$$\begin{aligned}
\hat{\mathcal{V}}_{jk}(t) &= E(\mathcal{V}_{jk}(t) | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&= 1 \times \mathbb{P}(\mathcal{V}_{jk}(t) = 1 | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&\quad + 0 \times \mathbb{P}(\mathcal{V}_{jk}(t) = 0 | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&= \mathbb{P}(Z_{t-1} = j, Z_t = k | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}),
\end{aligned} \tag{38}$$

e

$$\begin{aligned}
\hat{\mathcal{U}}_j(t) &= E(\mathcal{U}_j(t) | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&= 1 \times \mathbb{P}(\mathcal{U}_j(t) = 1 | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&\quad + 0 \times \mathbb{P}(\mathcal{U}_j(t) = 0 | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}) \\
&= \mathbb{P}(Z_t = j | \mathbf{C}^{(N)} = \mathbf{c}^{(N)}).
\end{aligned} \tag{39}$$

Pelas Equações (34) e (35) é possível reescrever estas esperanças como.

$$\hat{\mathcal{V}}_{jk}(t) = \frac{\alpha_{t-1}(j)p_{jk}p_k(c_t)\beta_t(k)}{\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)})}, \quad (40)$$

e

$$\hat{\mathcal{U}}_j(t) = \frac{\alpha_t(j)\beta_t(j)}{\mathbb{P}(\mathbf{C}^{(N)} = \mathbf{c}^{(N)})}. \quad (41)$$

O passo E consiste em trocar os valores $\mathcal{V}_{jk}(t)$ e $\mathcal{U}_j(t)$ na Equação (37) por $\hat{\mathcal{V}}_{jk}(t)$ e $\hat{\mathcal{U}}_j(t)$, respectivamente.

A nova FVC obtida no passo E é maximizada no passo M. Os termos A e B podem ser maximizados por multiplicador de Lagrange, os valores $u_j(1)$ e p_{jk} que maximizam estes dois termos são dados pela Proposição 3.

Proposição 3. *Considerando os termos A e B, na Equação (37), depois da substituição de $\mathcal{U}_j(t)$ e $\mathcal{V}_{jk}(t)$ por $\hat{\mathcal{U}}_j(t)$ e $\hat{\mathcal{V}}_{jk}(t)$. Isto é,*

$$A = \sum_{j=1}^2 \hat{\mathcal{U}}_j(1) \log u_j(1), \quad (42)$$

e

$$B = \sum_{j=1}^2 \sum_{k=1}^2 \left(\sum_{t=2}^N \hat{\mathcal{V}}_{jk}(t) \right) \log p_{jk}. \quad (43)$$

Os valores $u_j(1)$ e p_{jk} que maximizam a Equação (42) e (43) são, respectivamente, dados por

$$u_j(1) = \frac{\hat{\mathcal{U}}_j(1)}{\hat{\mathcal{U}}_1(1) + \hat{\mathcal{U}}_2(1)}, \quad (44)$$

e

$$p_{jk} = \frac{f_{jk}}{f_{j1} + f_{j2}}, \text{ onde } f_{jk} = \sum_{t=2}^N \hat{\mathcal{V}}_{jk}(t). \quad (45)$$

Demonstração. Prova da Equação (44): o objetivo é maximizar a função

$$F(u_1(1), u_2(1)) = \sum_{j=1}^2 \hat{\mathcal{U}}_j(1) \log u_j(1)$$

onde a função de restrição é

$$G(u_1(1), u_2(1)) = u_1(1) + u_2(1) = 1, u_1(1) \geq 0 \text{ e } u_2(1) \geq 0$$

Usando o método dos multiplicadores de Lagrange, pode-se introduzir $\lambda \in \mathbb{R}$ tal que

$$\nabla F = \lambda \nabla G$$

ou seja,

$$\left\langle \frac{\hat{\mathcal{U}}_1(1)}{u_1(1)}; \frac{\hat{\mathcal{U}}_2(1)}{u_2(1)} \right\rangle = \lambda \langle 1; 1 \rangle$$

o que implica

$$\lambda = \frac{\hat{\mathcal{U}}_1(1)}{u_1(1)} = \frac{\hat{\mathcal{U}}_2(1)}{u_2(1)}.$$

Ou, equivalentemente

$$u_1(1) = \frac{\hat{\mathcal{U}}_1(1)u_2(1)}{\hat{\mathcal{U}}_2(1)} = \frac{\hat{\mathcal{U}}_1(1)(1 - u_1(1))}{\hat{\mathcal{U}}_2(1)}.$$

Isto é,

$$u_1(1) = \frac{\hat{\mathcal{U}}_1(1) - \hat{\mathcal{U}}_1(1)u_1(1)}{\hat{\mathcal{U}}_2(1)}.$$

Separando $u_1(1)$ nesta Equação nós temos que

$$u_1(1) = \frac{\hat{\mathcal{U}}_1(1)}{\mathcal{U}_1(1) + \mathcal{U}_2(1)}.$$

Similarmente,

$$u_2(1) = \frac{\hat{\mathcal{U}}_2(1)}{\mathcal{U}_1(1) + \mathcal{U}_2(1)} \square$$

Prova da Equação (45): começa-se fixando o valor de j . Sendo assim, suponha, sem perda de generalidade que $j = 1$. Então, o objetivo é maximizar a seguinte função F

$$\begin{aligned} F(p_{11}, p_{12}) &= \sum_{k=1}^2 \left(\sum_{t=2}^N \mathcal{V}_{1k}(t) \right) \log p_{1k} \\ &= \left(\sum_{t=2}^N \mathcal{V}_{11}(t) \right) \log p_{11} + \left(\sum_{t=2}^N \mathcal{V}_{12}(t) \right) \log p_{12} \end{aligned}$$

Chamando $f_{jk} = \sum_{t=2}^N \mathcal{V}_{jk}$, tem-se que

$$F(p_{11}, p_{12}) = f_{11} \log p_{11} + f_{12} \log p_{12}.$$

A função de restrição G é dada por

$$G(p_{11}, p_{12}) = p_{11} + p_{12} = 1, p_{11} \geq 0, p_{12} \geq 0.$$

Utilizando o método dos multiplicadores de Lagrange chega-se a

$$\left\langle \frac{f_{11}}{p_{11}}, \frac{f_{12}}{p_{12}} \right\rangle = \langle \lambda; \lambda \rangle$$

Ou seja,

$$\frac{f_{11}}{p_{11}} = \frac{f_{12}}{p_{12}}.$$

Equivalentemente,

$$f_{12}p_{11} = p_{12}f_{11} = (1 - p_{11})f_{11}.$$

E então,

$$f_{12}p_{11} + f_{11}p_{11} = f_{11}.$$

Isto é,

$$p_{11}(f_{12} + f_{11}) = f_{11}.$$

O que permite concluir que

$$p_{11} = \frac{f_{11}}{f_{12} + f_{11}}.$$

Similarmente,

$$p_{12} = \frac{f_{12}}{f_{12} + f_{11}}.$$

Para $j = 2$ a demonstração é análoga.

□

A próxima proposição apresenta os valores de μ_j e σ_j^2 que maximizam o termo C na Equação (37).

Proposição 4. *Seja C como na Equação (37) com a substituição de $\mathcal{U}_j(t)$ por $\hat{\mathcal{U}}_j(t)$, isto é,*

$$C = \sum_{j=1}^2 \sum_{t=1}^N \hat{\mathcal{U}}_j(t) \log p_j(c_t). \quad (46)$$

Considerando um MOM gaussiano a dois estados e as definições anteriores, tem-se que os valores μ_j e σ_j^2 que maximizam a Equação (46) são dados por

$$\mu_j = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_j(t) c_t}{\sum_{t=1}^N \hat{\mathcal{U}}_j(t)}, \quad (47)$$

$$\sigma_j^2 = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_j(t)(c_t - \mu_j)^2}{\sum_{t=1}^N \hat{\mathcal{U}}_j(t)}. \quad (48)$$

Demonstração. Começa-se fixando o valor de j , sem perda de generalidade suponha que $j = 1$. Isto é, nós precisamos maximizar a função F dada por

$$\begin{aligned} F(\mu_1, \sigma_1) &= \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \log f_1(c_t) \\ &= \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \log \frac{1}{\sigma_1 \sqrt{2\pi}} e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}} \end{aligned}$$

Derivando a função F em relação à μ_1 , e igualando o resultado a zero tem-se que

$$\frac{\partial F}{\partial \mu_1} = \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \frac{\sigma_1 \sqrt{2\pi}}{e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}}} \frac{1}{\sigma_1 \sqrt{2\pi}} (e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}}) \frac{-2(c-\mu_1)(-1)}{2\sigma_1^2} = 0.$$

Ou, equivalentemente

$$\sum_{t=1}^N \frac{\hat{\mathcal{U}}_1(t)(c - \mu_1)}{\sigma_1^2} = 0.$$

Então,

$$\sum_{t=1}^N \hat{\mathcal{U}}_1(t)(c_t - \mu_1) = 0.$$

E então,

$$\sum_{t=1}^N \hat{\mathcal{U}}_1(t)c_t - \sum_{t=1}^N \hat{\mathcal{U}}_1(t)\mu_1 = 0.$$

Isto nos permite concluir que

$$\mu_1 = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_1(t)c_t}{\sum_{t=1}^N \hat{\mathcal{U}}_1(t)}.$$

Similarmente, nós temos que

$$\mu_2 = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_2(t)c_t}{\sum_{t=1}^N \hat{\mathcal{U}}_2(t)}.$$

Agora, derivando a função F em relação à σ_1 , e igualando o resultado a zero tem-se que

$$\frac{\partial F}{\partial \sigma_1} = \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \frac{\sigma_1 \sqrt{2\pi}}{e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}}} \left[\frac{-1}{\sigma_1^2 \sqrt{2\pi}} e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}} + \frac{1}{\sigma_1 \sqrt{2\pi}} e^{\frac{-(c-\mu_1)^2}{2\sigma_1^2}} \frac{2(c-\mu_1)^2}{2\sigma_1^3} \right] = 0.$$

Isto é,

$$\sum_{t=1}^N \hat{\mathcal{U}}_1(t) \left[\frac{-1}{\sigma_1} + \frac{(c - \mu_1)^2}{\sigma_1^3} \right] = 0.$$

Equivalentemente,

$$-\sum_{t=1}^N \frac{\hat{\mathcal{U}}_1(t)}{\sigma_1} + \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \frac{(c - \mu_1)^2}{\sigma_1^3} = 0.$$

Multiplicando ambos os lados desta equação por σ_1 tem-se que

$$-\sum_{t=1}^N \hat{\mathcal{U}}_1(t) + \sum_{t=1}^N \hat{\mathcal{U}}_1(t) \frac{(c - \mu_1)^2}{\sigma_1^2} = 0.$$

O que permite concluir que

$$\sigma_1^2 = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_1(t) (c - \mu_1)^2}{\sum_{t=1}^N \hat{\mathcal{U}}_1(t)}.$$

Similarmente, tem-se que

$$\sigma_2^2 = \frac{\sum_{t=1}^N \hat{\mathcal{U}}_2(t) (c - \mu_2)^2}{\sum_{t=1}^N \hat{\mathcal{U}}_2(t)}$$

□

Com os novos valores para $u_j(1)$, p_{jk} e (μ_j, σ_j) , $1 \leq j, k \leq 2$, obtidos no passo M é possível encontrar novos valores para $\hat{\mathcal{V}}_{jk}(t)$ e $\hat{\mathcal{U}}_j(t)$ e retornar para o passo E. O passo E e o passo M são iterativamente repetidos até que uma condição de convergência previamente estabelecida seja satisfeita.

Agora que foi concluída a discussão sobre o algoritmo ME, pode-se introduzir o algoritmo de Viterbi, o qual é utilizado para encontrar a sequência mais provável de estados ocultos z_1, \dots, z_N . Isso é feito usando os parâmetros previamente estimados no algoritmo ME de uma forma recorrente, como é possível ver na próxima seção.

2.5 Algoritmo de Viterbi

O propósito desta seção é introduzir a forma pela qual se encontra o máximo da seguinte expressão:

$$\mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)} | \mathbf{C}^{(N)} = \mathbf{c}^{(N)})$$

O algoritmo de Viterbi estima a sequência de estado ocultos mais prováveis. No caso deste trabalho, é como se uma pessoa imaginasse duas caixas. Uma destas caixas gera coeficientes de ondaletas de ruído, digamos a caixa 1. E a outra gera os coeficientes de ondaletas de sinal, digamos a caixa 2. Efetivamente, não é possível saber ainda quais dos estados (1 ou 2) são de fato o gerador de ruído ou sinal. Contudo já é possível imaginar

aqui uma clusterização dos coeficientes de ondaletas em dois grupos. Esta clusterização e a posterior decisão de qual estado gera sinal e qual gera ruído será detalhada no decorrer deste trabalho.

Uma vez que os valores $\mathbf{c}^{(N)}$ são fixos, o que se precisa realmente maximizar é o seguinte.

$$\mathbb{P}(\mathbf{Z}^{(N)} = \mathbf{z}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)}). \quad (49)$$

Esta maximização é feita por um método iterativo, o algoritmo de Viterbi, obtido a partir dos resultados apresentados na Proposição 5.

Proposição 5. *Seja,*

$$\xi_{1i} = \mathbb{P}(Z_1 = i, C_1 = c_1) = u_i(1)p_i(c_1), \quad (50)$$

e

$$\xi_{ti} = \max_{z_1, \dots, z_{t-1}} \mathbb{P}(\mathbf{Z}^{(t-1)} = \mathbf{z}^{(t-1)}, Z_t = i, \mathbf{C}^{(t)} = \mathbf{c}^{(t)}), \quad (51)$$

quando $t = 2, 3, \dots, N$, para todo $1 \leq i \leq 2$. Então,

$$\xi_{tj} = \left(\max_{1 \leq i \leq 2} (\xi_{t-1,i} p_{ij}) \right) p_j(c_t), \quad (52)$$

para todo $1 \leq j \leq 2$.

Demonstração. Para $t = 2$ temos:

$$\xi_{2j} = \max_{1 \leq z_1 \leq 2} \mathbb{P}(Z_1 = z_1, Z_2 = j, C_1 = c_1, C_2 = c_2),$$

substituindo z_1 por i

$$\xi_{2j} = \max_{1 \leq i \leq 2} \mathbb{P}(Z_1 = i, Z_2 = j, C_1 = c_1, C_2 = c_2) = \max_{1 \leq i \leq 2} u_i(1)p_{ij}p_i(c_1)p_j(c_2),$$

equivalentemente

$$\xi_{2j} = \left(\max_{1 \leq i \leq 2} u_i(1)p_i(c_1)p_{ij} \right) p_j(c_2) = \left(\max_{1 \leq i \leq 2} \xi_{1,i}p_{ij} \right) p_j(c_2).$$

Agora denominando $z_{t-1} = i$, onde $i \in \{1, 2\}$ e considerando $t > 2$. Pela Equação (51) tem-se que

$$\begin{aligned}
\xi_{tj} &= \max_{z_1, \dots, z_{t-1}} \mathbb{P}(\mathbf{Z}^{(t-1)} = \mathbf{z}^{(t-1)}, Z_t = j, \mathbf{C}^{(t)} = \mathbf{c}^{(t)}), \\
&\text{equivalentemente} \\
\xi_{tj} &= \max_{z_1, \dots, z_{t-1}} \mathbb{P}(\mathbf{Z}^{(t-2)} = \mathbf{z}^{(t-2)}, Z_{t-1} = i, Z_t = j, \mathbf{C}^{(t)} = \mathbf{c}^{(t)}), \\
&\text{segundo o mesmo raciocínio da Equação (36), tem-se as duas igualdades abaixo} \\
\xi_{tj} &= \left(\max_{z_1, \dots, z_{t-1}} \left(u_{z_1}(1) \prod_{n=2}^{t-1} p_{z_{n-1}, z_n} \prod_{n=1}^{t-1} p_{z_n}(c_n) p_{ij} \right) \right) p_j(c_t) \\
\xi_{tj} &= \left(\max_{z_1, \dots, z_{t-1}} \mathbb{P}(\mathbf{Z}^{(t-2)} = \mathbf{z}^{(t-2)}, Z_{t-1} = i, \mathbf{C}^{(t-1)} = \mathbf{c}^{(t-1)}) p_{ij} \right) p_j(c_t), \\
&\text{equivalentemente} \\
\xi_{tj} &= \left(\max_{1 \leq i \leq 2} \left(\left(\max_{z_1, \dots, z_{t-2}} \mathbb{P}(\mathbf{Z}^{(t-2)} = \mathbf{z}^{(t-2)}, Z_{t-1} = i, \mathbf{C}^{(t-1)} = \mathbf{c}^{(t-1)}) \right) p_{ij} \right) \right) p_j(c_t), \\
&\text{pela Equação (51) tem-se que} \\
\xi_{tj} &= \left(\max_{1 \leq i \leq 2} (\xi_{t-1, i} p_{ij}) \right) p_j(c_t).
\end{aligned}$$

□

A Proposição 5 permite resumir o algoritmo de Viterbi da seguinte forma.

O primeiro passo consiste em calcular os valores $\{\xi_{11}, \xi_{12}, \xi_{21}, \xi_{22}, \dots, \xi_{N1}, \xi_{N2}\}$ usando as Equações (50) e (52). O segundo passo consiste em encontrar o $\max\{\xi_{N1}, \xi_{N2}\}$. Pode-se notar que calcular o valor

$$\begin{aligned}
\max\{\xi_{N1}, \xi_{N2}\} = \max \{ & \max_{z_1, \dots, z_{N-1}} (\mathbb{P}(\mathbf{Z}^{(N-1)} = \mathbf{z}^{(N-1)}, Z_N = 1, \mathbf{C}^{(N)} = \mathbf{c}^{(N)})), \\
& \max_{z_1, \dots, z_{N-1}} (\mathbb{P}(\mathbf{Z}^{(N-1)} = \mathbf{z}^{(N-1)}, Z_N = 2, \mathbf{C}^{(N)} = \mathbf{c}^{(N)})) \},
\end{aligned}$$

é o mesmo que maximizar a Equação (49).

O terceiro passo consiste nas seguintes N iterações:

$$\begin{aligned}
z_N &= \operatorname{argmax}\{\xi_{N1}; \xi_{N2}\}; \\
z_{N-1} &= \operatorname{argmax}\{\xi_{N-1,1} p_{1,z_N}; \xi_{N-1,2} p_{2,z_N}\}; \\
&\vdots \\
z_2 &= \operatorname{argmax}\{\xi_{2,1} p_{1,z_3}; \xi_{2,2} p_{2,z_3}\}; \\
z_1 &= \operatorname{argmax}\{\xi_{1,1} p_{1,z_2}; \xi_{1,2} p_{2,z_2}\}.
\end{aligned}$$

Estas iterações encontram a sequência \mathbf{z}^N que maximiza a Equação (49). Isto é, esta é a sequência mais provável de estados ocultos do MOM.

2.6 Conclusões

Neste capítulo definimos o Modelo Oculto de Markov e mostramos como estimar os parâmetros deste modelo por meio do algoritmo de Maximização de Expectativas. A partir destes parâmetros previamente estimados utiliza-se o algoritmo de Viterbi para encontrar a sequência de estados ocultos mais provável. Agora que as discussões a respeito de MOM, algoritmo ME e algoritmo de Viterbi foram concluídas podemos apresentar as metodologias a partir do próximo capítulo.

Redução de ruído por meio do algoritmo de Viterbi e ondaletas

Neste capítulo é proposto o método de redução de ruído por meio do algoritmo de Viterbi e ondaletas (RRVO), este método baseia-se na sequência de estados ocultos mais prováveis em um Modelo Oculto de Markov. Esta sequência é obtida a partir do algoritmo de Viterbi. Para aplicar o método RRVO deve-se seguir os seguintes passos.

O primeiro passo consiste em: dado um sinal \mathbf{y} de tamanho N obtido a partir da adição de um ruído gaussiano branco \mathbf{e} a um sinal limpo \mathbf{x} , onde $N = 2^J$, $J \in \mathbb{N}$, aplica-se a TOD no sinal \mathbf{y} para obter os coeficientes de detalhe $\mathbf{d} = \{d_t\}_{1 \leq t \leq N-1}$ considerando a ordem de apresentação da escala mais fina para a mais grossa (veja o capítulo 1).

O algoritmo ME é utilizado no segundo passo para obter um MOM gaussiano de dois estados dada a sequência de observações $\mathbf{c}^{(N-1)} = \mathbf{d}$ (veja o capítulo 2), onde \mathbf{d} são os coeficientes de detalhe considerando a ordem de apresentação da escala mais fina para a mais grossa ($c_t = d_t$, $1 \leq t \leq N-1$).

No terceiro passo utiliza-se o algoritmo de Viterbi (veja o capítulo 2) para encontrar a sequência mais provável $\mathbf{z}^{(N)}$. Uma vez que está sendo considerado um MOM de 2 estados, então $z_t \in \{1, 2\}$, $1 \leq t \leq N-1$.

Um exemplo para estes três primeiros passos pode ser conferido na Figura 5. A Figura 5-a representa um sinal limpo \mathbf{x} de tamanho 32. Adicionando um RAGB ao \mathbf{x} , obtém-se o sinal ruído \mathbf{y} , representado na Figura 5-b. Os coeficientes de ondaletas de detalhe são então obtidos a partir da aplicação da TOD no sinal \mathbf{y} . Em seguida, por meio do algoritmo ME, um MOM é ajustado para estes coeficientes. Posteriormente, a partir do algoritmo de Viterbi, estes coeficientes são agrupados de acordo com a sequência de estados ocultos mais prováveis, como mostra a Figura 5-c.

O quarto passo consiste em gerar dois conjuntos $D_1 = \{b_{1i}\}_{1 \leq i \leq N-1}$ e $D_2 = \{b_{2i}\}_{1 \leq i \leq N-1}$

onde

$$b_{1i} = \begin{cases} c_i, & \text{if } z_i = 1, \\ 0, & \text{caso contrário.} \end{cases} \quad (53)$$

e

$$b_{2i} = \begin{cases} c_i, & \text{if } z_i = 2, \\ 0, & \text{caso contrário.} \end{cases} \quad (54)$$

Finalmente, no quinto passo, aplicando a ITOD nos conjuntos D_1 e D_2 , respectivamente, obtém-se $\hat{\mathbf{x}}_1$ e $\hat{\mathbf{x}}_2$. Denota-se por $\hat{\mathbf{x}}$ a redução de ruído do sinal \mathbf{y} a qual é dada por

$$\hat{\mathbf{x}} = \begin{cases} \hat{\mathbf{x}}_1, & \text{se } \text{MSE}(\hat{\mathbf{x}}_1, \mathbf{y}) < \text{MSE}(\hat{\mathbf{x}}_2, \mathbf{y}), \\ \hat{\mathbf{x}}_2, & \text{caso contrário,} \end{cases} \quad (55)$$

onde MSE denota o erro quadrático médio entre dois sinais.

A Figura 6 é uma continuação da Figura 5 e exemplifica os dois últimos passos do método RRVO. Na Figura 6-a tem-se os coeficientes apresentados na Figura 5-c respeitando a Equação (53), isto é, a Figura 6-a representa o conjunto D_1 . Analogamente, a Figura 6-b representa o conjunto D_2 . As Figuras 6-c e 6-d representam, respectivamente, os valores $\hat{\mathbf{x}}_1$ e $\hat{\mathbf{x}}_2$ mostrados na Equação (55). É possível perceber que a Figura 6-c é a que mais se aproxima da Figura 5-b. Assim, respeitando a regra estabelecida pela Equação (55) a estimativa $\hat{\mathbf{x}}$ para este exemplo é dada por $\hat{\mathbf{x}}_1$, conforme mostra a Figura 6-e.

Figura 5 – Representação gráfica do primeiro, segundo e terceiro passos do método RRVO.

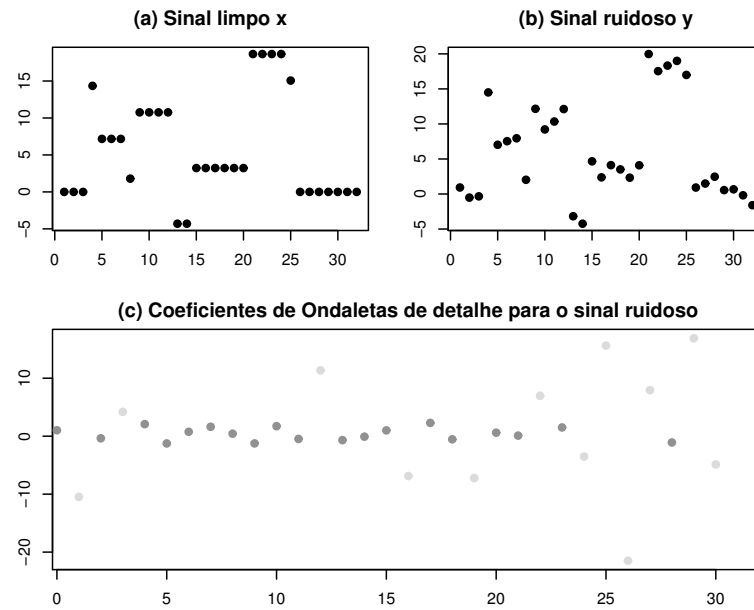
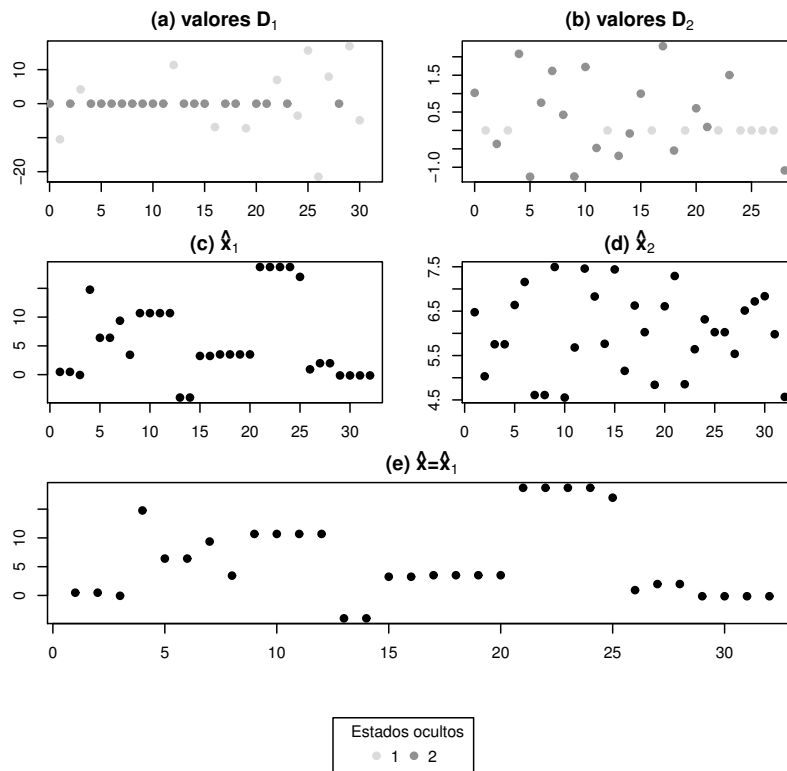


Figura 6 – Representação gráfica do quarto e quinto passos do método RRVO.



3.1 Conclusões

Neste capítulo apresentamos o método RRVO, este método agrupa os coeficientes de ondaleta de detalhe por meio do algoritmo de Viterbi e posteriormente decide qual destes conjuntos representa ruído e qual representa sinal. Este processo decisório é feito aplicando a ITOD em cada um destes conjuntos. Assim, obtemos dois candidatos a estimador do sinal limpo. O candidato que apresentar menor erro quadrático médio quando comparado ao sinal ruidoso será escolhido como estimador do sinal limpo.

Redução de ruído em sinal por clusterização e soft-thresholding

Aqui, será descrito o método RRSCST, o qual é baseado na TOD e MOM. O RRSCST é aplicado por meio dos seguintes passos (Figura 7).

Repita os quatro primeiros passos do método RRVO dados no capítulo 3. No quinto passo, deve-se decidir quais dos conjuntos (D_1 ou D_2) está estimando os coeficientes de ondaletas de sinal, denotado por $D_{\hat{\mathbf{x}}}$, e qual está estimando os coeficientes ondaletas de ruído, denotado por $D_{\hat{\mathbf{e}}}$. Esta classificação é feita da seguinte forma:

$$D_{\hat{\mathbf{x}}} = \begin{cases} D_1, & \text{se } \text{MSE}(\text{ITOD}(D_1), \mathbf{y}) < \text{MSE}(\text{ITOD}(D_2), \mathbf{y}), \\ D_2, & \text{caso contrário.} \end{cases}$$

e

$$D_{\hat{\mathbf{e}}} = \begin{cases} D_1, & \text{se } \text{MSE}(\text{ITOD}(D_1), \mathbf{y}) > \text{MSE}(\text{ITOD}(D_2), \mathbf{y}), \\ D_2, & \text{caso contrário.} \end{cases}$$

onde ITOD é a Inversa de Transformada de Ondaletas Discreta e o MSE é o erro quadrático médio.

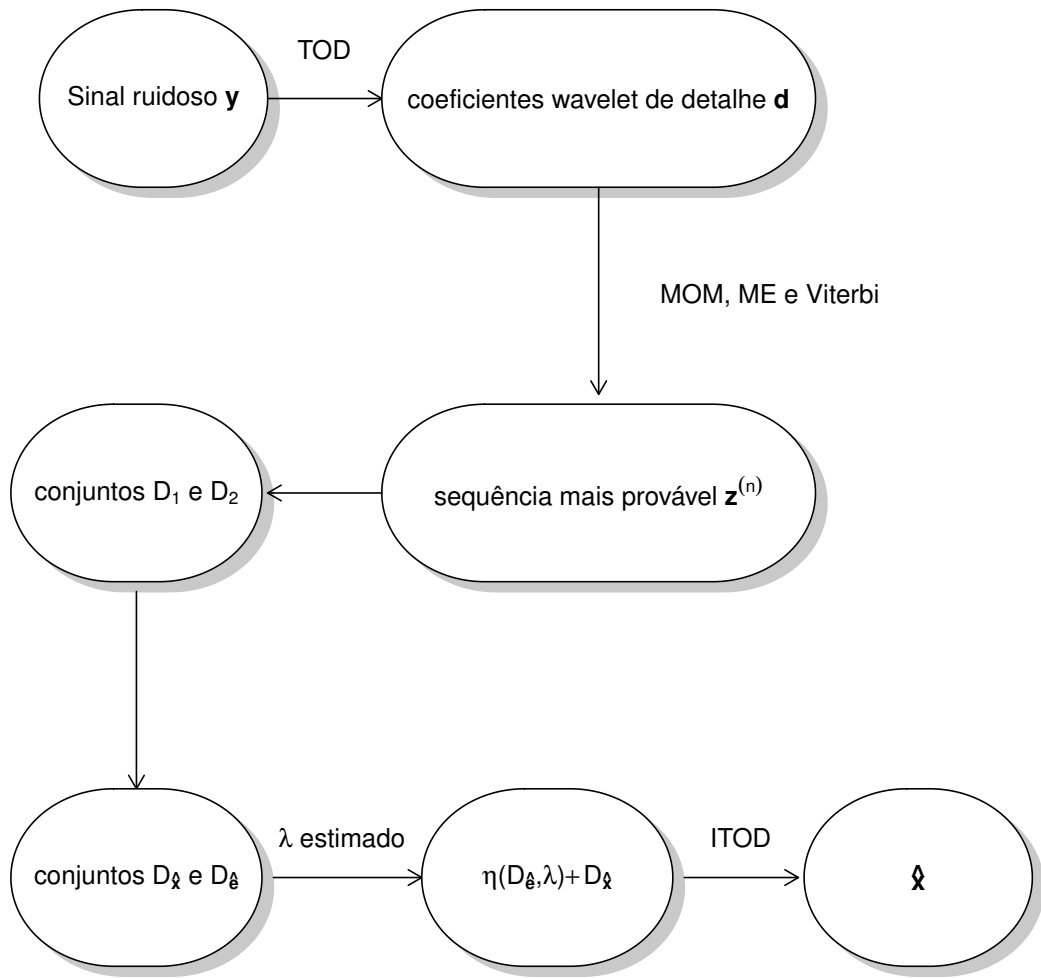
Seja $\mathcal{O} = \{o_j\}_{j=1}^{N-1}$ o conjunto ordenado de $D_{\hat{\mathbf{e}}}$, e $\mathcal{M} = \{m_i\}_{i=1}^{N-1}$ o conjunto com as somas acumuladas do quadrado de \mathcal{O} . No sexto passo, obtém-se o limiar λ dado por

$$\lambda = o_{\arg\min_{1 \leq i \leq N-1} \{N-1-2i+m_{i-1}+o_i^2(N-i)\}}.$$

Seja $\eta_S(D_{\hat{\mathbf{e}}}, \lambda)$ a imagem do conjunto $D_{\hat{\mathbf{e}}}$ sob a função de soft-thresholding $\eta_S(\cdot, \cdot)$ (veja Definição 5), e seja $\eta_S(D_{\hat{\mathbf{e}}}, \lambda) + D_{\hat{\mathbf{x}}}$ o conjunto representativo da soma dos elementos correspondentes em $\eta_S(D_{\hat{\mathbf{e}}}, \lambda)$ e $D_{\hat{\mathbf{x}}}$. Finalmente, no sétimo passo, aplica-se a ITOD no conjunto $\eta_S(D_{\hat{\mathbf{e}}}, \lambda) + D_{\hat{\mathbf{x}}}$, para obter a estimativa do sinal limpo $\hat{\mathbf{x}}$, que é dado por

$$\hat{\mathbf{x}} = \text{ITOD}(\eta_S(D_{\hat{\mathbf{e}}}, \lambda) + D_{\hat{\mathbf{x}}}).$$

Figura 7 – Fluxograma do método RRSCST



4.1 Conclusões

Neste capítulo apresentamos o método RRSCST, este método é a continuidade do método RRVO. O método RRSCST tenta identificar se não sobrou alguma informação de sinal no conjunto previamente definido como representante de ruído. Para isso é feita uma limiarização neste conjunto e os coeficientes previamente processados são somados aos respectivos coeficientes pertencentes ao conjunto representante de sinal.

Redução de ruído por meio do método SureShrink suavizado

Nesta seção será proposto um método de redução de ruído baseado no método SureShrink, chamado de SureShrinkWin. O método SureShrinkWin toma como base uma série de estimativas do sinal obtidas por uma aplicação iterativa do método SureShrink em janelas de sinal com diferentes tamanhos.

Seja \mathbf{y} um sinal ruidoso de tamanho N , obtido a partir da adição do ruído gaussiano branco \mathbf{e} no sinal limpo \mathbf{x} , onde $N = 2^J, J \in \mathbb{N}, N \geq 16$, para aplicar o método SureShrinkWin é necessário seguir os seguintes passos.

Para cada $h \in \{1, 2, 3, \dots, J-3\}$, defina $n = 16 \times 2^{(h-1)}$ e divida o sinal ruidoso \mathbf{y} em N/n janelas não sobrepostas, cada uma possuindo n observações.

Para cada janela, representada pelo intervalo $I_s = [(s-1)n+1, sn]$ com $s \in \{1, 2, 3, \dots, N/n\}$, aplique o método SureShrink no intervalo I_s . Para todo $k \in I_s \cap \mathbb{N}^*$, denota-se por $\hat{x}_h[k]$ o valor estimado de $x[k]$ no intervalo I_s .

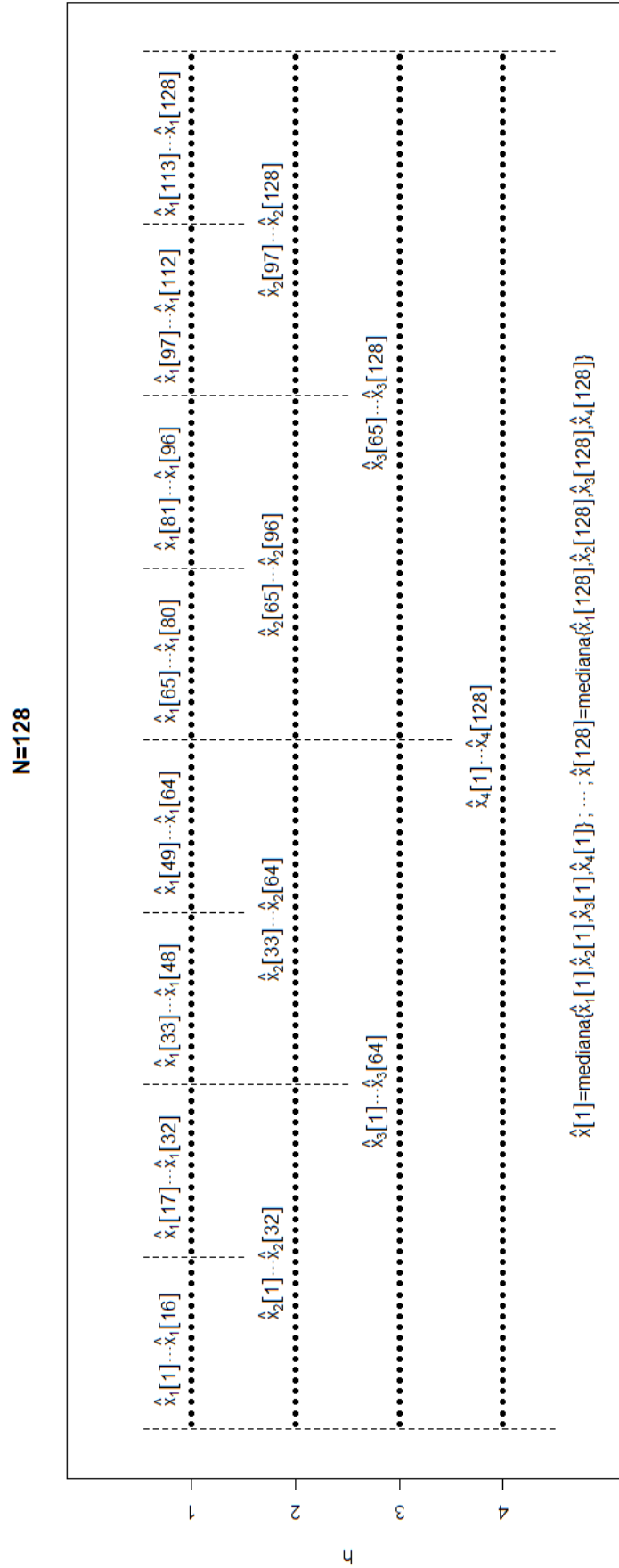
Então, a estimativa $\hat{\mathbf{x}} = (\hat{x}[1], \dots, \hat{x}[N])$ do sinal \mathbf{x} é obtida por

$$\hat{x}[k] = \text{mediana}\{\hat{x}_1[k], \dots, \hat{x}_{J-3}[k]\},$$

para todo $k \in \{1, 2, 3, \dots, N\}$.

A Figura 8 representa graficamente este processo para um sinal de tamanho $N = 128$. Como é possível perceber na imagem, inicialmente divide-se o sinal em 8 janelas, cada uma com tamanho 16, aplica-se o método SureShrink em cada uma das janelas obtendo-se assim as estimativas $\hat{x}_1[1], \dots, \hat{x}_1[128]$. No segundo janelamento tem-se 4 janelas, cada uma com tamanho 32, aplica-se o método SureShrink em cada uma das janelas obtendo-se assim as estimativas $\hat{x}_2[1], \dots, \hat{x}_2[128]$. O terceiro e quarto janelamentos são compostos respectivamente por 2 e 1 janelas com tamanhos respectivos de 64 e 128. Seguindo o mesmo procedimento obtemos os conjuntos de estimativas $\hat{x}_3[1], \dots, \hat{x}_3[128]$ e $\hat{x}_4[1], \dots, \hat{x}_4[128]$. Por fim, a estimativa final é dada pela mediana das respectivas estimativas obtidas em cada janelamento.

Figura 8 – Representação gráfica do método SureShrinkWin



5.1 Conclusões

Neste capítulo foi apresentado o método SureShrinkWin que consiste na aplicação do método SureShrink em janelas de sinal. Os respectivos estimadores prévios são processados por meio do cálculo da mediana obtendo-se assim a estimação pontual do sinal.

Redução de ruído por meio de dupla estimativa de limiares

Esta seção introduz o método de redução de ruído por meio da dupla estimativa de limiares (RRDEL) e demonstra matematicamente sua consistência. Basicamente, este método separa os coeficientes de ondaletas de detalhe em três grupos, o primeiro grupo consiste dos coeficientes que certamente representam informação de sinal, o segundo grupo consiste dos coeficientes que certamente representam informação de ruído e o terceiro grupo é formado pelos coeficientes pertencentes a uma região de indecisão, como podemos ver na figura 9. O processo decisório é feito por meio de uma função de classificação.

Considerando um sinal ruidoso \mathbf{y} com tamanho N obtido a partir da adição de um RAGB (denotado por \mathbf{e}), a aplicação do método RRDEL consiste dos seguintes passos.

Primeiramente, aplica-se a TOD no sinal \mathbf{y} para obter os coeficientes de ondaletas de detalhe $\{d_{j,k}\}$, onde $j \in \{0, \dots, J-1\}$ e $k \in \{0, \dots, 2^j-1\}$ (veja capítulo 1). No segundo passo obtém-se o limiar

$$\lambda_1 = \hat{\sigma} \sqrt{2 \log N}, \quad (56)$$

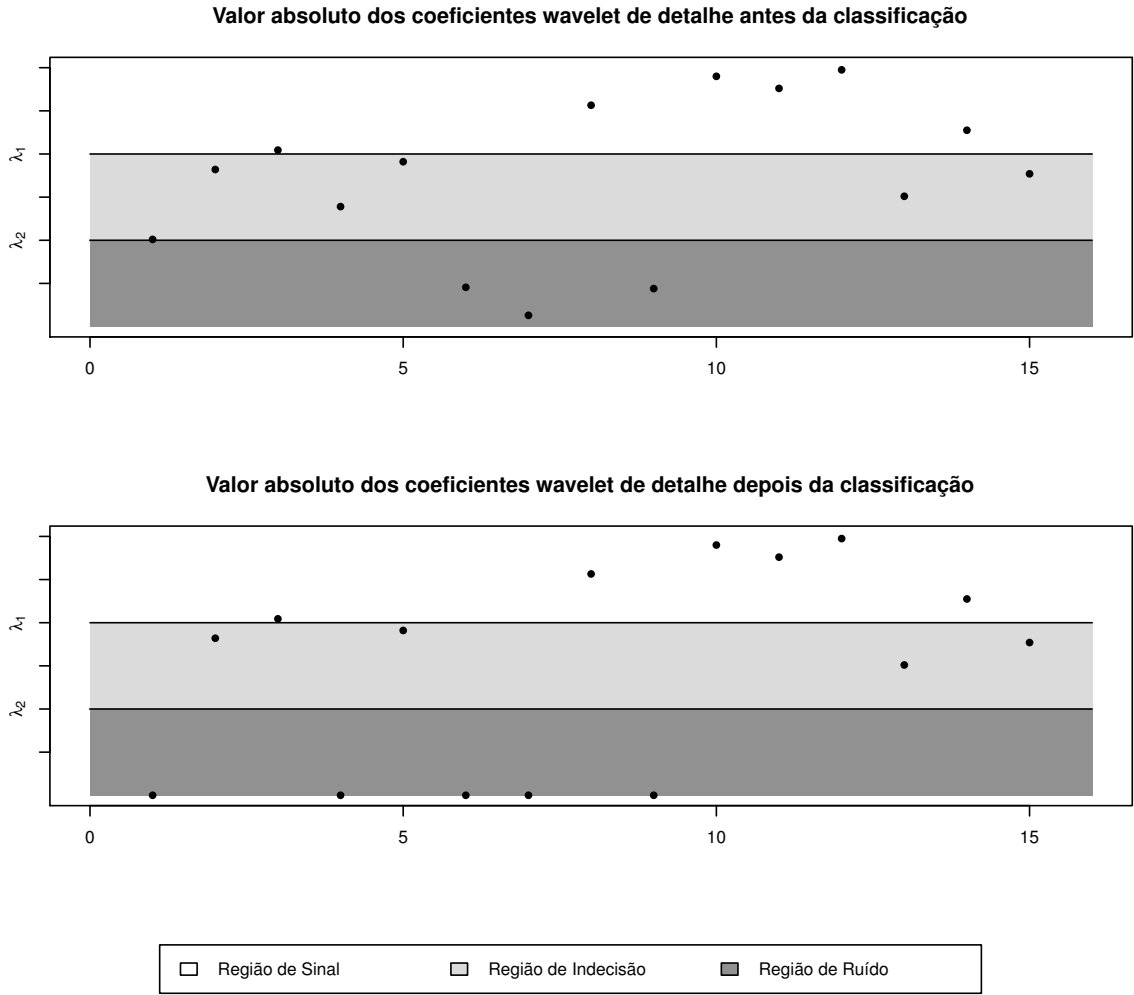
onde $\hat{\sigma}$ é a estimativa para a variância do ruído dada por $\hat{\sigma} = \text{DMA}(\{d_{J-1,k} : k = 0, 1, \dots, 2^{J-1} - 1\})$. O terceiro passo consiste em obter o limiar

$$\lambda_2 = \underset{\lambda \in [0, \lambda_1]}{\text{argmin}} \left\{ (N-1)\hat{\sigma} + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \min(d_{j,k}^2, \lambda^2) - 2\hat{\sigma} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{I}(|d_{j,k}| \leq \lambda) \right\} \quad (57)$$

onde \mathbb{I} é a função indicadora. No quarto passo aplica-se a função de limiarização dada por

$$\eta(d_{j,k}, \lambda_1, \lambda_2) = \begin{cases} d_{j,k}, & \text{se } |d_{j,k}| > \lambda_1, \\ d_{j,k} \mathbb{I}(|d_{j,k}| - \lambda_1 < |d_{j,k}| - \lambda_2), & \text{se } \lambda_2 \leq |d_{j,k}| \leq \lambda_1, \\ 0, & \text{se } |d_{j,k}| < \lambda_2. \end{cases}$$

Figura 9 – Representação gráfica do método RRDEL.



Finalmente, no quinto passo obtém-se a estimativa do sinal \hat{x} , denotada por \hat{x} , aplicando a ITOD em $\eta(d_{j,k}, \lambda_1, \lambda_2)$ no lugar de $\{d_{j,k}\}$, para todos os valores de j e k .

Para provar uma importante propriedade do método RRDEL é necessário primeiramente apresentar a desigualdade de Mill, ver Definição 17.

Definição 17 (Desigualdade de Mill). *Seja Z uma variável aleatória tal que $Z \sim \mathcal{N}(0, 1)$, então*

$$\mathbb{P}(|Z| > t) \leq \frac{2e^{-\frac{t^2}{2}}}{t\sqrt{2\pi}}, \quad (58)$$

para todo $t > 0$.

O teorema 2 demonstra matematicamente a consistência do método RRDEL. Em termos gerais, este teorema mostra que a probabilidade de classificar um coeficiente ruidoso

como representativo de sinal vai pra zero a medida que n vai para infinito, onde n é o tamanho do sinal.

Teorema 2. Seja D uma variável aleatória tal que $D \sim \mathcal{N}(0, \sigma^2)$, representando os coeficientes de ondaletas de detalhe que são considerados como informação de ruído. Então,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2) = 0, \quad (59)$$

onde n é o tamanho do sinal que está sendo processado pelo método RRDEL, λ_1 e λ_2 são dados, respectivamente, nas equações 56 e 57.

Demonstração. Seja D uma variável aleatória tal que $D \sim \mathcal{N}(0, \sigma^2)$, representando os coeficientes de ondaletas de detalhe que são considerados como informação de ruído, então a probabilidade de classificar um coeficiente ruidoso como um sinal é calculada por

$$\begin{aligned} \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2) &= \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2, |D| > \lambda_1) \\ &+ \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2, \lambda_2 \leq |D| \leq \lambda_1) \\ &+ \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2, 0 \leq |D| < \lambda_2) \\ &= A + B + C. \end{aligned}$$

Sendo assim, para provar o teorema basta mostrar que $\lim_{n \rightarrow +\infty} A = \lim_{n \rightarrow +\infty} B = \lim_{n \rightarrow +\infty} C = 0$.

Para provar que $\lim_{n \rightarrow +\infty} A = 0$, é necessário observar que

$$\mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2, |D| > \lambda_1) \leq \mathbb{P}(|D| > \lambda_1) = \mathbb{P}\left(\frac{|D|}{\sigma} > \frac{\lambda_1}{\sigma}\right), \quad (60)$$

então, pela desigualdade de Mill (veja Equação (58)) obtém-se

$$\begin{aligned} \mathbb{P}\left(\frac{|D|}{\sigma} > \frac{\lambda_1}{\sigma}\right) &\leq \frac{2e^{\frac{-\lambda_1^2}{2\sigma^2}}}{\frac{\lambda_1\sqrt{2\pi}}{\sigma}} \\ &= \frac{2e^{\frac{-\sigma^2 2 \log n}{2\sigma^2}}}{\frac{\sigma\sqrt{2 \log(n) 2\pi}}{\sigma}} \\ &= \frac{2e^{\log n^{-1}}}{2\sqrt{\log(n)\pi}} \\ &= \frac{1}{n\sqrt{\log(n)\pi}} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned} \quad (61)$$

Para provar que $\lim_{n \rightarrow +\infty} B = 0$, deve-se considerar o fato de que $\lambda_2 \leq |D| \leq \lambda_1$, equivalentemente

- uma vez que $\lambda_2 \leq |D|$, nós temos que $|D| - \lambda_2 \geq 0$, e consequentemente $||D| - \lambda_2| = |D| - \lambda_2$;
- uma vez que $\lambda_1 > |D|$, nós temos que $|D| - \lambda_1 < 0$, e consequentemente $||D| - \lambda_1| = \lambda_1 - |D|$.

Então, obtém-se

$$\begin{aligned}
B &= \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2, \lambda_2 \leq |D| \leq \lambda_1) \\
&= \mathbb{P}(\lambda_1 - |D| < |D| - \lambda_2, \lambda_2 \leq |D| \leq \lambda_1) \\
&\leq \mathbb{P}(\lambda_1 - |D| < |D| - \lambda_2) \\
&= \mathbb{P}(-2|D| < -\lambda_1 - \lambda_2) \\
&= \mathbb{P}(2|D| > \lambda_1 + \lambda_2) \\
&= \mathbb{P}\left(|D| > \frac{\lambda_1 + \lambda_2}{2}\right) \\
&= \mathbb{P}\left(\frac{|D|}{\sigma} > \frac{\lambda_1 + \lambda_2}{2\sigma}\right) \\
&\leq \frac{2e^{-\frac{(\lambda_1 + \lambda_2)^2}{8\sigma^2}}}{\frac{(\lambda_1 + \lambda_2)\sqrt{2\pi}}{2\sigma}} \text{ (pela desigualdade de Mill)} \\
&\leq \frac{2e^{-\frac{\lambda_1^2}{8\sigma^2}}}{\frac{\lambda_1\sqrt{2\pi}}{2\sigma}} \text{ (uma vez que } \lambda_2 \in [0; \lambda_1]) \\
&= \frac{2e^{-\frac{\sigma^2 2 \log n}{8\sigma^2}}}{\frac{\sigma\sqrt{2 \log(n)2\pi}}{2\sigma}} \\
&= \frac{2e^{\log n^{-\frac{1}{4}}}}{\sqrt{\pi \log(n)}} \\
&= \frac{2}{n^{\frac{1}{4}}\sqrt{\pi \log(n)}} \xrightarrow{n \rightarrow +\infty} 0.
\end{aligned}$$

Para provar que $\lim_{n \rightarrow +\infty} C = 0$, considerando o fato de que $\lambda_2 \leq \lambda_1$ (veja Equação 57). Uma vez que $0 \leq |D| < \lambda_2 \leq \lambda_1$, tem-se que

$$||D| - \lambda_2| \leq ||D| - \lambda_1|,$$

então $C = \mathbb{P}(|D| - \lambda_2 > |D| - \lambda_1, 0 < |D| < \lambda_2) = 0$.

Então, provou-se que $\lim_{n \rightarrow +\infty} A = \lim_{n \rightarrow +\infty} B = \lim_{n \rightarrow +\infty} C = 0$, equivalentemente $\lim_{n \rightarrow +\infty} \mathbb{P}(|D| - \lambda_1 < |D| - \lambda_2) = 0$. □

6.1 Conclusões

Neste capítulo foi apresentado o método RRDEL, este método propõe a estimativa de dois limiares separando os coeficientes ondaleta de detalhe em grupos de sinal, ruído e indecisão. O processo decisório é feito por meio da função de limiarização. Por meio da desigualdade de Mill é possível mostrar a robustez do método RRDEL, uma vez que a probabilidade de classificarmos um coeficiente proveniente de ruído como sendo um coeficiente de sinal vai pra zero à medida que o tamanho do sinal aumenta.

Aplicações para sinais corrompidos por ruído aditivo gaussiano branco

Neste capítulo são aplicadas simulações de Monte Carlo em três conjuntos distintos de sinais: os sinais de teste de Donoho & Johnstone, em sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH e em traços sísmicos obtidos do pacote R *astsa*. Os traços sísmicos possuem um tamanho de 2048, enquanto que os outros sinais possuem um tamanho de 1024, estes tamanhos são aceitos e utilizados na maior parte das simulações presentes na literatura.

Os sinais de teste de Donoho & Johnstone possuem características como descontinuidades e variações de frequência que os tornam amplamente utilizados na avaliação da performance de métodos de redução de ruído, esta larga utilização na literatura é motivação suficiente para sua escolha na avaliação de novos métodos. Os quatro sinais de teste de Donoho & Johnstone são denominados por: Blocks, Bumps, Doppler e Heavisine.

O sinal Blocks é uma sequência de funções escada variando em largura e altura. O sinal Bumps se caracteriza por ser um sinal com picos repentinos. O sinal Doppler é uma senoide que possui variação gradativa de frequência ao longo do tempo. O sinal Heavisine é a soma de uma senoide com um função escada.

Adicionalmente, os erros de diagnóstico (no caso dos sinais de eletrocardiograma) ou de análises (no caso de traços sísmicos) obtidos a partir da interpretação de artefatos ou atributos provenientes de sinais ruidosos faz com que a redução de ruído em ECG e em traços sísmicos seja uma preocupação recorrente na literatura associada a este tema. Por isso a escolha dos sinais provenientes da base de dados MIT-BIH e do pacote R *astsa*.

Os sinais de ECG obtidos a partir da base MIT-BIH estão subcategorizados como *normal sinus rhythm* e sua escolha foi baseada no fato destes sinais serem amplamente utilizados na literatura. Já os traços sísmicos obtidos do pacote *astsa* dividem-se em sinais provenientes de terremotos “EQ” e sinais provenientes de explosão “EX”, estes sinais também são amplamente aceitos na comunidade científica.

São consideradas 100 replicações de adições de ruído gaussiano branco em cada sinal,

com o objetivo de comparar a performance de cada um dos novos métodos propostos (RRVO, RRSCST, SureShrinkWin e RRDEL) com métodos já conhecidos na literatura: RRVt (CONDAT, 2013), HTLU (PERCIVAL; WALDEN, 2006), SureShrink (SURE) (DONOHO; JOHNSTONE, 1995) e (NLOA) (HAN; XU, 2016).

Seja \mathbf{x} um sinal não ruído com tamanho N , e seja \mathbf{y} o sinal ruído obtido a partir da adição de um ruído gaussiano branco \mathbf{e} a \mathbf{x} . Para comparar a performance dos novos métodos propostos com os outros métodos nós consideramos o SNR de saída, em decibéis (dB), dado por

$$SNR(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\sum_i^N x_i^2}{\sum_i^N (x_i - \hat{x}_i)^2}, \quad (62)$$

e o *Percentage Root Mean Square Difference* (PRD) dado por

$$PRD(\mathbf{x}, \hat{\mathbf{x}}) = 100 \sqrt{\frac{\sum_i^N (x_i - \hat{x}_i)^2}{\sum_i^N x_i^2}}, \quad (63)$$

onde $\hat{\mathbf{x}}$ é a estimativa para o sinal limpo \mathbf{x} .

Quanto maior for o SNR de saída e menor o PRD, melhor é a performance do método.

Para cada sinal \mathbf{x} , são geradas 100 replicações de séries temporais do modelo de ruído gaussiano branco, cada uma dessas séries é adicionada ao sinal \mathbf{x} , obtendo assim 100 sinais ruídos \mathbf{y} . Aplica-se cada método considerado em cada um dos sinais ruídos \mathbf{y} , calculando o SNR de saída e o PRD. Finalmente, nós obtemos a média dos SNR de saída e a média dos PRDs. Este tipo de abordagem é conhecida por simulações de Monte Carlo (MOONEY, 1997). Os resultados das simulações são mostrados nas Tabelas 1, 2 e 3. A representação gráfica destes resultados podem ser conferidos nas figuras 11, 12, 13, 14, 15, 16.

As primeiras colunas de cada tabela apresentam os sinais que estão sendo considerados. A segunda coluna de cada tabela, mostra o quão ruído é o sinal de entrada para cada uma das 100 replicações mencionadas anteriormente. Observe que o sinal mais ruído é aquele que possui relação sinal-ruído (SNR) de entrada igual a zero decibéis, e que o sinal menos ruído é aquele que possui relação sinal-ruído (SNR) de entrada igual a quinze decibéis. Da segunda à nona coluna de cada tabela encontram-se as médias do SNR de saída agrupados por método de redução de ruído. Da décima até a décima sétima coluna estão apresentadas as médias do parâmetro de performance PRD, também agrupadas por método de redução de ruído.

Considerando todos os eventos apresentados nas referidas tabelas, é possível verificar que o método RRVO apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVt em, respectivamente, 90,10%, 74,47%, 78,64% e 80,72% dos casos. Já o método RRSCST apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVt em, respectivamente, 74,47%, 75%, 63,54% e 76,04% dos casos.

Enquanto isso o método SureShrinkWin apresentou uma performance melhor que os métodos HTLU, NLOA, SURE e RRVt em, respectivamente, 92,18%, 74,47%, 76,04%

e 73,95% dos casos. Ao passo que o método RRDEL apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVT em, respectivamente, 81,25%, 66,66%, 52,6% e 65,10% dos casos.

Fazendo uma análise mais específica, percebe-se que para o sinal Blocks, o método SureShrinkWin mostrou-se mais eficaz se forem considerados os SNRs de entrada entre 0 e 12 dB. Isso, possivelmente se deve ao fato do método ter como base uma limiarização que guarda características semelhantes a da função que está sendo suavizada. Ainda assim, o método RRVO mostrou-se melhor para os SNRs de entrada entre 4 e 12 dB. Perdendo para o método RRVT apenas em 3 casos, com os SNRs de entrada variando entre 13 e 15 dB.

Para SNRs de entrada entre 3 e 7 dB, o método RRVO supera todos os métodos quando considerado o sinal Heavisine. Enquanto que o método SureShrinkWin apresenta uma melhor performance para os SNRs entre 8 e 15 dB. Assim, pode-se ver que estes métodos reduzem o ruído de sinais suaves com descontinuidade repentina de uma forma bastante eficaz.

Também é possível perceber a excelente performance do método RRVO para o sinal Doppler, considerando os SNRs de entrada entre 1 e 15 dB, mostrando sua eficácia na redução de ruído de sinais cuja frequência muda de forma gradual.

O método RRVO mostrou-se também um bom identificador de ruídos em funções que possuem picos repentinos, como é o caso do sinal Bumps. Haja vista que a performance deste método supera a de todos os outros quando considerado um alto nível de ruído, SNRs de entrada entre 1 e 12 dB.

O sinal Bumps guarda características muito próximas aos dos sinais de ECG. Por esta razão, o método RRVO foi extremamente superior aos métodos já estabelecidos pela literatura quando aplicado neste tipo de sinal. Basta ver que se forem considerados apenas os resultados dados pela Tabela 2, então o método RRVO supera os métodos HTLU, NLOA, SURE e RRVT em 95,31%, 96,87%, 95,31% e 84,37%.

Cabe destacar também a performance do método RRSCST que, se comparado ao método RRVO, supera-o em 100% quando aplicado nos sinais de ECG ruidosos. Este resultado mostra que, o método RRVO identifica como ruidosos coeficientes que, passado pela limiarização do método RRSCST, mostram possuir pelo menos um resquício de informação de sinal.

Uma análise visual interessante pode ser feita através de uma nuvem de palavras como apresentado na figura 10, para produzir este gráfico leva-se em consideração o número de vezes que cada método apresenta a melhor performance. Dito de outra forma é contada a quantidade de vezes que o PRD ou o SNR de saída do método aparece em negrito nas Tabelas 1, 2 e 3. Quanto maior for essa quantidade, maior o tamanho da palavra.

Tabela 1 – Resultados da redução de ruído Gaussiano nos sinais de teste de Donoho & Johnstone, as melhores performances estão apresentadas em negrito.

Sinal	SNR de entrada (dB)	Média-SNR de saída (dB)								Média-PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
Blocks	0	7.94	0.35	7.79	-1.67	-0.79	6.84	10.15	1.16	40.10	96.22	40.83	121.15	109.60	45.53	31.12	87.51
	1	8.59	1.21	8.45	-1.07	-0.02	7.19	10.83	2.29	37.21	87.05	37.82	113.16	100.24	43.73	28.76	76.86
	2	9.25	2.22	9.03	-0.23	0.81	7.51	11.20	3.39	34.50	77.44	35.41	104.31	91.42	42.13	27.56	67.67
	3	9.47	3.18	9.24	5.60	3.79	7.75	11.50	4.49	33.70	69.40	34.59	63.48	67.23	40.98	26.62	59.67
	4	9.97	4.26	9.70	11.12	7.06	8.14	11.98	5.62	31.76	61.37	32.75	27.80	44.41	39.18	25.18	52.41
	5	10.42	5.12	10.14	11.35	8.02	8.64	12.31	6.84	30.15	55.47	31.12	27.12	39.76	36.98	24.27	45.52
	6	10.81	6.26	10.59	11.99	9.11	9.04	12.79	8.09	28.82	48.67	29.55	25.17	35.07	35.34	22.93	39.44
	7	11.02	7.18	10.73	12.34	9.85	9.35	13.11	9.21	28.13	43.80	29.09	24.16	32.18	34.09	22.11	34.65
	8	11.39	8.03	11.58	12.78	10.82	11.24	13.50	10.44	26.97	39.70	26.42	22.98	28.79	27.94	21.15	30.07
	9	11.96	9.24	13.02	13.65	12.12	13.57	14.50	11.83	25.25	34.55	22.37	20.82	24.80	21.02	18.85	25.62
	10	12.42	10.35	13.71	14.33	13.09	14.26	14.99	13.14	23.95	30.45	20.66	19.24	22.17	19.40	17.80	22.04
	11	13.03	11.26	14.46	15.26	14.19	15.09	15.69	14.57	22.32	27.38	18.93	17.29	19.53	17.63	16.44	18.69
	12	13.60	12.05	15.31	16.09	15.36	15.77	16.23	15.91	20.89	25.00	17.16	15.70	17.07	16.30	15.45	16.03
	13	14.29	13.25	16.14	16.50	16.48	16.40	16.99	17.34	19.32	21.78	15.60	14.97	15.01	15.15	14.16	13.60
	14	15.16	14.12	16.99	17.28	17.58	17.46	18.03	18.76	17.47	19.70	14.16	13.68	13.22	13.42	12.58	11.55
	15	15.78	15.17	18.02	18.06	18.50	18.24	18.99	20.29	16.26	17.45	12.56	12.52	11.90	12.25	11.25	9.68
Bumps	0	3.16	0.05	2.91	1.88	0.56	1.65	4.46	1.18	69.56	99.46	71.59	88.97	96.15	82.74	59.85	87.36
	1	3.67	1.19	3.47	6.30	3.45	1.94	4.60	2.50	65.55	87.29	67.08	48.74	67.35	80.03	58.95	74.99
	2	4.29	2.13	4.38	7.29	4.60	3.30	5.45	3.60	61.08	78.29	60.77	43.28	58.98	69.91	53.65	66.11
	3	4.93	3.16	5.77	7.98	5.58	5.76	6.78	4.76	56.73	69.52	51.79	39.93	52.65	53.54	46.01	57.83
	4	5.58	4.23	7.32	9.27	6.78	7.69	7.97	6.00	52.62	61.51	43.07	34.44	45.87	41.31	40.01	50.14
	5	6.15	5.36	8.16	10.04	7.76	8.64	8.92	7.15	49.29	54.03	39.12	31.50	40.97	37.02	35.91	43.93
	6	6.92	6.28	9.26	11.10	8.81	9.33	9.93	8.39	45.11	48.58	34.49	27.89	36.26	34.17	31.90	38.05
	7	7.72	7.21	10.04	12.06	9.89	10.12	10.73	9.58	41.18	43.64	31.52	24.98	32.05	31.24	29.11	33.22
	8	8.56	8.41	11.32	13.14	11.34	11.04	11.83	11.09	37.36	38.02	27.21	22.04	27.12	28.10	25.67	27.92
	9	9.49	9.23	12.18	13.94	12.26	11.81	12.45	12.24	33.56	34.60	24.61	20.10	24.39	25.70	23.86	24.44
	10	10.57	10.26	13.03	14.62	13.35	12.73	13.30	13.50	29.65	30.71	22.33	18.59	21.52	23.12	21.64	21.16
	11	11.59	11.28	13.94	15.49	14.51	13.55	14.32	14.76	26.36	27.32	20.09	16.82	18.84	21.05	19.25	18.29
	12	12.30	12.12	14.78	16.31	15.57	14.27	14.95	15.94	24.30	24.79	18.26	15.31	16.66	19.38	17.90	15.97
	13	13.42	13.17	15.87	17.12	16.61	15.31	15.98	17.25	21.34	21.95	16.10	13.95	14.79	17.18	15.89	13.72
	14	14.07	14.15	16.48	17.93	17.69	15.92	16.60	18.36	19.81	19.62	15.01	12.70	13.06	16.02	14.81	12.08
	15	15.03	15.14	17.73	18.92	18.86	17.00	17.71	19.75	17.73	17.51	13.00	11.34	11.41	14.14	13.03	10.30
Heavisine	0	16.45	0.22	14.53	-0.74	0.14	16.90	14.13	1.29	15.30	97.59	18.97	124.84	99.25	14.35	19.70	86.21
	1	17.05	1.46	15.19	17.31	4.93	17.56	14.75	2.54	14.16	84.65	17.53	13.69	56.74	13.28	18.32	74.63
	2	17.18	2.16	15.82	16.98	5.82	17.50	15.52	3.71	13.92	78.00	16.42	14.28	51.21	13.36	16.80	65.28
	3	17.64	3.15	16.96	17.96	7.15	17.93	16.29	4.88	13.16	69.61	14.32	12.66	43.97	12.70	15.35	57.01
	4	17.97	4.32	17.90	18.19	8.40	18.09	17.18	6.10	12.65	60.89	12.77	12.34	38.05	12.47	13.88	49.56
	5	18.24	5.42	18.21	18.71	9.43	18.45	18.05	7.49	12.26	53.67	12.32	11.63	33.81	11.96	12.54	42.22
	6	18.57	6.47	18.57	18.96	10.50	18.81	18.63	8.63	11.80	47.60	11.80	11.29	29.87	11.48	11.73	37.03
	7	19.09	7.49	19.09	19.62	11.63	18.88	19.49	9.86	11.15	42.36	11.15	10.46	26.28	11.38	10.64	32.16
	8	19.26	8.25	19.26	19.94	12.98	19.15	20.14	11.23	10.91	38.72	10.91	10.12	22.49	11.03	9.85	27.45
	9	19.50	9.28	19.50	20.04	14.33	19.16	20.80	12.52	10.62	34.40	10.62	9.97	19.23	11.03	9.14	23.66
	10	20.13	10.20	20.13	20.51	15.82	19.41	21.68	14.09	9.86	30.94	9.86	9.46	16.20	10.71	8.25	19.76
	11	20.98	11.19	20.98	21.21	17.02	19.81	22.12	15.44	8.96	27.58	8.96	8.72	14.13	10.23	7.85	16.92
	12	21.23	12.19	21.23	21.55	18.60	19.97	22.70	16.90	8.71	24.60	8.71	8.40	11.78	10.04	7.34	14.30
	13	21.81	13.19	21.81	22.24	20.36	20.38	23.18	18.23	8.13	21.93	8.13	7.76	9.64	9.59	6.94	12.27
	14	22.26	14.26	22.26	22.88	21.90	20.71	23.68	19.73	7.72	19.40	7.72	7.21	8.10	9.23	6.55	10.32
	15	23.00	15.18	23.00	23.37	23.38	21.13	24.28	21.20	7.09	17.44	7.09	6.81	8.81	8.78	6.12	8.73
Doppler	0	8.79	0.27	8.56	1.73	-0.14	6.07	10.09	1.33	36.51	97.03	37.48	102.59	105.84	49.76	31.30	85.84
	1	9.58	1.43	9.22	11.92	4.71	6.66	10.72	2.53	33.24	84.93	34.68	25.55	58.22	46.50	29.15	74.76
	2	10.44	2.18	10.15	12.91	5.71	7.28	11.49	3.73	30.15	77.84	31.19	22.73	51.87	43.32	26.67	65.14
	3	11.18	3.36	10.97	13.51	6.72	7.93	11.75	4.82	27.71	68.21	28.40	21.17	46.15	40.32	25.85	57.43
	4	12.10	4.23	12.09	14.51	7.92	10.57	12.89	6.17	24.88	61.50	24.95	18.94	40.23	30.34	22.71	49.15
	5	12.83	5.30	12.92	15.05	9.09	12.62	13.77	7.36	22.88	54.34	22.70	17.73	35.15	23.42	20.54	42.89
	6	14.10	6.11	13.65	15.95	10.36	13.61	14.76	8.59	19.80	49.51	20.87	15.95	30.39	20.91	18.29	37.20
	7	15.25	7.30	14.38	16.76	11.51	14.49	15.65	9.93	17.33	43.19	19.16	14.57	26.60	18.90	16.53	31.87
	8	15.89	8.38	15.62	17.48	12.79	15.22	16.36	11.24	16.08	38.17	16.60	13.40	22.96	17.36	15.23	27.42
	9	16.73	9.01	16.37	17.87	13.79	15.69	16.95	12.40	14.60	35.44	15.23	12.82	20.48	16.44	14.23	24.00
	10	17.34	10.15	17.11	18.85	15.39	16.98	17.89	13.71	13.60	31.10	13.98	11.44	17.01	14.17	12.76	20.64
	11	18.03	11.24	18.31	19.44	16.50	17.43	18.29	15.01	12.58	27.44	12.19	10.69	15.00	13.44	12.18	17.76
	12	18.84	12.16	19.05	20.47	18.01	18.32	19.11	16.26	11.46	24.67	11.18	9.50	12.62	12.15	11.08	15.40
	13	19.60	13.18	20.18	21.65	19.84	19.33	20.09	17.55	10.48	21.94	9.84	8.29	10.22	10.81	9.91	13.27
	14	20.10	14.42	21.01	22.42	21.37	20.18	20.76	18.73	9.90	19.07	8.95	7.59	8.56	9.81	9.18	11.57
	15	20.97	15.21	21.69	23.49	23.02	20.93	21.35	19.74	8.96	17.37	8.24	6.71	7.08	9.00	8.57	10.31

Tabela 2 – Resultados da redução de ruído Gaussiano nos sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH, as melhores performances estão apresentadas em negrito.

Sinal	SNR de entrada (dB)	Média-SNR de saída (dB)								Média - PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
16265 ECG1	0	0.64	0.16	0.66	3.50	3.52	0.21	1.98	6.97	92.90	98.31	92.68	66.89	66.68	97.60	79.65	44.86
	1	1.06	1.19	1.07	4.62	4.65	0.34	2.24	7.48	88.57	87.27	88.39	58.83	58.67	96.16	77.30	42.28
	2	1.67	2.17	1.68	5.74	5.74	0.58	2.80	8.18	82.62	77.93	82.49	51.69	51.68	93.56	72.52	39.01
	3	2.08	3.02	2.16	6.80	6.80	0.86	3.34	8.50	78.73	70.68	78.17	45.76	45.76	91.01	68.31	37.60
	4	2.79	4.18	4.82	7.82	7.82	5.03	5.63	8.69	72.57	61.90	58.14	40.67	40.67	58.22	52.94	36.82
	5	3.36	5.12	6.36	8.88	8.88	6.96	7.25	9.06	67.94	55.49	48.15	35.99	35.99	44.96	43.54	35.27
	6	4.25	6.39	7.29	9.70	9.70	7.86	8.60	9.32	61.36	48.00	43.26	32.77	32.77	40.53	37.17	34.21
	7	4.80	7.22	8.07	10.73	10.73	8.40	9.21	9.37	57.65	43.61	39.52	29.07	29.07	38.14	34.70	34.03
	8	5.85	8.19	9.29	11.68	11.68	9.55	10.42	9.51	51.04	39.02	34.36	26.08	26.08	33.33	30.14	33.47
	9	6.85	9.21	10.02	12.28	12.28	9.89	10.96	9.58	45.50	34.67	31.58	24.34	24.34	32.06	28.33	33.18
	10	7.63	9.96	10.93	12.94	12.94	10.74	11.84	9.64	41.57	31.78	28.43	22.55	22.55	29.07	25.62	32.95
	11	8.68	11.30	11.84	13.72	13.72	11.83	12.68	9.57	36.84	27.26	25.59	20.62	20.62	25.66	23.25	33.22
	12	9.70	12.16	12.53	14.45	14.45	12.38	13.45	9.65	32.80	24.66	23.64	18.96	18.96	24.08	21.26	32.93
	13	10.62	13.16	13.26	14.98	14.98	13.19	14.24	9.67	29.48	21.97	21.74	17.84	17.83	21.96	19.43	32.87
	14	11.57	14.10	13.99	15.59	15.59	14.22	14.99	9.69	26.42	19.74	19.99	16.63	16.62	19.50	17.83	32.76
	15	12.17	15.10	14.53	16.32	16.32	14.53	15.72	9.73	24.63	17.59	18.79	15.29	15.29	18.80	16.38	32.63
16265 ECG2	0	1.69	0.18	1.66	-0.78	-0.16	1.05	2.17	1.46	82.32	97.97	82.64	111.17	103.10	88.61	77.89	84.49
	1	2.11	1.38	2.06	0.37	1.06	1.19	2.46	1.46	78.48	85.48	78.94	98.97	90.44	87.22	75.36	84.51
	2	2.59	2.20	2.53	4.50	4.56	1.41	2.76	1.49	74.24	77.71	74.80	60.45	59.67	85.07	72.81	84.21
	3	2.90	3.23	2.85	5.45	5.45	1.64	2.94	1.50	71.60	69.01	72.04	53.44	53.44	82.92	71.34	84.12
	4	3.29	4.12	3.57	6.17	6.17	2.79	3.58	1.48	68.48	62.27	66.60	49.20	49.19	74.00	66.37	84.32
	5	3.75	5.25	4.98	7.34	7.34	5.50	4.96	1.50	64.96	54.65	56.75	42.99	42.99	55.10	56.84	84.16
	6	4.12	6.26	6.18	8.08	8.08	7.44	6.15	1.49	62.24	48.68	49.15	39.50	39.49	42.60	49.47	84.26
	7	4.61	7.12	6.78	9.01	9.01	7.98	7.58	1.48	58.87	44.09	45.86	35.45	35.45	39.94	41.87	84.33
	8	5.14	8.16	7.50	9.96	9.96	8.73	9.04	1.48	55.33	39.09	42.20	31.78	31.77	36.67	35.42	84.32
	9	5.46	9.14	7.99	10.86	10.87	9.11	9.76	1.47	53.36	34.91	39.89	28.67	28.66	35.11	32.57	84.38
	10	5.99	10.23	8.94	11.74	11.75	10.16	10.72	1.49	50.20	30.84	35.73	25.88	25.87	31.06	29.13	84.22
	11	6.34	11.19	9.65	12.58	12.59	10.76	11.31	1.49	48.22	27.64	32.95	23.50	23.49	29.00	27.21	84.20
	12	6.86	12.14	10.26	13.49	13.50	11.36	12.02	1.48	45.39	24.74	30.73	21.17	21.15	27.08	25.09	84.35
	13	7.71	13.21	11.33	14.42	14.42	12.29	12.86	1.48	41.19	21.86	27.16	19.03	19.02	24.33	22.77	84.29
	14	8.26	14.05	11.82	14.99	15.01	12.70	13.25	1.47	38.67	19.84	25.65	17.80	17.78	23.21	21.76	84.39
	15	9.03	15.13	12.49	15.70	15.73	13.64	14.22	1.49	35.37	17.53	23.76	16.41	16.37	20.83	19.46	84.27
16272 ECG1	0	1.71	0.18	1.70	2.93	3.18	0.91	2.77	3.41	82.19	98.03	82.27	75.80	72.42	90.03	72.72	67.56
	1	2.05	1.14	2.05	5.38	5.39	1.13	3.15	3.35	79.00	87.80	79.02	54.11	53.96	87.81	69.58	68.06
	2	2.47	2.08	2.46	6.04	6.04	1.25	3.37	3.25	75.27	78.75	75.39	49.93	49.93	86.56	67.84	68.82
	3	2.98	3.33	3.23	7.04	7.04	2.16	4.07	3.41	70.96	68.23	69.18	44.49	44.49	79.15	62.77	67.59
	4	3.54	4.19	5.16	8.16	8.16	5.33	5.70	3.52	66.54	61.78	55.84	39.17	39.17	56.48	52.33	66.70
	5	4.13	5.24	6.62	9.16	9.16	7.57	7.66	3.39	62.17	54.76	46.70	34.85	34.85	41.87	41.51	67.71
	6	4.84	6.09	7.60	10.24	10.24	8.38	8.88	3.45	57.31	49.63	41.73	30.78	30.78	38.14	36.02	67.22
	7	5.55	7.36	8.75	11.19	11.19	9.24	9.97	3.49	52.82	42.88	36.60	27.61	27.61	34.59	31.78	66.91
	8	6.43	8.30	9.71	12.05	12.05	9.79	10.68	3.43	47.72	38.49	32.72	24.97	24.97	32.43	29.27	67.42
	9	7.31	9.47	10.71	13.04	13.04	10.71	11.66	3.40	43.13	33.64	29.16	22.30	22.30	29.15	26.12	67.59
	10	8.07	10.11	11.79	13.90	13.90	11.39	12.37	3.45	39.53	31.24	25.76	20.18	20.18	26.98	24.08	67.25
	11	9.44	11.21	12.79	14.67	14.68	12.32	13.22	3.41	33.76	27.53	22.94	18.48	18.48	24.24	21.85	67.53
	12	10.29	12.12	13.58	15.34	15.34	13.02	13.95	3.41	30.64	24.80	20.95	17.11	17.11	22.36	20.08	67.50
	13	11.46	13.08	14.51	16.22	16.22	13.96	14.81	3.41	26.77	22.20	18.83	15.46	15.46	20.07	18.19	67.54
	14	12.36	14.20	15.26	17.02	17.02	14.89	15.58	3.39	24.12	19.50	17.26	14.09	14.09	18.02	16.65	67.66
	15	13.36	15.19	15.89	17.63	17.63	15.43	16.36	3.41	21.53	17.41	16.06	13.15	13.15	16.95	15.21	67.51
16272 ECG2	0	2.60	0.16	2.40	0.84	1.18	2.17	4.62	3.82	74.15	98.28	75.91	95.93	91.49	77.91	58.74	64.40
	1	2.99	1.25	2.78	5.11	5.22	2.29	5.01	3.86	70.93	86.68	72.67	58.59	57.24	76.85	56.16	64.16
	2	3.55	2.14	3.40	6.98	6.98	2.49	5.41	3.86	66.49	78.20	67.65	44.84	44.83	75.06	53.65	64.17
	3	4.18	3.26	4.06	8.08	8.08	2.74	5.82	3.83	61.87	68.80	62.73	39.48	39.48	50.52	72.92	51.19
	4	4.64	4.42	4.84	8.97	8.97	3.74	6.27	3.79	58.69	60.17	57.67	35.64	35.64	66.44	48.68	64.63
	5	5.55	5.56	6.73	9.97	9.97	6.10	7.52	3.75	52.81	52.82	46.71	31.79	31.78	51.87	42.38	64.96
	6	6.38	6.36	8.73	10.70	10.70	8.83	9.06	3.72	48.02	48.30	36.65	29.18	29.18	36.23	35.34	65.21
	7	7.35	7.22	9.81	11.42	11.42	9.83	10.22	3.83	42.93	43.60	32.34	26.88	26.88	32.29	30.87	64.37
	8	8.16	8.36	10.71	12.44	12.44	10.70	11.31	3.88	39.12	38.22	29.15	23.89	23.89	29.21	27.25	63.96
	9	9.40	9.49	11.52	13.06	13.06	11.44	12.19	3.80	33.91	33.59	26.57	22.25	22.25	26.82	24.59	64.56
	10	10.26	10.59	12.34	13.75	13.75	12.29	13.08	3.80	30.70	29.59	24.16	20.54	20.54	24.30	22.19	64.54
	11	11.04	11.71	12.97	14.35	14.35	12.76	13.71	3.77	28.09	26.03	22.48	19.18	19.18	23.05	20.65	64.77
	12	11.66	12.74	13.82	14.91	14.91	13.65	14.49	3.83	26.14	23.11	20.38	17.98	17.98	20.80	18.86	64.34
	13	12.27	13.63	14.28	15.53	15.53	14.24	15.08	3.83	24.35	20.86	19.33	16.75	16.75	19.45	17.62	64.32
	14	13.05	14.64	15.00	16.35	16.35	15.08	15.83	3.79	22.27	18.58	17.79	15.24	15.24	17.64	16.17	64.67
	15	13.57	15.92	15.78	17.05	17.06	15.79	16.51	3.80	20.97	16.06	16.26	14.05	14.05	16.25	14.96	64.56

Tabela 3 – Resultados da redução de ruído Gaussiano nos traços sísmicos obtidos a partir do pacote R astsa, as melhores performances estão apresentadas em negrito.

Sinal	SNR de entrada (dB)	Média – SNR de saída (dB)								Média - PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
EQ2	0	3,74	0,06	3,77	-2,39	-1,34	2,02	4,47	5,19	65,08	99,33	64,84	132,13	117,06	79,22	59,90	55,06
	1	4,24	1,14	4,42	6,56	6,63	2,83	5,29	5,92	61,45	87,71	60,22	48,00	47,32	72,91	54,42	50,60
	2	4,78	2,13	5,85	7,12	7,15	6,13	6,37	6,57	57,73	78,24	51,06	44,07	43,90	49,58	48,07	46,97
	3	5,26	3,13	6,25	7,43	7,48	6,54	6,78	7,06	54,57	69,79	48,72	42,49	42,26	47,10	45,84	44,35
	4	5,74	4,10	6,75	7,74	7,77	7,13	7,27	7,44	51,66	62,40	45,99	41,03	40,87	44,01	43,34	42,48
	5	6,24	5,13	7,19	7,96	8,05	7,77	7,95	7,81	48,77	55,46	43,69	40,00	39,57	40,90	40,08	40,68
	6	6,50	6,13	7,58	8,18	8,24	8,12	8,48	8,01	47,35	49,37	41,79	39,01	38,74	39,29	37,68	39,76
	7	6,81	7,05	8,01	8,48	8,55	8,58	8,88	8,14	45,68	44,41	39,75	37,67	37,37	37,27	36,00	39,20
	8	7,10	8,03	8,40	8,70	8,76	9,23	9,40	8,31	44,18	39,67	38,04	36,74	36,48	34,58	33,91	38,43
	9	7,31	9,07	8,84	9,14	9,20	9,66	9,74	8,40	43,09	35,19	36,16	34,93	34,71	32,91	32,60	38,02
	10	7,66	10,07	9,28	9,53	9,59	10,27	10,16	8,48	41,38	31,38	34,37	33,40	33,19	30,69	31,07	37,68
	11	7,99	11,23	9,78	9,93	10,00	10,92	10,60	8,53	39,89	27,47	32,45	31,88	31,65	28,46	29,52	37,47
	12	8,26	12,08	10,21	10,45	10,51	11,49	10,99	8,57	38,62	24,90	30,87	30,04	29,83	26,67	28,23	37,26
	13	8,50	13,02	10,55	10,71	10,80	11,83	11,28	8,60	37,57	22,33	29,68	29,14	28,86	25,68	27,31	37,16
	14	8,87	14,02	11,04	11,05	11,15	12,44	11,54	8,63	36,01	19,91	28,05	28,02	27,71	23,89	26,48	37,02
	15	9,01	15,07	11,52	11,52	11,66	12,88	11,81	8,63	35,44	17,64	26,55	26,57	26,13	22,71	25,67	37,03
EQ4	0	1,92	0,04	1,94	-2,14	-1,18	0,89	2,10	4,47	80,22	99,55	80,00	128,01	114,61	90,27	78,55	59,77
	1	2,20	1,06	2,21	-1,62	-0,46	1,08	2,39	5,26	77,67	88,54	77,54	120,57	105,50	88,31	76,06	54,59
	2	2,53	2,05	2,54	2,58	3,23	1,29	3,78	5,80	74,74	79,00	74,62	79,02	71,78	86,21	65,16	51,29
	3	2,83	3,14	3,25	5,22	5,42	2,84	5,02	6,25	72,24	69,69	68,98	54,84	53,57	73,72	56,15	48,68
	4	3,10	4,12	4,34	5,35	5,58	5,02	5,63	6,51	69,96	62,23	60,77	54,05	52,62	56,55	52,34	47,25
	5	3,30	5,01	4,77	5,65	5,86	5,81	5,98	6,71	68,36	56,21	57,78	52,21	50,91	51,27	50,28	46,18
	6	3,54	6,12	5,17	5,99	6,18	6,06	6,30	6,91	66,53	49,46	55,17	50,19	49,08	49,82	48,43	45,13
	7	3,81	7,02	5,62	6,54	6,67	6,40	6,61	7,00	64,47	44,58	52,38	47,12	46,43	47,90	46,75	44,68
	8	4,12	8,14	5,99	7,02	7,09	7,04	7,01	7,19	62,21	39,18	50,17	44,60	44,22	44,48	44,61	43,68
	9	4,35	8,95	6,29	7,29	7,37	7,37	7,23	7,19	60,62	35,69	48,49	43,23	42,82	42,85	43,50	43,71
	10	4,60	10,06	6,57	7,85	7,89	7,73	7,50	7,31	58,88	31,41	46,92	40,53	40,35	41,15	42,18	43,10
	11	4,73	10,98	6,84	8,15	8,19	8,13	7,73	7,33	58,01	28,26	45,51	39,16	38,98	39,25	41,07	43,00
	12	5,05	12,03	7,19	8,32	8,39	8,57	7,98	7,37	55,94	25,03	43,71	38,37	38,06	37,30	39,91	42,81
	13	5,19	13,01	7,46	8,74	8,79	8,94	8,24	7,42	55,01	22,38	42,36	36,59	36,34	35,77	38,75	42,54
	14	5,39	14,05	7,63	8,89	8,93	9,17	8,36	7,45	53,76	19,84	41,57	35,95	35,80	34,83	38,21	42,40
	15	5,55	15,08	7,88	9,07	9,10	9,65	8,52	7,47	52,81	17,62	40,36	35,22	35,09	32,97	37,50	42,30
EX2	0	2,06	0,11	2,08	-2,03	-1,28	1,24	1,80	3,91	78,89	98,74	78,68	126,31	115,84	86,70	81,27	63,74
	1	2,35	1,03	2,37	-0,91	-0,09	1,42	2,09	4,41	76,34	88,83	76,18	113,14	102,45	84,94	78,65	60,20
	2	2,56	2,16	2,57	0,80	1,55	1,58	2,33	4,85	74,51	77,97	74,37	95,28	86,20	83,41	76,50	57,24
	3	2,81	3,10	3,31	4,90	5,10	3,19	3,18	5,19	72,40	69,98	68,49	57,20	55,77	70,54	69,71	55,02
	4	3,03	4,20	4,20	5,57	5,68	4,96	3,49	5,38	70,54	61,67	61,79	52,68	52,05	57,04	67,10	53,83
	5	3,19	5,07	4,68	6,11	6,15	5,70	4,12	5,59	69,23	55,76	58,32	49,50	49,29	51,92	62,41	52,53
	6	3,50	6,00	5,04	6,55	6,56	6,07	4,93	5,68	66,85	50,10	56,00	47,04	47,00	49,76	57,04	52,02
	7	3,76	7,31	5,52	6,96	6,97	6,67	6,07	5,76	64,87	43,11	52,96	44,91	44,82	46,47	49,72	51,51
	8	3,99	8,01	5,80	7,28	7,29	7,06	6,36	5,80	63,21	39,78	51,30	43,24	43,19	44,42	48,11	51,32
	9	4,18	9,10	6,14	7,69	7,72	7,38	6,63	5,86	61,81	35,08	49,31	41,25	41,15	42,80	46,66	50,95
	10	4,55	10,14	6,54	8,08	8,09	7,92	7,03	5,85	59,24	31,13	47,09	39,45	39,42	40,21	44,51	50,97
	11	4,71	11,04	6,83	8,35	8,36	8,17	7,18	5,87	58,17	28,07	45,54	38,25	38,19	39,08	43,76	50,90
	12	4,84	11,99	7,00	8,37	8,38	8,46	7,35	5,84	57,30	25,15	44,67	38,17	38,12	37,79	42,91	51,06
	13	5,02	13,06	7,25	8,70	8,72	8,79	7,75	5,87	56,13	22,24	43,42	36,74	36,65	36,40	41,01	50,87
	14	5,19	14,08	7,51	8,96	8,98	9,02	8,02	5,87	55,03	19,76	42,14	35,66	35,58	35,47	39,77	50,85
	15	5,28	15,10	7,63	9,18	9,20	9,31	8,04	5,86	54,43	17,58	41,57	34,75	34,69	34,29	39,64	50,92
EX4	0	0,83	0,06	0,85	-1,62	-0,92	0,33	0,69	2,94	90,87	99,32	90,68	120,99	111,49	96,23	92,36	71,29
	1	0,99	1,07	1,00	0,83	1,36	0,43	0,83	3,41	89,27	88,40	89,10	94,19	87,62	95,16	90,94	67,52
	2	1,19	1,99	1,20	3,21	3,45	0,54	0,93	3,66	87,24	79,51	87,10	70,61	68,07	93,93	89,83	65,58
	3	1,37	3,04	1,38	4,22	4,33	0,62	1,11	4,01	85,38	70,47	85,27	61,57	60,77	93,10	87,98	63,01
	4	1,61	4,08	1,62	4,70	4,77	0,74	1,27	4,12	83,10	62,53	83,03	58,28	57,80	91,88	86,41	62,22
	5	1,71	5,01	1,77	5,28	5,32	0,91	1,62	4,23	82,14	56,21	81,66	54,55	54,26	90,31	83,11	61,43
	6	2,05	6,07	3,21	5,90	5,95	3,61	2,62	4,31	79,01	49,74	69,46	50,73	50,48	67,58	74,20	60,87
	7	2,19	7,04	3,73	6,29	6,31	4,60	2,96	4,35	77,68	44,46	65,27	48,53	48,45	59,65	71,16	60,58
	8	2,44	8,02	4,13	6,94	6,95	5,27	3,70	4,41	75,51	39,73	62,14	45,01	44,97	54,54	65,41	60,21
	9	2,66	9,05	4,39	7,55	7,56	5,65	4,02	4,41	73,66	35,27	60,36	41,98	41,93	52,28	63,00	60,19
	10	2,84	10,02	4,63	7,86	7,86	6,05	4,74	4,41	72,13	31,56	58,70	40,50	40,47	49,94	58,00	60,18
	11	2,97	11,02	4,78	8,20	8,21	6,36	4,99	4,45	71,06	28,11	57,66	38,92	38,88	48,14	56,35	59,91
	12	3,14	12,03	4,98	8,49	8,49	6,56	5,43	4,46	69,64	25,03	56,36	37,62	37,62	47,06	53,60	59,84
	13	3,25	13,04	5,16	8,80	8,80	6,76	5,74	4,47	68,82	22,29	55,22	36,32	36,31	45,99	51,67	59,77
	14	3,31	14,08	5,29	9,01	9,01	7,14	5,95	4,49	68,32	19,77	54,41	35,45	35,44	44,04	50,42	59,66
	15	3,39	14,96	5,26	9,14	9,14	6,84	5,85	4,48	67,71	17,87	54,56	34,93	34,92	45,53	51,03	59,67

Figura 10 – Nuvem de palavras representado o número de vezes que cada método obteve a melhor performance



7.1 Conclusões

Neste capítulo foram feitas simulações com sinais corrompidos por ruído gaussiano, os resultados foram interpretados e apresentados tanto na forma de tabela quanto na forma gráfica. Foi possível verificar que os métodos propostos por este trabalho foram eficazes quando comparados a métodos tradicionais na literatura. O tempo computacional médio - para redução de ruído em apenas um sinal - dos métodos RRVO, RRSCST, SureShrinkWin, RRDEL foi respectivamente de: 0,08439428 segundos; 0,9008196 segundos; 0,2003816 segundos e 0,01245578 segundos.

Figura 11 – Representação gráfica dos resultados da tabela 1 - SNR de saída

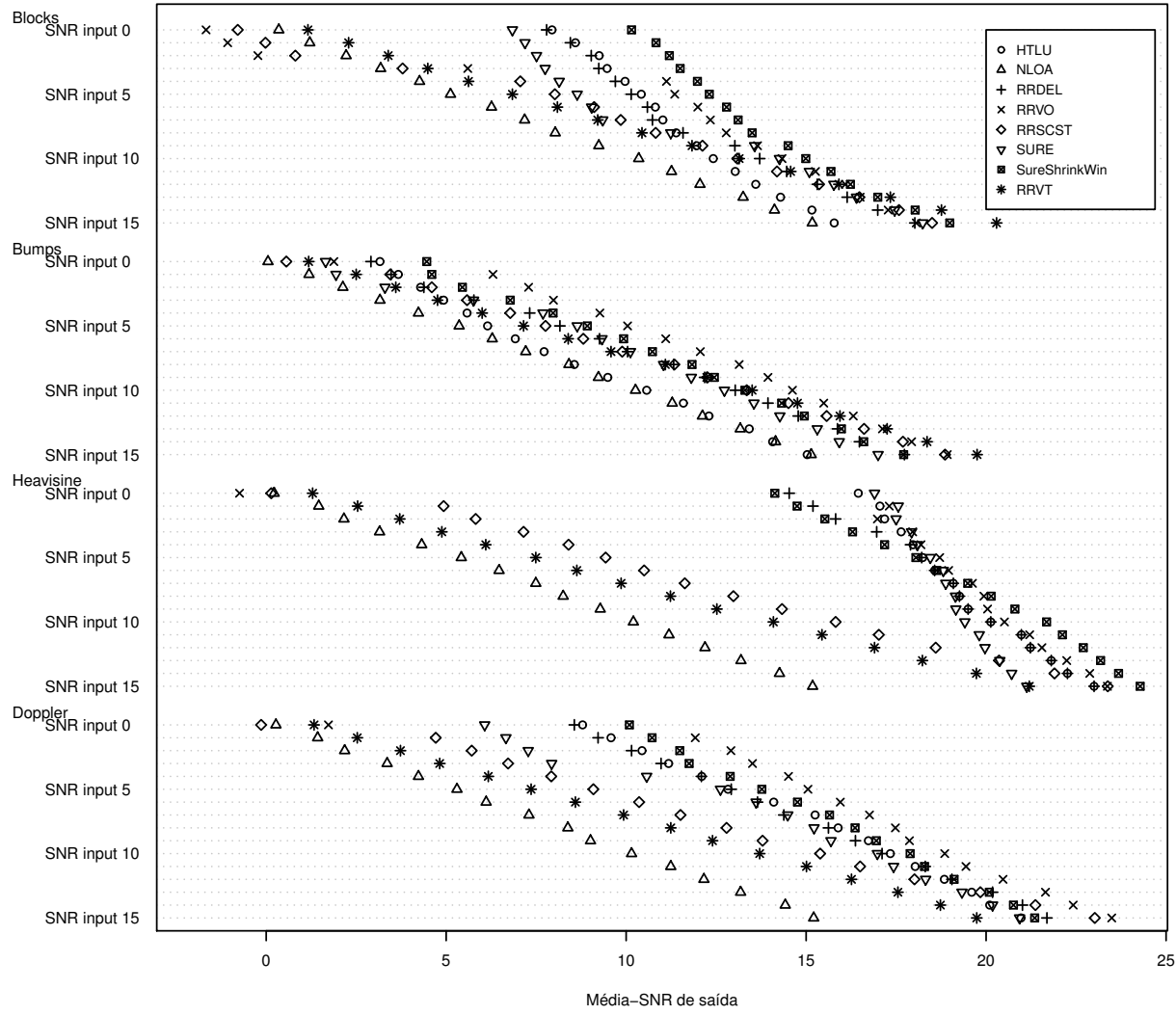


Figura 12 – Representação gráfica dos resultados da tabela 1 - PRD

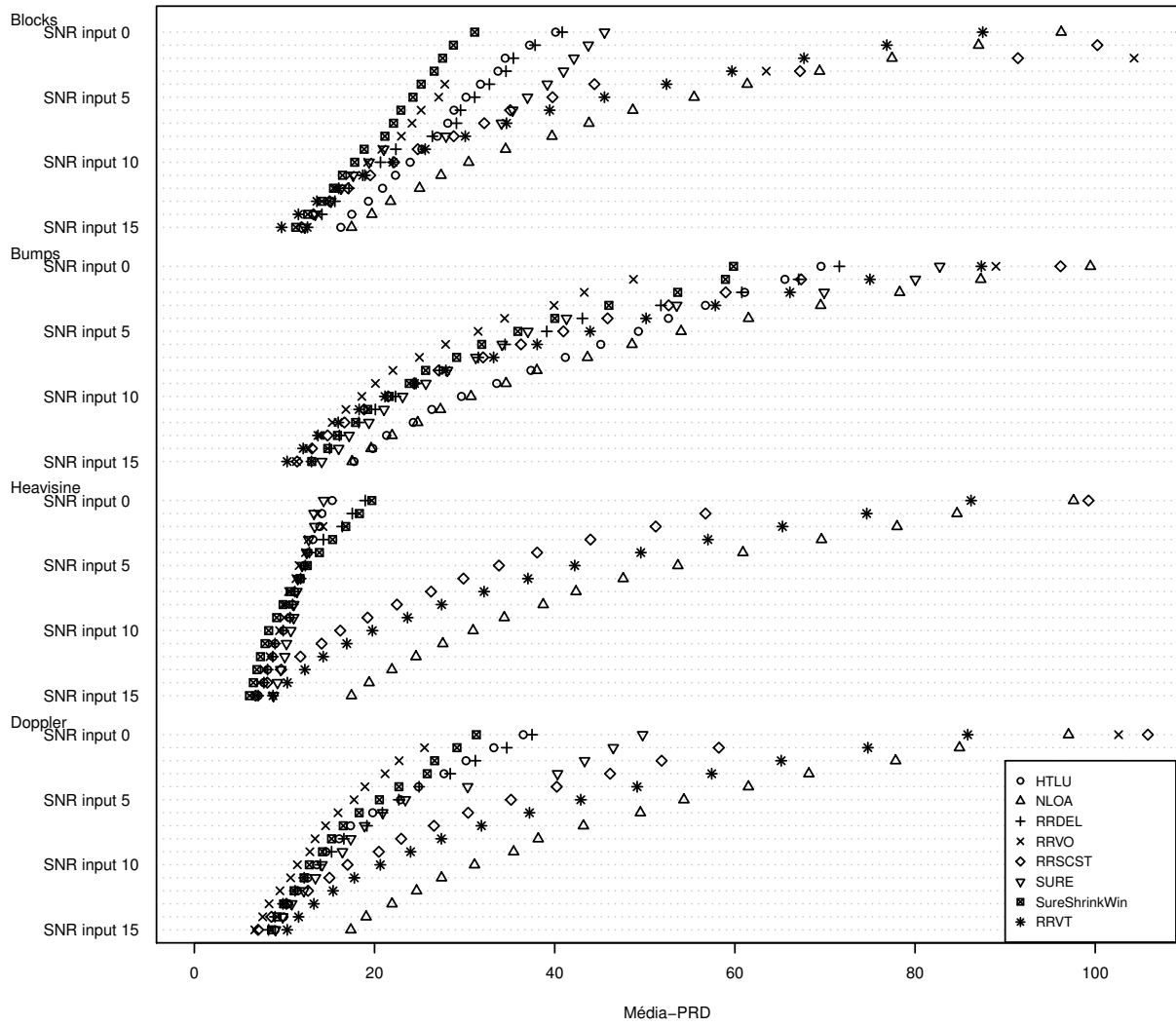


Figura 13 – Representação gráfica dos resultados da tabela 2 - SNR de saída

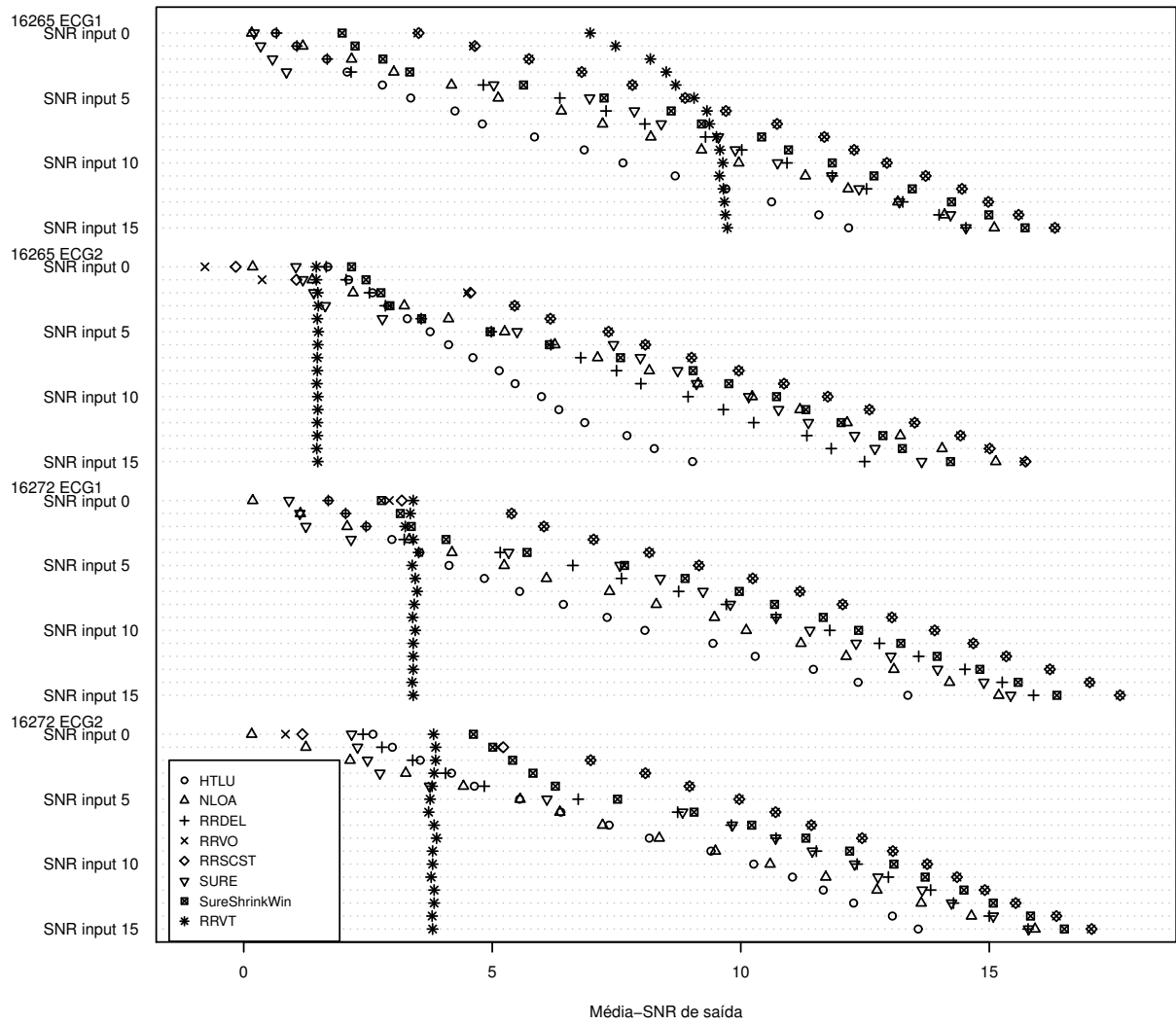


Figura 14 – Representação gráfica dos resultados da tabela 2 - PRD

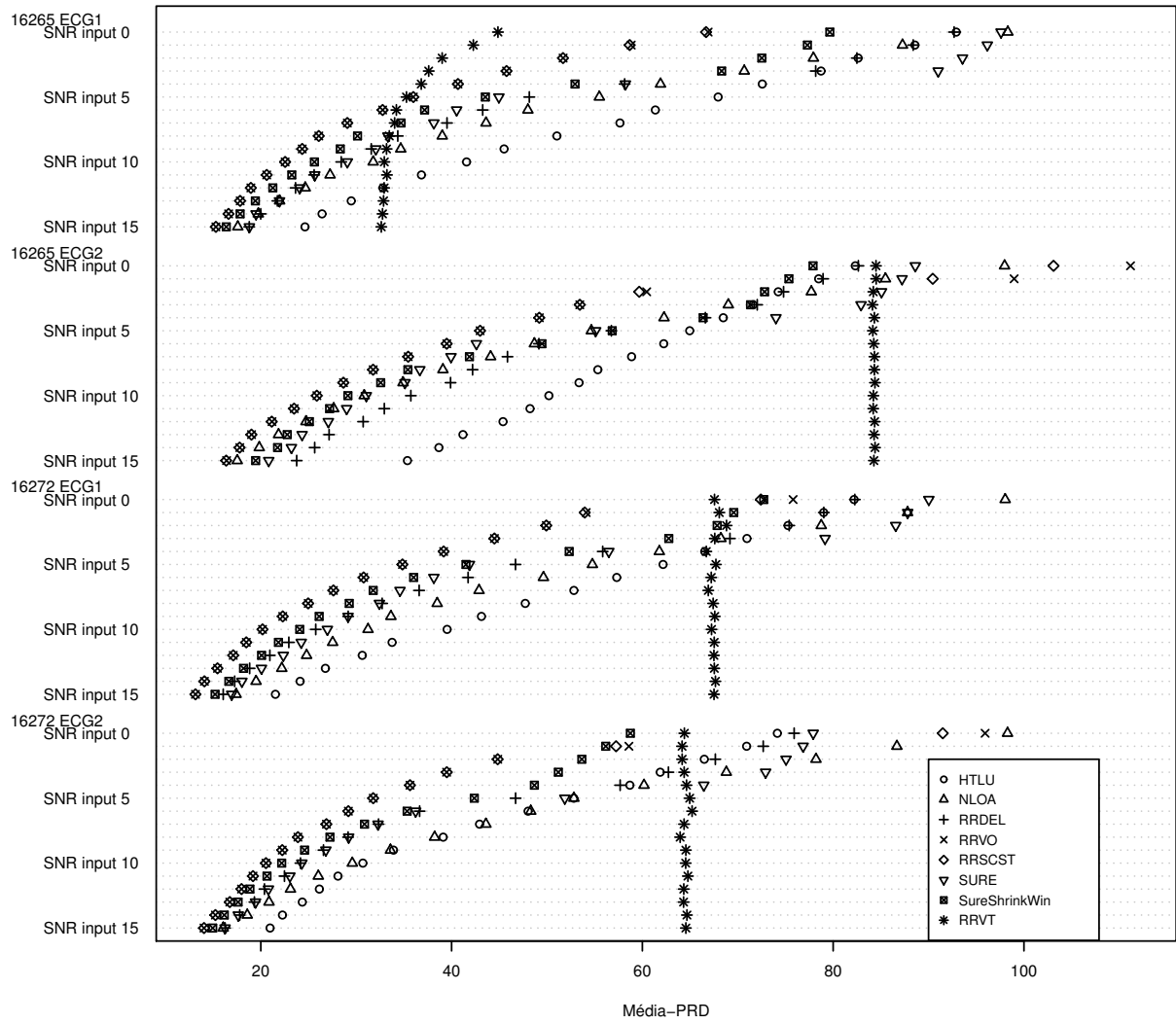


Figura 15 – Representação gráfica dos resultados da tabela 3 - SNR de saída

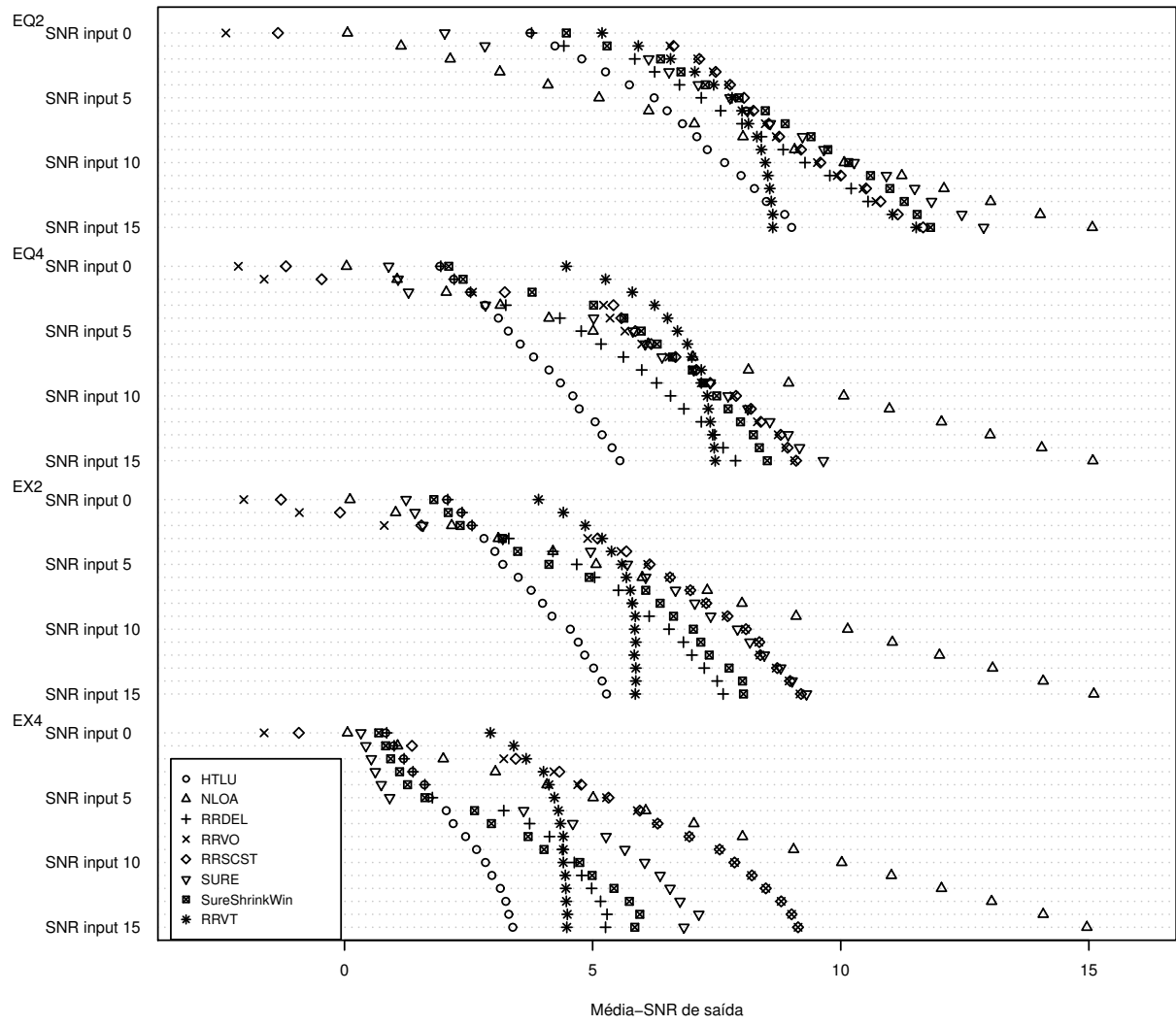
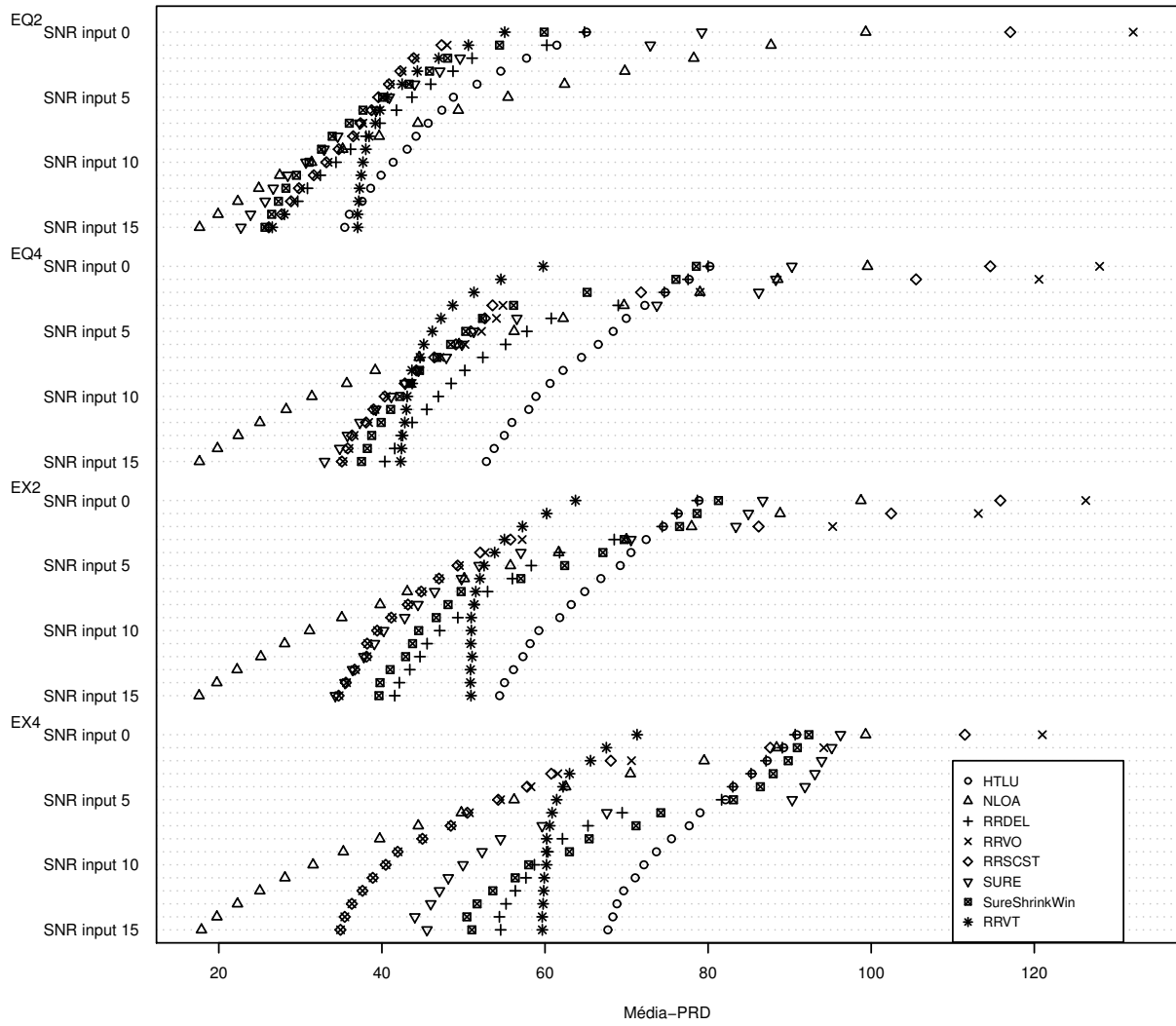


Figura 16 – Representação gráfica dos resultados da tabela 3 - PRD



Aplicações para sinais corrompidos por ruído aditivo de Rayleigh

Neste capítulo será feita a mesma abordagem do capítulo 7 porém, desta vez, será considerado o Ruído aditivo de Rayleigh (veja Definição 4). Os resultados das simulações são mostrados nas Tabelas 4, 5 e 6. A representação gráfica destes resultados podem ser conferidos nas figuras 18, 19, 20, 21, 22 e 23.

Considerando todos os eventos apresentados nas referidas tabelas, é possível verificar que o método RRVO apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVT em, respectivamente, 100%, 90,1%, 97,91% e 89,06% dos casos. Já o método RRSCST apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVT em, respectivamente, 90,62%, 94,27%, 88,54% e 85,93% dos casos.

Enquanto isso o método SureShrinkWin apresentou uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVT em, respectivamente, 86,97%, 65,62%, 64,58% e 88,54% dos casos. Ao passo que o método RRDEL apresenta uma performance igual ou superior aos métodos HTLU, NLOA, SURE e RRVT em, respectivamente, 95,83%, 65,10%, 42,7% e 83,85% dos casos.

Fazendo uma análise mais específica, percebe-se que para o sinal Blocks, o método SureShrinkWin mostrou-se mais eficaz se forem considerados os SNRs de entrada entre 0 e 6 dB. Para SNRs de entrada entre 0 e 6 dB, o método RRVO supera todos os métodos quando considerado o sinal Bumps. Enquanto que o método SureShrinkWin apresenta uma melhor performance para os SNRs entre 0 e 10 dB, quando considerado o sinal Heavisine.

Também é possível perceber a excelente performance do método RRVO para o sinal Doppler, considerando os SNRs de entrada entre 0 e 9 dB. Sendo superado pelo método RRSCST quando considerados os SNRs de entrada entre 10 e 15 dB.

Para os sinais de ECG, os melhores resultados foram obtidos pelos métodos RRVO e RRSCST. O mesmo acontecendo para os traços sísmicos, sendo o método RRSCST ligeiramente melhor que o RRVO na maioria dos casos apresentados na tabela 6.

Uma análise visual interessante pode ser feita através de uma nuvem de palavras como apresentado na figura 17, para produzir este gráfico leva-se em consideração o número de vezes que cada método apresenta a melhor performance. Dito de outra forma é contada a quantidade de vezes que o PRD ou o SNR de saída do método aparece em negrito nas Tabelas 4, 5 e 6. Quanto maior for essa quantidade, maior o tamanho da palavra.

Figura 17 – Nuvem de palavras representado o número de vezes que cada método obteve a melhor performance



8.1 Conclusões

Neste capítulo foram feitas simulações com sinais corrompidos por ruído Rayleigh, os resultados foram interpretados e apresentados tanto na forma de tabela quanto na forma gráfica. Foi possível verificar que os métodos propostos por este trabalho foram eficazes quando comparados a métodos tradicionais na literatura. O tempo computacional médio - para redução de ruído em apenas um sinal - dos métodos RRVO, RRSCST, SureShrinkWin, RRDEL foi respectivamente de: 0,1712724 segundos; 0,9008196 segundos; 0,2130886 segundos e 0,01391753 segundos.

Tabela 4 – Resultados da redução de ruído Rayleigh nos sinais de teste de Donoho & Johnstone, as melhores performances estão apresentadas em negrito.

Sinal	SNRdeentrada(dB)	Média-SNRdesaída(dB)								Média-PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
Blocks	0	0,54	-0,05	0,52	0,65	0,4	0,35	0,72	0,29	93,95	100,55	94,19	92,81	95,56	96,01	92,02	96,71
	1	1,5	0,98	1,5	1,63	1,42	1,39	1,71	1,35	84,11	89,37	84,11	82,88	84,92	85,22	82,15	85,58
	2	2,48	2,02	2,58	2,64	2,46	2,64	2,73	2,44	75,13	79,22	74,3	73,79	75,32	73,84	73,05	75,52
	3	3,44	3,04	3,57	3,64	3,48	3,64	3,71	3,49	67,34	70,51	66,3	65,79	66,97	65,79	65,24	66,93
	4	4,35	4,02	4,53	4,61	4,48	4,59	4,66	4,5	60,59	62,99	59,38	58,82	59,68	58,96	58,52	59,55
	5	5,41	5,14	5,63	5,71	5,62	5,68	5,75	5,67	53,66	55,35	52,33	51,84	52,39	51,98	51,61	52,06
	6	6,21	5,99	6,46	6,52	6,47	6,51	6,58	6,55	48,92	50,21	47,56	47,2	47,48	47,27	46,9	47,05
	7	7,36	7,2	7,65	7,68	7,68	7,69	7,77	7,81	42,88	43,67	41,47	41,31	41,29	41,24	40,89	40,7
	8	8,09	7,97	8,41	8,42	8,46	8,45	8,53	8,61	39,39	39,94	37,96	37,93	37,76	37,8	37,46	37,1
	9	9,29	9,25	9,68	9,66	9,73	9,7	9,77	9,93	34,32	34,49	32,82	32,88	32,61	32,75	32,46	31,88
	10	10,03	10,02	10,45	10,44	10,52	10,46	10,53	10,73	31,51	31,56	30,02	30,05	29,78	30,01	29,74	29,08
	11	11,07	11,08	11,52	11,54	11,6	11,5	11,59	11,83	27,97	27,92	26,54	26,5	26,29	26,62	26,35	25,63
	12	12,15	12,17	12,62	12,65	12,7	12,56	12,66	12,94	24,69	24,64	23,4	23,3	23,17	23,55	23,29	22,54
	13	13,25	13,24	13,69	13,76	13,79	13,62	13,72	14,05	21,76	21,78	20,67	20,52	20,45	20,85	20,62	19,85
	14	14,18	14,17	14,62	14,69	14,71	14,53	14,63	14,99	19,54	19,58	18,58	18,42	18,38	18,77	18,56	17,8
15	15,16	15,13	15,58	15,66	15,67	15,48	15,57	15,98	17,46	17,53	16,64	16,49	16,47	16,83	16,65	15,89	
Bumps	0	0,11	0,01	0,47	0,66	0,43	0,49	0,6	0,39	98,71	99,86	94,77	92,67	95,13	94,55	93,38	95,6
	1	1,21	1,13	1,59	1,78	1,58	1,59	1,7	1,55	86,98	87,77	83,26	81,47	83,4	83,26	82,21	83,67
	2	2,22	2,16	2,61	2,79	2,61	2,59	2,7	2,6	77,47	78,01	74,05	72,56	74,05	74,25	73,26	74,11
	3	3,09	3,04	3,49	3,66	3,5	3,45	3,57	3,51	70,06	70,49	66,95	65,66	66,84	67,24	66,31	66,76
	4	4	3,95	4,39	4,55	4,42	4,34	4,46	4,45	63,11	63,5	60,36	59,24	60,11	60,65	59,82	59,95
	5	5,24	5,19	5,61	5,77	5,68	5,56	5,68	5,72	54,68	55,05	52,41	51,48	52,02	52,75	51,99	51,77
	6	6,12	6,08	6,49	6,65	6,58	6,43	6,56	6,64	49,42	49,68	47,36	46,52	46,87	47,71	47,01	46,59
	7	6,81	6,78	7,19	7,34	7,29	7,12	7,25	7,35	45,66	45,82	43,7	42,94	43,19	44,07	43,43	42,89
	8	8,15	8,15	8,55	8,7	8,68	8,46	8,59	8,75	39,14	39,13	37,37	36,75	36,82	37,74	37,18	36,5
	9	9,01	9,04	9,44	9,58	9,58	9,34	9,47	9,66	35,45	35,33	33,73	33,2	33,21	34,13	33,62	32,89
	10	10,03	10,08	10,49	10,62	10,63	10,36	10,5	10,72	31,52	31,32	29,91	29,44	29,42	30,33	29,87	29,12
	11	11,23	11,3	11,71	11,85	11,85	11,56	11,69	11,94	27,45	27,24	25,98	25,57	25,56	26,42	26,04	25,29
	12	12,08	12,15	12,56	12,7	12,7	12,4	12,53	12,8	24,91	24,69	23,54	23,18	23,17	23,98	23,64	22,91
	13	12,92	12,99	13,41	13,55	13,55	13,24	13,36	13,64	22,6	22,4	21,35	21,02	21,02	21,79	21,48	20,79
	14	14,25	14,31	14,74	14,85	14,85	14,53	14,65	14,94	19,38	19,26	18,33	18,09	18,09	18,77	18,52	17,9
15	15,06	15,11	15,54	15,65	15,65	15,32	15,43	15,73	17,66	17,57	16,72	16,5	16,5	17,15	16,92	16,35	
Heavisine	0	0,96	0,03	0,96	0,97	0,63	0,96	0,98	0,44	89,58	99,63	89,58	89,48	92,97	89,53	89,34	95,07
	1	1,87	0,96	1,87	1,88	1,57	1,87	1,9	1,4	80,68	89,6	80,68	80,56	83,44	80,65	80,39	85,16
	2	2,91	2,01	2,91	2,93	2,65	2,91	2,95	2,49	71,54	79,32	71,54	71,41	73,71	71,56	71,22	75,05
	3	3,95	3,06	3,95	3,96	3,73	3,94	3,99	3,59	63,47	70,29	63,47	63,37	65,11	63,57	63,16	66,18
	4	4,91	4,03	4,91	4,92	4,72	4,88	4,95	4,59	56,84	62,87	56,84	56,78	58,09	57,01	56,54	58,94
	5	6,02	5,16	6,02	6,03	5,88	5,98	6,07	5,76	50,01	55,22	50,01	49,97	50,82	50,23	49,71	51,5
	6	6,84	5,99	6,84	6,85	6,73	6,79	6,89	6,62	45,5	50,19	45,5	45,47	46,07	45,78	45,22	46,65
	7	7,92	7,08	7,92	7,93	7,87	7,85	7,98	7,76	40,19	44,24	40,19	40,13	40,44	40,51	39,92	40,94
	8	9,04	8,23	9,04	9,07	9,04	8,95	9,11	8,94	35,32	38,79	35,32	35,21	35,32	35,69	35,06	35,75
	9	9,73	8,93	9,73	9,77	9,76	9,63	9,8	9,66	32,61	35,77	32,61	32,48	32,5	33,01	32,37	32,89
	10	10,91	10,13	10,91	10,96	10,97	10,78	10,98	10,89	28,47	31,17	28,47	28,33	28,29	28,92	28,26	28,56
	11	11,97	11,2	11,97	12,02	12,04	11,81	12,03	11,98	25,2	27,55	25,2	25,07	25,01	25,68	25,03	25,18
	12	12,75	11,99	12,75	12,79	12,82	12,57	12,8	12,78	23,04	25,16	23,05	22,92	22,86	23,54	22,9	22,97
	13	13,61	12,85	13,6	13,65	13,67	13,45	13,65	13,65	20,88	22,77	20,89	20,78	20,72	21,27	20,77	20,77
	14	14,96	14,23	14,99	15,01	15,03	14,99	15,01	15,03	17,86	19,43	17,8	17,77	17,72	17,8	17,77	17,72
15	15,74	15,01	15,77	15,78	15,8	15,77	15,77	15,81	16,33	17,76	16,28	16,26	16,22	16,27	16,27	16,21	
Doppler	0	1,05	0,23	1,03	1,11	0,81	1,02	1,07	0,65	88,67	97,37	88,86	88,04	91,13	88,93	88,44	92,85
	1	1,81	1	1,78	1,86	1,59	1,77	1,82	1,43	81,23	89,18	81,44	80,71	83,32	81,53	81,1	84,81
	2	2,88	2,08	2,86	2,93	2,69	2,84	2,89	2,55	71,81	78,74	71,96	71,36	73,38	72,08	71,69	74,57
	3	3,72	2,93	3,7	3,78	3,57	3,69	3,74	3,43	65,15	71,36	65,28	64,73	66,33	65,4	65,05	67,37
	4	4,86	4,08	4,84	4,92	4,75	4,83	4,87	4,61	57,14	62,52	57,26	56,79	57,92	57,36	57,07	58,79
	5	5,7	4,94	5,69	5,76	5,63	5,67	5,72	5,49	51,88	56,66	51,93	51,51	52,33	52,05	51,79	53,15
	6	6,61	5,87	6,62	6,69	6,59	6,6	6,63	6,44	46,71	50,9	46,68	46,29	46,85	46,8	46,6	47,63
	7	7,6	6,87	7,62	7,7	7,63	7,59	7,63	7,47	41,68	45,33	41,59	41,22	41,55	41,72	41,55	42,32
	8	8,7	7,98	8,73	8,81	8,78	8,69	8,72	8,59	36,75	39,9	36,61	36,26	36,41	36,76	36,63	37,19
	9	9,64	8,94	9,68	9,77	9,75	9,64	9,66	9,55	32,95	35,75	32,8	32,49	32,54	32,97	32,88	33,31
	10	10,62	9,92	10,66	10,75	10,74	10,61	10,62	10,53	29,45	31,93	29,31	29,03	29,04	29,47	29,43	29,76
	11	11,53	10,84	11,58	11,66	11,66	11,52	11,53	11,44	26,51	28,73	26,38	26,13	26,12	26,54	26,52	26,8
	12	12,79	12,1	12,83	12,9	12,9	12,77	12,77	12,68	22,93	24,85	22,82	22,66	22,64	22,98	22,99	23,24
	13	13,82	13,13	13,86	13,92	13,92	13,79	13,79	13,68	20,37	22,07	20,28	20,15	20,14	20,44	20,45	20,72
	14	14,91	14,22	14,95	15,01	15,02	14,88	14,87	14,73	17,97	19,45	17,88	17,76	17,75	18,03	18,06	18,35
15	15,6	14,91	15,64	15,71	15,71	15,56	15,55	15,38	16,61	17,97	16,52	16,4	16,39	16,67	16,7	17,02	

Tabela 5 – Resultados da redução de ruído Rayleigh nos sinais de eletrocardiograma obtidos a partir da base de dados MIT-BIH, as melhores performances estão apresentadas em negrito,

Sinal	SNR de entrada (dB)	Média-SNR de saída (dB)								Média - PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
16265 ECG1	0	-0,41	0,12	0,32	0,64	0,64	0,39	0,52	0,52	104,83	98,62	96,41	92,93	92,93	95,62	94,2	94,21
	1	0,47	1,03	1,22	1,54	1,54	1,27	1,41	1,32	94,78	88,8	86,93	83,81	83,8	86,44	85	85,94
	2	1,31	1,87	2,05	2,35	2,35	2,07	2,23	2,03	85,97	80,63	79,01	76,29	76,29	78,76	77,36	79,18
	3	2,43	3,02	3,17	3,46	3,46	3,18	3,34	2,96	75,6	70,66	69,4	67,12	67,12	69,38	68,08	71,11
	4	3,38	3,98	4,11	4,39	4,39	4,1	4,27	3,71	67,79	63,29	62,3	60,34	60,34	62,36	61,15	65,27
	5	4,16	4,78	4,88	5,16	5,16	4,86	5,04	4,3	61,95	57,71	57,06	55,22	55,21	57,14	55,97	60,97
	6	5,84	6,52	6,53	6,82	6,82	6,53	6,72	5,47	51,05	47,2	47,17	45,62	45,62	47,15	46,13	53,26
	7	6,05	6,74	6,73	7,02	7,02	6,73	6,93	5,6	49,87	46,05	46,1	44,58	44,58	46,07	45,06	52,45
	8	7,08	7,84	7,75	8,06	8,06	7,78	7,98	6,25	44,26	40,55	40,97	39,56	39,56	40,82	39,91	48,69
	9	7,96	8,81	8,64	8,97	8,97	8,7	8,9	6,76	40	36,28	36,99	35,61	35,61	36,72	35,9	45,92
	10	8,59	9,53	9,31	9,65	9,66	9,4	9,59	7,1	37,19	33,39	34,26	32,91	32,91	33,89	33,17	44,14
	11	9,09	10,12	9,84	10,21	10,22	9,94	10,14	7,36	35,13	31,2	32,22	30,85	30,85	31,85	31,12	42,84
	12	11,28	12,86	12,34	12,85	12,86	12,51	12,7	8,32	27,29	22,74	24,16	22,77	22,76	23,68	23,17	38,37
	13	11,05	12,56	12,06	12,57	12,57	12,23	12,42	8,23	28,04	23,55	24,94	23,53	23,52	24,46	23,93	38,75
	14	12,36	14,26	13,61	14,21	14,21	13,8	14	8,67	24,1	19,36	20,87	19,48	19,47	20,43	19,96	36,86
16265 ECG2	15	12,47	14,4	13,74	14,34	14,35	13,94	14,12	8,7	23,8	19,05	20,57	19,19	19,17	20,11	19,67	36,73
	0	0,1	0,89	0,77	1,22	1,22	1,02	0,98	-1,32	98,87	90,28	91,5	86,95	86,94	88,9	89,36	116,42
	1	0,33	1,18	1,04	1,5	1,5	1,29	1,28	-1,19	96,27	87,36	88,77	84,16	84,16	86,22	86,29	114,65
	2	0,68	1,6	1,43	1,92	1,92	1,69	1,72	-1	92,49	83,18	84,79	80,19	80,18	82,34	82,08	112,15
	3	1,78	2,94	2,69	3,25	3,25	2,97	3,05	-0,46	81,52	71,26	73,34	68,83	68,83	71,07	70,38	105,46
	4	2,46	3,81	3,5	4,09	4,09	3,77	3,88	-0,17	75,33	64,54	66,85	62,47	62,46	64,81	64	101,98
	5	2,77	4,19	3,86	4,47	4,47	4,14	4,25	-0,05	72,7	61,73	64,11	59,8	59,8	62,1	61,33	100,6
	6	5,07	6,99	6,42	7,17	7,18	6,71	6,86	0,61	55,83	44,73	47,76	43,79	43,78	46,19	45,39	93,22
	7	4,25	6	5,53	6,22	6,22	5,8	5,95	0,41	61,35	50,1	52,89	48,87	48,87	51,28	50,44	95,36
	8	5,04	6,96	6,39	7,14	7,15	6,68	6,83	0,6	56,01	44,89	47,9	43,94	43,93	46,33	45,54	93,28
	9	5,17	7,12	6,55	7,3	7,3	6,83	6,98	0,63	55,16	44,06	47,08	43,16	43,15	45,54	44,76	92,97
	10	7,54	10,07	9,18	10,12	10,13	9,53	9,68	1,03	41,98	31,36	34,76	31,18	31,16	33,39	32,82	88,81
	11	6,85	9,18	8,39	9,27	9,27	8,74	8,87	0,93	45,44	34,77	38,05	34,41	34,4	36,58	36,01	89,81
	12	8,15	10,87	9,88	10,89	10,89	10,24	10,39	1,1	39,14	28,61	32,09	28,55	28,54	30,77	30,23	88,08
	13	9,28	12,46	11,23	12,41	12,41	11,66	11,8	1,21	34,38	23,83	27,45	23,96	23,95	26,13	25,7	86,95
	14	11	15,2	13,34	14,97	14,97	13,9	14,07	1,34	28,19	17,37	21,54	17,86	17,84	20,2	19,79	85,73
	15	12,33	17,56	14,91	17	17	15,62	15,83	1,4	24,2	13,25	17,97	14,14	14,12	16,57	16,17	85,15
16272 ECG1	0	-0,24	0,15	0,41	0,72	0,72	0,49	0,59	-0,88	102,78	98,31	95,45	92,11	92,11	94,56	93,42	110,67
	1	1,14	1,58	1,84	2,14	2,14	1,88	2,01	-0,04	87,7	83,4	80,88	78,21	78,2	80,58	79,34	100,43
	2	1,71	2,15	2,43	2,7	2,7	2,44	2,58	0,27	82,11	78,06	75,62	73,26	73,26	75,54	74,33	96,95
	3	2,81	3,25	3,54	3,79	3,79	3,51	3,65	0,8	72,39	68,8	66,57	64,66	64,66	66,8	65,69	91,21
	4	3,17	3,61	3,9	4,14	4,14	3,86	4	0,96	69,47	66	63,85	62,07	62,07	64,16	63,08	89,55
	5	3,64	4,07	4,36	4,6	4,6	4,3	4,46	1,15	65,81	62,57	60,54	58,88	58,88	60,95	59,88	87,57
	6	5,67	6,09	6,34	6,57	6,57	6,25	6,41	1,86	52,08	49,61	48,19	46,93	46,92	48,7	47,82	80,74
	7	6	6,41	6,65	6,88	6,88	6,55	6,71	1,95	50,14	47,83	46,54	45,3	45,3	47,07	46,18	79,89
	8	7,32	7,73	7,91	8,16	8,16	7,82	7,99	2,29	43,07	41,09	40,22	39,09	39,09	40,65	39,87	76,85
	9	8,49	8,93	9,05	9,32	9,32	8,96	9,14	2,53	37,61	35,77	35,29	34,19	34,19	35,64	34,9	74,71
	10	8,52	8,96	9,07	9,35	9,35	8,99	9,17	2,54	37,51	35,66	35,18	34,09	34,09	35,54	34,8	74,67
	11	11,74	12,64	12,51	12,83	12,83	12,42	12,62	3,01	25,88	23,34	23,7	22,82	22,82	23,94	23,38	70,73
	12	10,52	11,18	11,16	11,46	11,46	11,07	11,26	2,86	29,79	27,62	27,67	26,74	26,74	27,97	27,35	71,94
	13	11,62	12,48	12,36	12,69	12,69	12,28	12,48	2,99	26,26	23,77	24,09	23,21	23,21	24,33	23,77	70,85
	14	10,12	10,71	10,72	11,02	11,02	10,63	10,83	2,8	31,2	29,14	29,1	28,13	28,12	29,4	28,75	72,41
	15	11,88	12,8	12,65	12,98	12,98	12,57	12,77	3,02	25,49	22,92	23,31	22,44	22,44	23,53	22,99	70,62
16272 ECG2	0	-0,27	-0,26	0,11	0,3	0,3	0,14	0,16	-1,02	103,14	103,06	98,71	96,62	96,62	98,44	98,24	112,49
	1	1,29	1,34	1,68	1,86	1,86	1,68	1,74	-0,02	86,16	85,69	82,43	80,73	80,73	82,39	81,83	100,21
	2	2,65	2,74	3,02	3,2	3,2	3,02	3,1	0,74	73,72	72,98	70,62	69,21	69,21	70,68	69,99	91,85
	3	2,42	2,5	2,8	2,97	2,97	2,79	2,87	0,62	75,72	75,02	72,49	71,07	71,07	72,56	71,9	93,15
	4	3,55	3,67	3,91	4,09	4,09	3,9	4,01	1,19	66,46	65,55	63,73	62,46	62,46	63,8	63,07	87,24
	5	3,83	3,97	4,2	4,37	4,37	4,19	4,29	1,32	64,33	63,34	61,67	60,46	60,46	61,72	61,02	85,93
	6	5,61	5,87	6	6,18	6,18	6	6,1	2,04	52,42	50,89	50,15	49,11	49,11	50,11	49,52	79,04
	7	7,68	8,2	8,19	8,4	8,4	8,21	8,31	2,69	41,33	38,9	38,98	38,04	38,04	38,86	38,42	73,38
	8	4,37	4,54	4,74	4,91	4,91	4,74	4,84	1,56	60,48	59,3	57,95	56,8	56,8	57,97	57,3	83,6
	9	7,6	8,11	8,1	8,31	8,31	8,13	8,23	2,67	41,69	39,3	39,35	38,41	38,41	39,24	38,8	73,55
	10	9,25	10,05	9,93	10,16	10,16	9,94	10,04	3,04	34,47	31,44	31,89	31,06	31,06	31,85	31,48	70,44
	11	8,74	9,45	9,36	9,58	9,58	9,37	9,48	2,94	36,57	33,71	34,05	33,2	33,2	33,99	33,59	71,28
	12	13,62	15,15	14,65	14,89	14,89	14,61	14,72	3,56	20,84	17,47	18,52	18,01	18,01	18,6	18,37	66,37
	13	11,08	12,19	11,94	12,18	12,18	11,92	12,02	3,33	27,94	24,59	25,31	24,62	24,62	25,36	25,06	68,19
	14	13,97	15,55	15	15,25	15,25	14,96	15,07	3,58	20,04	16,69	17,79	17,28	17,28	17,87	17,63	66,2
	15	13,71	15,25	14,74	14,98	14,98	14,7	14,81	3,57	20,63	17,28	18,33	17,82	17,82	18,41	18,19	66,32

Tabela 6 – Resultados da redução de ruído Rayleigh nos traços sísmicos obtidos a partir do pacote R astsa, as melhores performances estão apresentadas em negrito,

Sinal	SNR de entrada (dB)	Média-SNR de saída (dB)								Média - PRD							
		HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT	HTLU	NLOA	RRDEL	RRVO	RRSCST	SURE	SureShrinkWin	RRVT
EQ2	0	0,05	0,03	0,39	0,55	0,55	0,48	0,33	0,16	99,38	99,69	95,65	93,89	93,88	94,66	96,24	98,18
	1	1,14	1,23	1,54	1,69	1,69	1,63	1,44	1,13	87,71	86,79	83,8	82,29	82,28	82,92	84,76	87,81
	2	1,84	2,01	2,28	2,45	2,45	2,37	2,17	1,73	80,91	79,34	76,92	75,44	75,43	76,08	77,93	81,96
	3	3,08	3,39	3,59	3,78	3,78	3,69	3,46	2,72	70,16	67,71	66,12	64,69	64,69	65,37	67,16	73,14
	4	3,72	4,11	4,28	4,49	4,49	4,38	4,13	3,2	65,15	62,34	61,1	59,66	59,65	60,4	62,19	69,22
	5	4,78	5,3	5,42	5,65	5,65	5,53	5,19	3,93	57,67	54,31	53,57	52,17	52,16	52,92	55,03	63,59
	6	5,57	6,21	6,29	6,52	6,53	6,4	5,98	4,44	52,64	48,94	48,48	47,18	47,17	47,88	50,25	60,01
	7	6,34	7,07	7,1	7,33	7,34	7,21	6,7	4,87	48,2	44,33	44,14	42,98	42,97	43,61	46,22	57,09
	8	7,18	8	8	8,21	8,22	8,12	7,48	5,29	43,77	39,79	39,82	38,85	38,83	39,28	42,29	54,36
	9	8,31	9,28	9,2	9,42	9,42	9,34	8,48	5,8	38,44	34,35	34,66	33,82	33,81	34,14	37,67	51,32
	10	9,47	10,57	10,4	10,63	10,64	10,56	9,44	6,21	33,62	29,61	30,21	29,41	29,38	29,65	33,74	48,91
	11	9,52	10,62	10,44	10,69	10,69	10,61	9,47	6,23	33,44	29,43	30,05	29,23	29,2	29,49	33,6	48,82
	12	10,93	12,22	11,92	12,23	12,24	12,12	10,56	6,63	28,42	24,48	25,34	24,45	24,42	24,78	29,67	46,6
	13	10,09	11,27	11,04	11,3	11,31	11,22	9,93	6,4	31,3	27,33	28,04	27,22	27,19	27,49	31,87	47,84
	14	12,26	13,81	13,41	13,8	13,82	13,62	11,49	6,93	24,39	20,39	21,37	20,41	20,38	20,85	26,63	45
EQ4	15	13,41	15,25	14,74	15,23	15,25	14,99	12,23	7,14	21,36	17,28	18,32	17,33	17,28	17,81	24,45	43,96
	0	0,11	0,09	0,44	0,6	0,6	0,54	0,39	0,21	98,77	99,02	95,02	93,3	93,29	94,03	95,64	97,63
	1	0,87	0,94	1,26	1,42	1,42	1,35	1,17	0,9	90,44	89,75	86,52	84,96	84,96	85,6	87,38	90,17
	2	2,14	2,34	2,6	2,77	2,77	2,69	2,48	1,97	78,17	76,37	74,16	72,69	72,68	73,37	75,18	79,67
	3	3,07	3,38	3,59	3,78	3,78	3,69	3,46	2,71	70,2	67,75	66,16	64,73	64,72	65,41	67,19	73,17
	4	3,38	3,72	3,91	4,11	4,11	4,01	3,77	2,94	67,8	65,18	63,76	62,32	62,31	63,06	64,81	71,28
	5	4,46	4,94	5,07	5,3	5,3	5,18	4,87	3,72	59,85	56,64	55,76	54,35	54,34	55,11	57,08	65,2
	6	5,33	5,94	6,03	6,27	6,27	6,14	5,75	4,29	54,12	50,49	49,95	48,61	48,6	49,34	51,61	61,02
	7	6,1	6,8	6,85	7,08	7,09	6,96	6,48	4,74	49,52	45,73	45,46	44,24	44,23	44,9	47,44	57,95
	8	7,13	7,95	7,95	8,16	8,17	8,06	7,43	5,27	44,02	40,04	40,06	39,07	39,06	39,54	42,53	54,5
	9	8,28	9,26	9,18	9,39	9,4	9,31	8,46	5,79	38,53	34,45	34,75	33,92	33,9	34,23	37,74	51,37
	10	9,17	10,24	10,09	10,32	10,33	10,24	9,21	6,11	34,8	30,75	31,29	30,48	30,46	30,76	34,65	49,47
	11	10,43	11,66	11,41	11,68	11,69	11,59	10,19	6,5	30,09	26,11	26,89	26,05	26,02	26,32	30,96	47,3
	12	10,59	11,84	11,58	11,86	11,87	11,77	10,3	6,55	29,53	25,58	26,38	25,53	25,5	25,8	30,54	47,07
	13	11,33	12,69	12,36	12,69	12,71	12,57	10,85	6,73	27,13	23,19	24,09	23,19	23,16	23,53	28,67	46,07
	14	11,12	12,44	12,13	12,45	12,46	12,33	10,69	6,68	27,81	23,88	24,75	23,87	23,84	24,19	29,2	46,35
	15	14,81	17,04	16,41	16,95	17	16,68	13	7,32	18,18	14,06	15,11	14,2	14,13	14,65	22,4	43,03
EX2	0	-0,05	-0,08	0,28	0,44	0,44	0,37	0,23	0,07	100,55	100,98	96,83	95,04	95,03	95,85	97,38	99,23
	1	0,97	1,05	1,36	1,52	1,52	1,45	1,27	0,98	89,45	88,66	85,53	83,98	83,97	84,62	86,43	89,3
	2	1,63	1,79	2,06	2,23	2,23	2,16	1,96	1,56	82,86	81,43	78,85	77,39	77,38	77,98	79,81	83,59
	3	2,73	3	3,23	3,41	3,41	3,33	3,1	2,45	73,02	70,78	68,97	67,54	67,53	68,2	69,99	75,43
	4	4,11	4,54	4,7	4,91	4,91	4,8	4,52	3,47	62,33	59,29	58,23	56,83	56,82	57,55	59,45	67,04
	5	4,75	5,27	5,39	5,62	5,62	5,49	5,16	3,91	57,89	54,54	53,78	52,38	52,37	53,13	55,24	63,75
	6	5,65	6,29	6,37	6,61	6,61	6,48	6,05	4,48	52,17	48,45	48,03	46,73	46,72	47,43	49,83	59,69
	7	5,92	6,6	6,66	6,9	6,9	6,77	6,31	4,64	50,56	46,78	46,45	45,2	45,19	45,88	48,35	58,62
	8	6,82	7,61	7,62	7,84	7,84	7,73	7,15	5,12	45,6	41,66	41,6	40,56	40,55	41,08	43,92	55,46
	9	7,85	8,77	8,72	8,93	8,94	8,85	8,09	5,6	40,49	36,44	36,63	35,76	35,75	36,11	39,41	52,45
	10	8,82	9,86	9,74	9,96	9,97	9,87	8,93	5,99	36,21	32,13	32,59	31,76	31,74	32,09	35,79	50,16
	11	9,93	11,08	10,87	11,13	11,14	11,04	9,8	6,36	31,89	27,91	28,6	27,78	27,75	28,05	32,36	48,11
	12	10,32	11,53	11,28	11,55	11,56	11,47	10,1	6,47	30,49	26,52	27,28	26,45	26,42	26,71	31,28	47,48
	13	11,76	13,21	12,84	13,2	13,22	13,05	11,16	6,83	25,82	21,85	22,8	21,87	21,84	22,26	27,69	45,55
	14	12,48	14,08	13,65	14,06	14,08	13,87	11,63	6,98	23,78	19,78	20,77	19,81	19,77	20,24	26,2	44,79
	15	12,49	14,09	13,66	14,08	14,09	13,88	11,64	6,98	23,75	19,75	20,74	19,78	19,74	20,22	26,18	44,78
EX4	0	0,08	0,05	0,41	0,57	0,57	0,5	0,36	0,18	99,12	99,4	95,39	93,64	93,63	94,39	95,98	97,95
	1	0,68	0,72	1,05	1,21	1,21	1,14	0,97	0,72	92,49	92,07	88,64	87,04	87,03	87,73	89,41	92,02
	2	2	2,18	2,44	2,62	2,62	2,54	2,33	1,86	79,45	77,77	75,48	73,98	73,97	74,67	76,49	80,75
	3	2,58	2,83	3,07	3,25	3,25	3,16	2,95	2,33	74,31	72,19	70,27	68,83	68,82	69,49	71,24	76,48
	4	3,68	4,06	4,23	4,44	4,44	4,33	4,08	3,16	65,47	62,69	61,42	59,98	59,97	60,72	62,49	69,47
	5	4,55	5,04	5,17	5,39	5,4	5,27	4,95	3,78	59,24	56	55,15	53,74	53,74	54,51	56,53	64,75
	6	5,04	5,6	5,71	5,94	5,94	5,82	5,45	4,1	56,01	52,5	51,84	50,48	50,48	51,2	53,41	62,36
	7	6,11	6,8	6,85	7,09	7,09	6,96	6,48	4,74	49,5	45,7	45,44	44,22	44,21	44,88	47,42	57,94
	8	6,93	7,72	7,73	7,95	7,95	7,84	7,24	5,17	45,06	41,11	41,06	40,06	40,04	40,54	43,45	55,13
	9	7,86	8,77	8,73	8,94	8,94	8,85	8,09	5,61	40,48	36,43	36,62	35,75	35,73	36,1	39,41	52,45
	10	8,7	9,72	9,6	9,83	9,83	9,74	8,81	5,95	36,75	32,67	33,1	32,27	32,24	32,6	36,26	50,44
	11	9,91	11,07	10,86	11,11	11,12	11,03	9,79	6,35	31,94	27,96	28,65	27,83	27,8	28,1	32,4	48,13
	12	11,22	12,56	12,24	12,56	12,57	12,44	10,77	6,7	27,49	23,56	24,45	23,54	23,51	23,89	28,95	46,22
	13	11,62	13,03	12,68	13,02	13,04	12,89	11,05	6,8	26,26	22,31	23,23	22,33	22,3	22,68	28,01	45,72
	14	12,27	13,83	13,42	13,82	13,83	13,64	11,5	6,94	24,35	20,35	21,33	20,38	20,35	20,81	26,62	44,99
	15	13,8	15,75	15,21	15,71	15,74	15,46	12,47	7,2	20,42	16,32	17,37	16,38	16,33	16,87	23,81	43,66

Figura 18 – Representação gráfica dos resultados da tabela 4 - SNR de saída

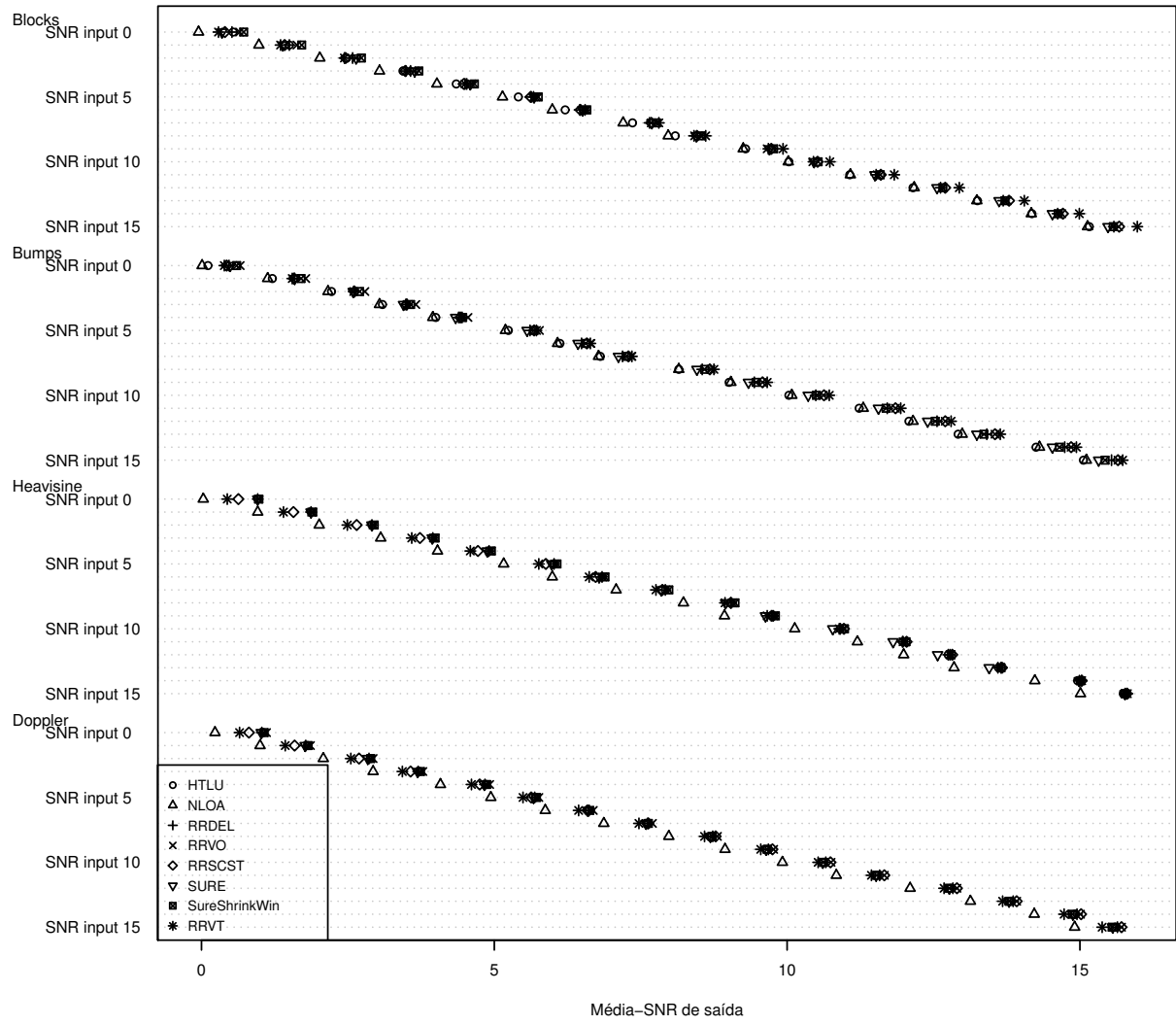


Figura 19 – Representação gráfica dos resultados da tabela 4 - PRD

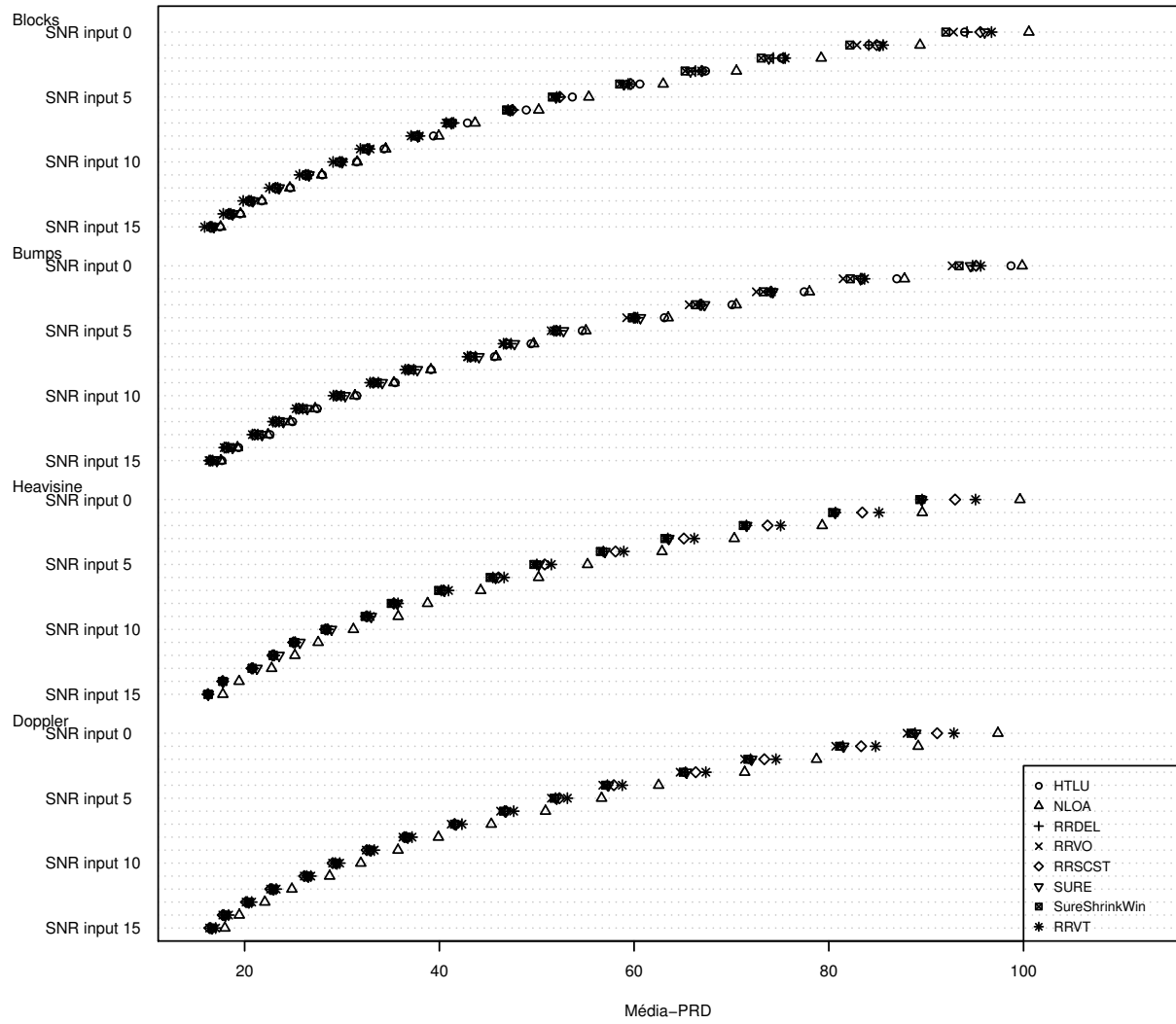


Figura 20 – Representação gráfica dos resultados da tabela 5 - SNR de saída

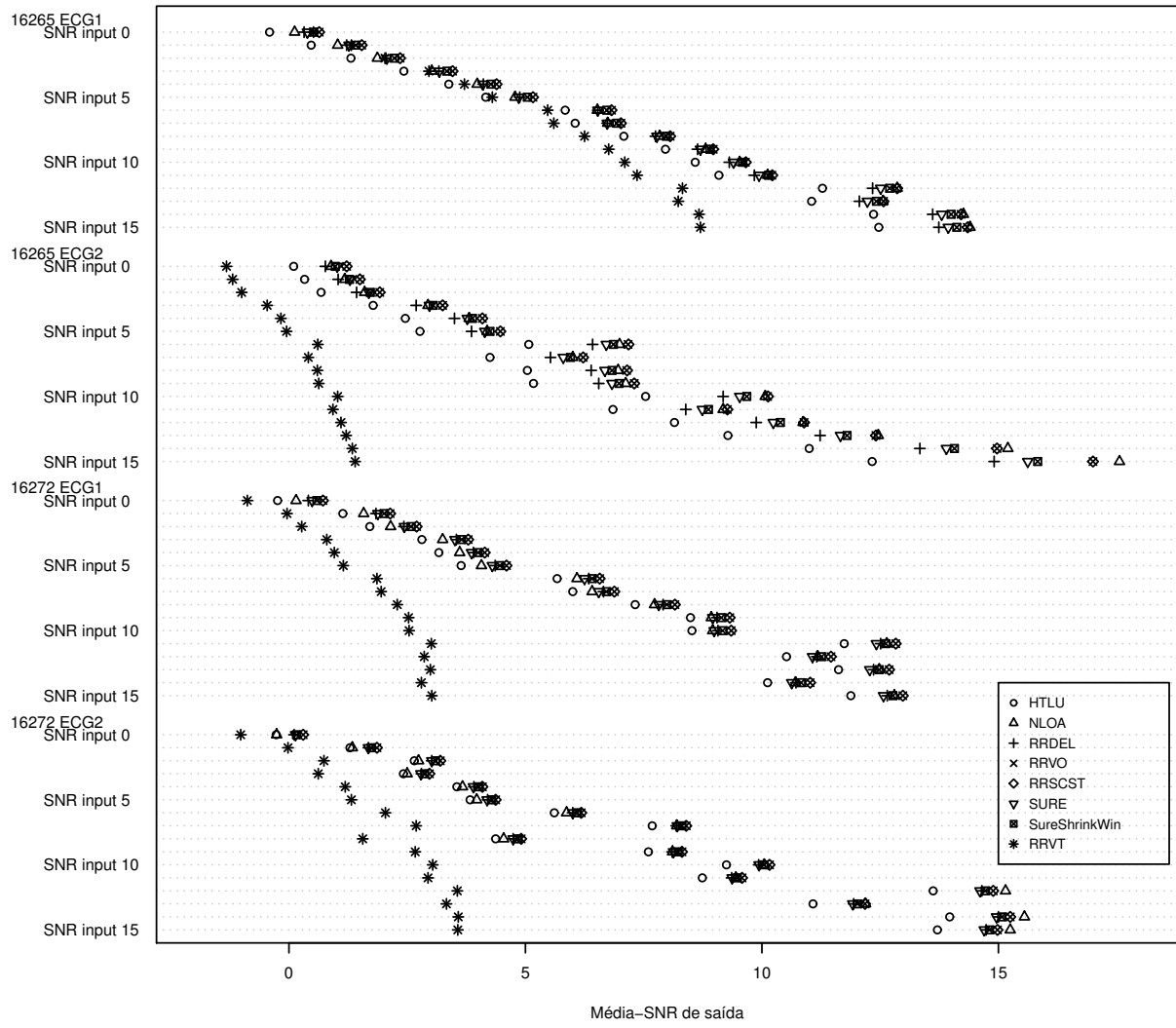


Figura 22 – Representação gráfica dos resultados da tabela 6 - SNR de saída

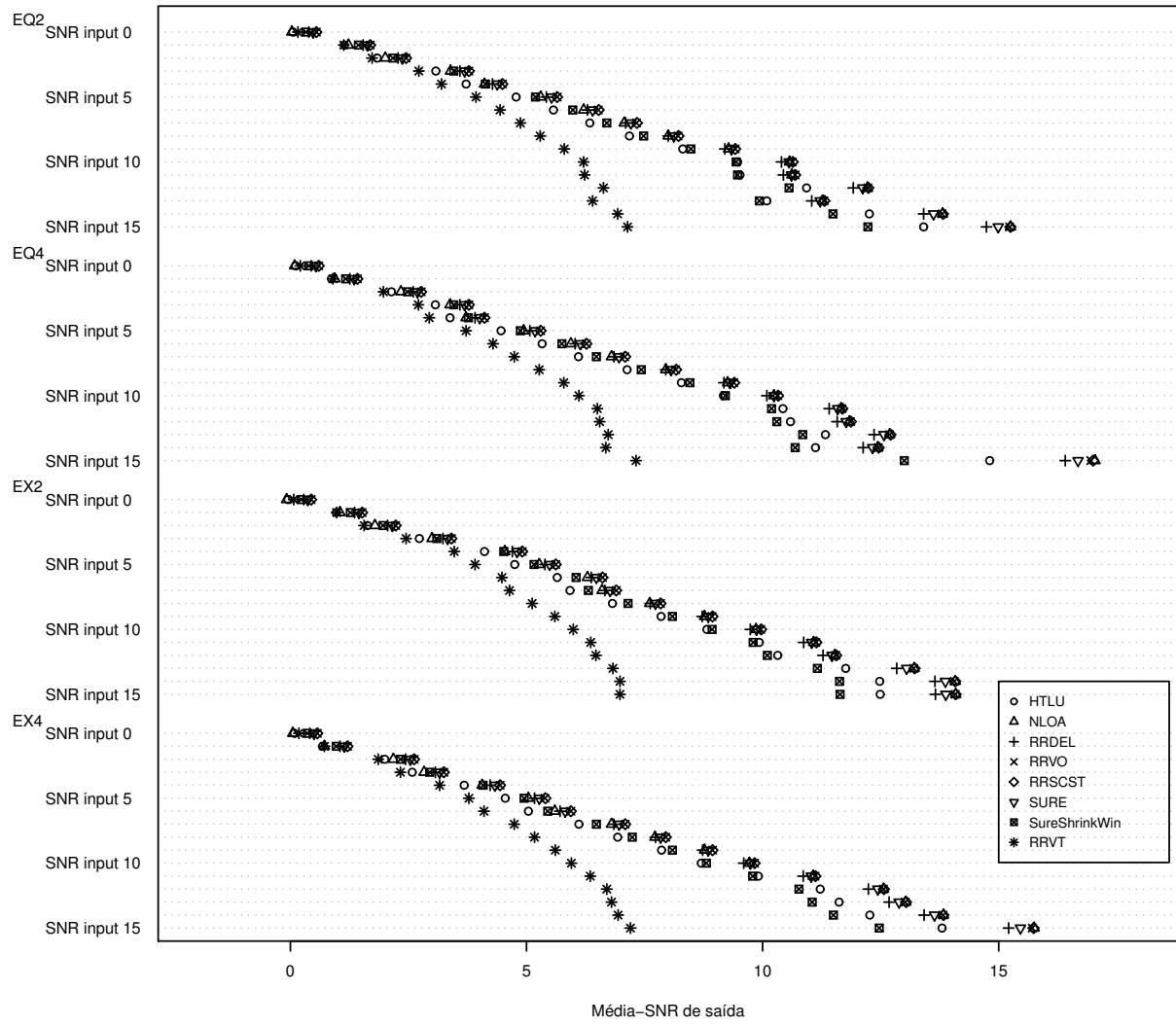
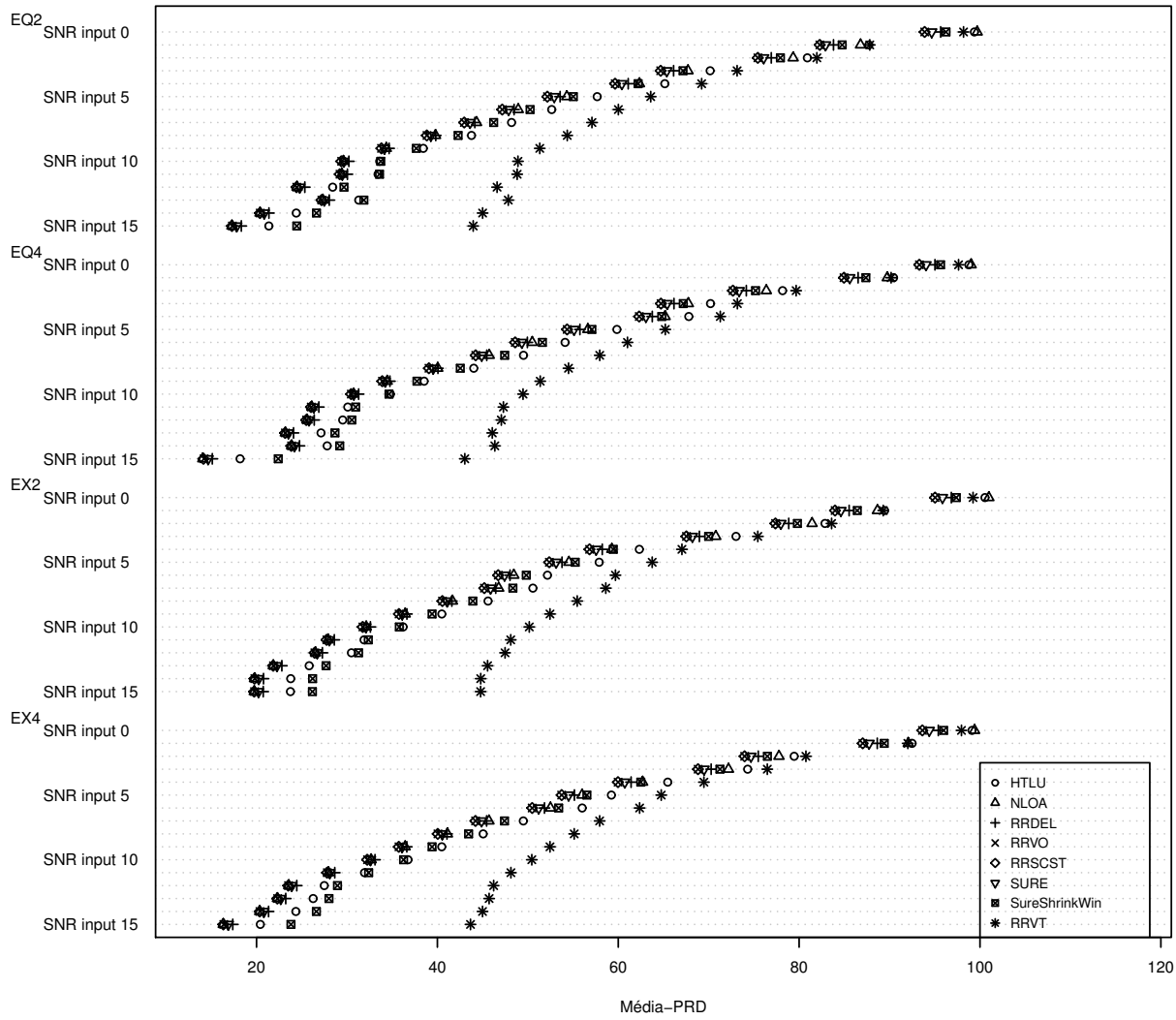


Figura 23 – Representação gráfica dos resultados da tabela 6 - PRD



Conclusões

Este trabalho busca contribuir com a área de pesquisa associada à redução de ruído em sinais, em especial sinais não estacionários. Inicialmente, contribuiu-se com a utilização do algoritmo de Viterbi aplicado no domínio de ondaletas e com a introdução de um processo decisório baseado no erro quadrático médio. Este tipo de abordagem permitiu separar os coeficientes de ondaletas de detalhe em dois grupos, um representativo de sinal, e outro representativo de ruído. Este método foi denominado por redução de ruído por meio do algoritmo de Viterbi e ondaletas (RRVO).

Uma outra colaboração deste trabalho foi o método Redução de ruído em sinal por clusterização e soft-thresholding (RRSCST). Este método busca verificar se o RRVO não deixou para trás alguma informação de sinal. A forma de se fazer esta verificação consiste na estimativa de um limiar e na aplicação, com base neste limiar, de uma função de soft-thresholding no conjunto de coeficientes de ondaletas de detalhe previamente classificados como ruidosos.

Além destes, também foram propostos os métodos SureShrinkWin e RRDEL. O método SureShrinkWin aplica o método SureShrink em janelas de diferentes tamanhos e estima o sinal limpo por meio da mediana das estimativas obtidas em cada janela. Enquanto que o método RRDEL aplica um processo decisório nos coeficientes de ondaletas de detalhe a partir da estimativa dupla de limiares.

Cabe destacar a novidade destes métodos que surgem a partir de teorias que raramente aparecem na literatura relacionada ao *denoising*. Sendo assim, o presente trabalho pode ser interpretado como uma nova fonte de pesquisa que, além de propor métodos novos, aplica-os de forma rigorosa na redução de ruídos de diversos sinais e compara, seguindo simulações de Monte Carlo, a performance destes novos métodos com a performance de métodos já bem estabelecidos na literatura.

As implementações exigiram conhecimento de programação na linguagem R. Os dois métodos propostos foram implementados por meio desta linguagem, o que consistiu de um trabalho de pesquisa extenso e de várias horas de programação. Esforço que se repetiu também na implementação dos métodos já consagrados na literatura e na busca da base

de dados onde foram feitas as simulações. As quais por sua vez, tomaram bastante tempo de processamento devido às várias replicações que devem ser feitas quando métodos de redução de ruído são aplicados seguindo a simulação de Monte Carlo. Este trabalho rendeu duas publicações (VARGAS; VEIGA, 2017; VARGAS; VEIGA, 2018) e espera-se como pesquisa futura a proposição de um novo método para redução de ruído em sinais de eletrocardiograma por meio de uma nova função de risco, a ser minimizada por algoritmos genéticos.

Referências

ANDERSSON, F. N. Core inflation. 2008.

BENAZZA-BENYAHIA, A.; PESQUET, J.-C. An extended sure approach for multicomponent image denoising. In: IEEE. **Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on**. [S.l.], 2004. v. 2, p. ii-945.

CHANDRAKAR, C.; KOWAR, M. Denoising ecg signals using adaptive filter algorithm. **International Journal of Soft Computing and Engineering (IJSCE)**, v. 2, n. 1, p. 120-123, 2012.

CONDAT, L. A direct algorithm for 1d total variation denoising. **IEEE Signal Processing Letters**, v. 20, n. 11, p. 1054-1057, 2013. Disponível em: <<https://doi.org/10.1109/LSP.2013.2278339>>.

DONOHU, D. L.; JOHNSTONE, I. M. Adapting to unknown smoothness via wavelet shrinkage. **Journal of the american statistical association**, Taylor & Francis Group, v. 90, n. 432, p. 1200-1224, 1995.

FENG, L. et al. Implementation of de-noise dwf chip based on adaptive soft-threshold. In: IEEE. **Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on**. [S.l.], 2000. v. 1, p. 614-618.

FLIGGE, M.; SOLANKI, S. Noise reduction in astronomical spectra using wavelet packets. **Astronomy and Astrophysics Supplement Series**, EDP Sciences, v. 124, n. 3, p. 579-587, 1997.

HAN, G.; XU, Z. Electrocardiogram signal denoising based on a new improved wavelet thresholding. **Review of Scientific Instruments**, AIP Publishing, v. 87, n. 8, p. 084303, 2016.

KOLACZYK, E. D. A wavelet shrinkage approach to tomographic image reconstruction. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 91, n. 435, p. 1079-1090, 1996.

KOZAKEVICIUS, A. de J. et al. Adaptive ecg filtering and qrs detection using orthogonal wavelet transform. **Biomedical Engineering, Innsbruck, Austria**, 2005.

- LIU, Y. et al. Applications of savitzky-golay filter for seismic random noise reduction. **Acta Geophysica**, v. 64, n. 1, p. 101–124, 2016. Disponível em: <<https://doi.org/10.1515/acgeo-2015-0062>>.
- LOTRIČ, U.; DOBNIKAR, A. Wavelet based smoothing in time series prediction with neural networks. In: SPRINGER. **Artificial Neural Nets and Genetic Algorithms**. [S.l.], 2001. p. 43–46.
- MEYER, Y. Wavelets-algorithms and applications. **Wavelets-Algorithms and applications Society for Industrial and Applied Mathematics Translation.**, 142 p., v. 1, 1993.
- MOONEY, C. Z. **Monte carlo simulation**. Sage Publications, 1997. v. 116. Disponível em: <<https://doi.org/10.4135/9781412985116>>.
- MOYO, P.; BROWNJOHN, J. Detection of anomalous structural behaviour using wavelet analysis. **Mechanical Systems and Signal Processing**, Elsevier, v. 16, n. 2-3, p. 429–445, 2002. Disponível em: <<https://doi.org/10.1006/mssp.2001.1449>>.
- NING, X.; SELESNICK, I. W. Ecg enhancement and qrs detection based on sparse derivatives. **Biomedical Signal Processing and Control**, Elsevier, v. 8, n. 6, p. 713–723, 2013. Disponível em: <<https://doi.org/10.1016/j.bspc.2013.06.005>>.
- PERCIVAL, D. B.; WALDEN, A. T. **Wavelet methods for time series analysis**. [S.l.]: Cambridge university press, 2006. v. 4.
- PERRONE, D.; FAVARO, P. A clearer picture of total variation blind deconvolution. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 38, n. 6, p. 1041–1055, 2016.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>.
- RAMSEY, J. B. The contribution of wavelets to the analysis of economic and financial data. **Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 357, n. 1760, p. 2593–2606, 1999.
- REMÉNYI, N.; VIDA KOVIC, B. Wavelet shrinkage with double weibull prior. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 44, n. 1, p. 88–104, 2015. Disponível em: <<https://doi.org/10.1080/03610918.2013.765470>>.
- SADASIVAN, J.; MUKHERJEE, S.; SEELAMANTULA, C. S. An optimum shrinkage estimator based on minimum-probability-of-error criterion and application to signal denoising. In: IEEE. **Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on**. [S.l.], 2014. p. 4249–4253.
- SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications: with R examples**. [S.l.]: Springer Science & Business Media, 2010.

SINGH, K.; RANADE, S. K.; SINGH, C. Comparative performance analysis of various wavelet and nonlocal means based approaches for image denoising. **Optik-International Journal for Light and Electron Optics**, Elsevier, v. 131, p. 423–437, 2017. Disponível em: <<https://doi.org/10.1016/j.ijleo.2016.11.055>>.

STEIN, C. M. Estimation of the mean of a multivariate normal distribution. **The annals of Statistics**, JSTOR, p. 1135–1151, 1981.

VARGAS, R. N.; VEIGA, A. C. P. Seismic trace noise reduction by wavelets and double threshold estimation. **IET Signal Processing**, IET, v. 11, n. 9, p. 1069–1075, 2017. Disponível em: <<http://dx.doi.org/10.1049/iet-spr.2017.0061>>.

_____. Electrocardiogram signal denoising by clustering and soft thresholding. **IET Signal Processing**, IET, 2018. Disponível em: <<https://doi.org/10.1049/iet-spr.2018.5162>>.

WASSERMANN, L. **All of nonparametric statistics**. [S.l.]: Springer Science+Business Media, New York, 2006.

WEI, G.; WU, S. Denoising radar signals using complex wavelet. In: IEEE. **Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on**. [S.l.], 2003. v. 1, p. 341–344.

YOUNG, K.; SOHER, B. J.; MAUDSLEY, A. A. Automated spectral analysis ii: Application of wavelet shrinkage for characterization of non-parameterized signals. **Magnetic resonance in medicine**, Wiley Online Library, v. 40, n. 6, p. 816–821, 1998.

ZUCCHINI, W.; MACDONALD, I. L. **Hidden Markov models for time series: an introduction using R**. [S.l.]: CRC press, 2009. v. 150.